Discrete Time Stochastic Processes

Joseph C. Watkins

May 5, 2007

Contents

1 Basic Concepts for Stochastic Processes

In this section, we will introduce three of the most versatile tools in the study of random processes - conditional expectation with respect to a σ-algebra, stopping times with respect to a filtration of σ-algebras, and the coupling of two stochastic processes.

1.1 Conditional Expectation

Information will come to us in the form of σ-algebras. The notion of conditional expectation $E[Y|\mathcal{G}]$ is to make the best estimate of the value of Y given a σ-algebra $\mathcal{G}$.

For example, let $\{C_i; i \geq 1\}$ be a countable partitiion of Ω, i. e., $C_i \cap C_j = \emptyset$, whenever $i \neq j$ and $\bigcup_{i\geq 1} C_i = \Omega$. Then, the σ-algebra, $\mathcal{G}$, generated by this partition consists of unions from all possible subsets of the partition. For a random variable Z to be $\mathcal{G}$-measurable, then the sets $\{Z \in B\}$ must be such a union. Consequently, Z is $\mathcal{G}$-measurable if and only if it is constant on each of the C_i's.

The natural manner to set $E[Y|\mathcal{G}](\omega)$ is to check which set in the partition contains ω and average Y over that set. Thus, if $\omega \in C_i$ and $P(C_i) > 0$, then

$$E[Y|\mathcal{G}](\omega) = E[Y|C_i] = \frac{E[Y; C_i]}{P(C_i)}.$$

If $P(C_i) = 0$, then any value $E[Y|\mathcal{G}](\omega)$ (say 0) will suffice, In other words,

$$E[Y|\mathcal{G}] = \sum_i E[Y|C_i] I_{C_i}. \tag{1.1}$$

We now give the definition. Check that the definition of $E[Y|\mathcal{G}]$ in (1.1) has the two given conditions.

Definition 1.1. *Let Y be an integrable random variable on $(\Omega, \mathcal{F}, P)$ and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. The* conditional expectation *of Y given $\mathcal{G}$, denoted $E[Y|\mathcal{G}]$ is the a.s. unique random variable satisfying the following two conditions.*

1. *$E[Y|\mathcal{G}]$ is $\mathcal{G}$-measurable.*
2. *$E[E[Y|\mathcal{G}]; A] = E[Y; A]$ for any $A \in \mathcal{G}$.*

To see that that $E[Y|\mathcal{G}]$ as defined in (1.1) follows from these two conditions, note that by condition 1, $E[Y|\mathcal{G}]$ must be constant on each of the C_i. By condition 2, for $\omega \in C_i$,

$$E[Y|\mathcal{G}](\omega)P(C_i) = E[E[Y|\mathcal{G}]; C_i] = E[Y; C_i] \quad \text{or} \quad E[Y|\mathcal{G}](\omega) = \frac{E[Y; C_i]}{P(C_i)} = E[Y|C_i].$$

For the general case, the definition states that $E[Y|\mathcal{G}]$ is essentially the only random variable that uses the information provided by $\mathcal{G}$ and gives the same averages as Y on events that are in $\mathcal{G}$. The existence and uniqueness is provided by the Radon-Nikodym theorem. For Y positive, define the measure

$$\nu(A) = E[Y; A] \qquad \text{for} \qquad A \in \mathcal{G}.$$

Then ν is a measure defined on $\mathcal{G}$ that is absolutely continuous with respect to the underlying probability P restricted to $\mathcal{G}$. Set $E[Y|\mathcal{G}]$ equal to the Radon-Nikodym derivative $d\nu/dP|_{\mathcal{G}}$.

The general case follows by taking positive and negative parts of Y.

For $B \in \mathcal{F}$, the *conditional probability* of B given $\mathcal{G}$ is defined to be $P(B|\mathcal{G}) = E[I_B|\mathcal{G}]$.

Exercise 1.2. *1. Let $C \in \mathcal{F}$, then $P(B|\sigma(C)) = P(B|C)I_C + P(B|C^c)I_{C^c}$.*

2. If $\mathcal{G} = \sigma\{C_1, \cdots, C_n\}$, the σ-algebra generated by a finite partition, then

$$P(B|\mathcal{G}) = \sum_{i=1}^{n} P(B|C_i)I_{C_i}.$$

In other words, if $\omega \in C_i$ then $P(A|\mathcal{G})(\omega) = P(A|C_i)$.

Theorem 1.3 (Bayes Formula)**.** *Let $B \in \mathcal{G}$, then*

$$P(B|A) = \frac{E[P(A|\mathcal{G})I_B]}{E[P(A|\mathcal{G})]}.$$

Proof. $E[P(A|\mathcal{G})I_B] = E[I_A I_B] = P(A \cap B)$ and $E[P(A|\mathcal{G})] = P(A)$. □

Exercise 1.4. *Show the Bayes formula for a finite partition $\{C_1, \cdots, C_n\}$ is*

$$P(C_j|A) = \frac{P(A|C_j)P(C_j)}{\sum_{i=1}^{n} P(A|C_i)P(C_i)}.$$

If $\mathcal{G} = \sigma(X)$, then we usually write $E[Y|\mathcal{G}] = E[Y|X]$. For these circumstances, we have the following theorem which can be proved using the standard machine.

Theorem 1.5. *Let X be a random variable. Then Z is a measurable function on $(\Omega, \sigma(X))$ if and only if there exists a measurable function h on the range of X so that $Z = h(X)$.*

This shows that the deifinition of conditional expectation with respect to a σ-algebra extends the definition of conditional expectation with respect to a random variable.

We now summarize the properties of conditional expectation.

Theorem 1.6. *Let $Y, Y_1, Y_2, \cdots$ have finite absolute mean on $(\Omega, \mathcal{F}, P)$ and let $a_1, a_2 \in \mathbb{R}$. In addition, let $\mathcal{G}$ and $\mathcal{H}$ be σ-algebras contained in $\mathcal{F}$. Then*

1. *If Z is any version of $E[Y|\mathcal{G}]$, then $EZ = EY$. ($E[E[Y|\mathcal{G}]] = EY$).*
2. *If Y is $\mathcal{G}$ measurable, then $E[Y|\mathcal{G}] = Y$, a.s.*
3. **(linearity)** *$E[a_1Y_1 + a_2Y_2|\mathcal{G}] = a_1E[Y_1|\mathcal{G}] + a_2E[Y_2|\mathcal{G}]$, a.s.*
4. **(positivity)** *If $Y \geq 0$, then $E[Y|\mathcal{G}] \geq 0$, a.s.*
5. **(conditional monotone convergence theorem)** *If $Y_n \uparrow Y$, then $E[Y_n|\mathcal{G}] \uparrow E[Y|\mathcal{G}]$, a.s.*
6. **(conditional Fatous's lemma)** *If $Y_n \geq 0$, then $E[\liminf_{n\to\infty} Y_n|\mathcal{G}] \leq \liminf_{n\to\infty} E[Y_n|\mathcal{G}]$.*
7. **(conditional dominated convergence theorem)** *If $\lim_{n\to\infty} Y_n(\omega) = Y(\omega)$, a.s., if $|Y_n(\omega)| \leq V(\omega)$ for all n, and if $EV < \infty$, then*

$$\lim_{n\to\infty} E[Y_n|\mathcal{G}] = E[Y|\mathcal{G}],$$

almost surely.

8. **(conditional Jensen's inequality)** *If $\phi : \mathbb{R} \to \mathbb{R}$ is convex, and $E|\phi(Y)| < \infty$, then*
$$E[\phi(Y)|\mathcal{G}] \geq \phi(E[Y|\mathcal{G}]),$$
almost surely. In particular,
9. **(contraction)** $||E[Y|\mathcal{G}]||_p \leq ||Y||_p$ *for* $p \geq 1$.
10. **(tower property)** *If $\mathcal{H} \subset \mathcal{G}$, then*
$$E[E[Y|\mathcal{G}]|\mathcal{H}] = E[Y|\mathcal{H}],$$
almost surely.
11. **(conditional constants)** *If Z is $\mathcal{G}$-measurable, and $E|ZY| < \infty$, then*
$$E[ZY|\mathcal{G}] = ZE[Y|\mathcal{G}].$$
12. **(role of independence)** *If $\mathcal{H}$ is independent of $\sigma(\sigma(Y), \mathcal{G})$, then*
$$E[Y|\sigma(\mathcal{G}, \mathcal{H})] = E[Y|\mathcal{G}],$$
almost surely. In particular,
13. *if Y is independent of $\mathcal{H}$, then $E[Y|\mathcal{H}] = EY$.*

Proof. 1. Take $A = \Omega$ in the definition.

2. Check that Y satisfies both 1 and 2 in the definition.
3. Clearly $a_1 E[Y_1|\mathcal{G}] + a_2 E[Y_2|\mathcal{G}]$ is $\mathcal{G}$-measurable. Let $A \in \mathcal{G}$. Then
$$\begin{aligned} E[a_1 E[Y_1|\mathcal{G}] + a_2 E[Y_2|\mathcal{G}]; A] &= a_1 E[E[Y_1|\mathcal{G}]; A] + a_2 E[E[Y_2|\mathcal{G}]; A] \\ &= a_1 E[Y_1; A] + a_2 E[Y_2; A] = E[a_1 Y_1 + a_2 Y_2; A], \end{aligned}$$
yielding property 2.
4. Because $Y \geq 0$ and $\{E[Y|\mathcal{G}] \leq -\frac{1}{n}\}]$ is $\mathcal{G}$-measurable, we have that
$$0 \leq E[Y; \{E[Y|\mathcal{G}] \leq -\frac{1}{n}\}] = E[E[Y|\mathcal{G}]; \{E[Y|\mathcal{G}] \leq -\frac{1}{n}\}] \leq -\frac{1}{n} P\{E[Y|\mathcal{G}] \leq -\frac{1}{n}\}$$
and
$$P\{E[Y|\mathcal{G}] \leq -\frac{1}{n}\} = 0.$$
Consequently,
$$P\{E[Y|\mathcal{G}] < 0\} = P\left(\bigcup_{n=1}^{\infty} \{E[Y|\mathcal{G}] \leq -\frac{1}{n}\}\right) = 0.$$

5. Write $Z_n = E[Y_n|\mathcal{G}]$. Then, by positivity of conditional expectation,
$$Z = \lim_{n\to\infty} Z_n$$
exists almost surely. Note that Z is $\mathcal{G}$-measurable. Choose $A \in \mathcal{G}$. By the monotone convergence theorem,
$$E[Z;G] = \lim_{n\to\infty} E[Z_n;G] = \lim_{n\to\infty} E[Y_n;G] = E[Y;G].$$
Thus, $Z = E[Y|\mathcal{G}]$.

Because $-|Y| \leq Y \leq |Y|$, we have $-E[|Y||\mathcal{G}] \leq E[Y|\mathcal{G}] \leq E[|Y||\mathcal{G}]$ and $|E[Y|\mathcal{G}]| \leq E[|Y||\mathcal{G}]$. Consequently,
$$E[\cdot|\mathcal{G}] : L^1(\Omega,\mathcal{F},P) \to L^1(\Omega,\mathcal{G},P)$$
is a continuous linear mapping.

6. Repeat the proof of Fatou's lemma replacing expectation with $E[\cdot|\mathcal{G}]$ and use both positivity and the conditional monotone convergence theorem.

7. Repeat the proof of the dominated convergence theorem from Fatou's lemma again replacing expectation with conditional expectation.

8. Follow the proof of Jensen's inequality.

9. Use the conditional Jensen's inequality on the convex function $\phi(y) = y^p, y \geq 0$,
$$||E[Y|\mathcal{G}]||_p^p = E[|E[Y|\mathcal{G}]|^p] \leq E[E[|Y||\mathcal{G}]^p] \leq E[E[|Y|^p|\mathcal{G}]] = E[|Y|^p] = ||Y||_p^p.$$

10. By definition $E[E[Y|\mathcal{G}]|\mathcal{H}]$ is $\mathcal{H}$-measurable. Let $A \in \mathcal{H}$, then $A \in \mathcal{G}$ and
$$E[E[E[Y|\mathcal{G}]|\mathcal{H}];A] = E[E[Y|\mathcal{G}];A] = E[Y;A].$$

11. Use the standard machine. The case Z an indicator function follows from the definition of conditional expectation.

12. We need only consider the case $Y \geq 0, EY > 0$. Let $Z = E[Y|\mathcal{G}]$ and consider the two probability measures
$$\mu(A) = \frac{E[Y;A]}{EY}, \quad \nu(A) = \frac{E[Z;A]}{EY}.$$
Note that $EY = EZ$ and define
$$\mathcal{C} = \{A : \mu(A) = \nu(A)\}.$$
If $A = B \cap C, B \in \mathcal{G}, C \in \mathcal{H}$, then YI_B and C are independent and
$$E[Y;B\cap C] = E[YI_B;C] = E[YI_B]P(C) = E[Y;B]P(C).$$
Z is $\mathcal{G}$-measurable and thus ZI_B and C are independent and
$$E[Z;B\cap C] = E[ZI_B;C] = E[ZI_B]P(C) = E[Z;B]P(C) = E[Y;B]P(C).$$

Consequently,

$$\mathcal{D} = \{B \cap C; B \in \mathcal{G}, C \in \mathcal{H}\} \subset \mathcal{C}.$$

Now, $\mathcal{D}$ is closed under pairwise intersection. Thus, by the the Sierpinski Class Theorem, μ and ν agree on $\sigma(\mathcal{D}) = \sigma(\mathcal{G}, \mathcal{H})$ and

$$E[Z; A] = E[Y; A], \text{ for all } A \in \sigma(\mathcal{G}, \mathcal{H}).$$

Because Z is $\sigma(\mathcal{G}, \mathcal{H})$-measurable, this completes the proof.

13. Take $\mathcal{G}$ to be the trivial σ-algebra $\{\emptyset, \Omega\}$. However, this property can be verified directly.

□

Exercise 1.7. *Let X be integrable, then the collection*

$$\{E[X|\mathcal{G}]; \mathcal{G} \text{ a sub-}\sigma\text{algebra of } \mathcal{F}\}.$$

is uniformly integrable.

Remark 1.8. *If the sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, then by Jensen's inequality,*

$$E[\cdot|\mathcal{G}] : L^2(\Omega, \mathcal{F}, P) \to L^2(\Omega, \mathcal{G}, P).$$

If $E[Y^2] < \infty$, we can realize the conditional expectation $E[Y|\mathcal{G}]$ as a Hilbert space projection.

Note that $L^2(\Omega, \mathcal{G}, P) \subset L^2(\Omega, \mathcal{F}, P)$. Let $\Pi_{\mathcal{G}}$ be the orthogonal projection operator onto the closed subspace $L^2(\Omega, \mathcal{G}, P)$. Thus $\Pi_{\mathcal{G}} Y$ is $\mathcal{G}$-measurable and $Y - \Pi_{\mathcal{G}} Y$ is orthogonal to any element in $X \in L^2(\Omega, \mathcal{G}, P)$ Consequently,

$$E[(Y - \Pi_{\mathcal{G}} Y)X] = 0, \quad \text{or} \quad E[\Pi_{\mathcal{G}} Y X] = E[YX].$$

take $X = I_A$, $A \in \mathcal{G}$ to see that $\Pi_{\mathcal{G}} Y = E[Y|\mathcal{G}]$.

As before, this can be viewed as a minimization problem

$$\min\{E[(Y - Z)^2]; Z \in L^2(\Omega, \mathcal{G}, P)\}.$$

The minimum is acheived uniquely by the choice $Z = E[Y|\mathcal{G}]$.

Exercise 1.9.

1. *Prove the conditional Chebyshev inequality: Let $g : \mathbb{R} \to [0, \infty)$ be increasing on $[a, \infty)$, then*

$$P\{g(X) > a|\mathcal{G}\} \leq \frac{E[g(X)|\mathcal{G}]}{a}.$$

2. *Let $\mathcal{G}$ be a sub-σ-algebra and let $EY^2 < \infty$, then*

$$Var(Y) = E[Var(Y|\mathcal{G})] + Var(E[Y|\mathcal{G}]).$$

where $Var(Y|\mathcal{G})] = E[(Y - E[Y|\mathcal{G}])^2|\mathcal{G}]$.

Example 1.10. *1. Let S_j be a binomial random variable, the number of heads in j flips of a biased coin with the probability of heads equal to p. Thus, there exist independent 0-1 valued random variables $\{X_j : j \geq 1\}$ so that $S_n = X_1 + \cdots + X_n$. Let $k < n$, then $S_k = X_1 + \cdots + X_k$ and $S_n - S_k = X_{k+1} + \cdots + X_n$ are independent. Therefore,*

$$E[S_n|S_k] = E[(S_n - S_k) + S_k|S_k] = E[(S_n - S_k)|S_k] + E[S_k|S_k] = E[S_n - S_k] + S_k = (n-k)p + S_k.$$

Now to find $E[S_k|S_n]$, note that

$$x = E[S_n|S_n = x] = E[X_1|S_n = x] + \cdots + E[X_n|S_n = x].$$

Each of these summands has the same value, namely x/n. Therefore

$$E[S_k|S_n = x] = E[X_1|S_n = x] + \cdots + E[X_k|S_n = x] = \frac{k}{n}x$$

and

$$E[S_k|S_n] = \frac{k}{n}S_n.$$

2. Let $\{X_n : n \geq 1\}$ be independent random variables having the same distribution. Let μ be their common mean and σ^2 their common variance. Let N be a non-negative valued random variable and define $S = \sum_{n=1}^{N} X_n$, then

$$ES = E[E[S|N]] = E[N\mu] = \mu EN.$$

$$Var(S) = E[Var(S|N)] + Var(E[S|N]) = E[N\sigma^2] + Var(N\mu) = \sigma^2 EN + \mu^2 Var(N).$$

Alternatively, for an $\mathbb{N}$-valued random variable X, define the generating function

$$G_X(z) = Ez^X = \sum_{x=0}^{\infty} z^x P\{X = x\}.$$

Then

$$G_X(1) = 1, \quad G'_X(1) = EX, \quad G''_X(1) = E[X(X-1)], \quad Var(X) = G''_X(1) + G'_X(1) - G'_X(1)^2.$$

Now,

$$\begin{aligned} G_S(z) &= Ez^S = E[E[z^S|N]] = \sum_{n=0}^{\infty} E[z^S|N = n]P\{N = n\} = \sum_{n=0}^{\infty} E[z^{(X_1+\cdots+X_N)}|N = n]P\{N = n\} \\ &= \sum_{n=0}^{\infty} E[z^{(X_1+\cdots+X_n)}]P\{N = n\} = \sum_{n=0}^{\infty} \left(E[z^{X_1}] \times \cdots \times E[z^{X_n}]\right) P\{N = n\} \\ &= \sum_{n=0}^{\infty} G_X(z)^n P\{N = n\} = G_N(G_X(z)) \end{aligned}$$

Thus,

$$G'_S(z) = G'_N(G_X(z))G'_X(z), \quad ES = G'_S(1) = G'_N(1)G'_X(1) = \mu EN$$

and

$$G''_S(z) = G''_N(G_X(z))G'_X(z)^2 + G'_N(G_X(z))G''_X(z)$$
$$G''_S(1) = G''_N(1)G'_X(1)^2 + G'_N(1)G''_X(1) = G''_N(1)\mu^2 + EN(\sigma^2 - \mu + \mu^2).$$

Now, substitute to obtain the formula above.

3. *The density with respect to Lebesgue measure for a bivariate standard normal is*

$$f_{(X,Y)}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right).$$

To check that $E[Y|X] = \rho X$, we must check that for any Borel set B

$$E[YI_B(X)] = E[\rho X I_B(X)] \quad or \quad E[(Y-\rho X)I_B(X)] = 0.$$

Complete the square to see that

$$\begin{aligned} E[(Y-\rho X)I_B(X)] &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_B \left(\int (y-\rho x)\exp\left(-\frac{(y-\rho x)^2}{2(1-\rho^2)}\right) dy\right) e^{-x^2/2}\ dx \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_B \left(\int z \exp\left(-\frac{z^2}{2(1-\rho^2)}\right) dz\right) e^{-x^2/2} dx = 0,\ z = y - \rho x \end{aligned}$$

because the integrand for the z integration is an odd function. Now

$$Cov(X,Y) = E[XY] = E[E[XY|X]] = E[XE[Y|X]] = E[\rho X^2] = \rho.$$

Definition 1.11. *On a probability space $(\Omega, \mathcal{F}, P)$, let $\mathcal{G}_1, \mathcal{G}_2, \mathcal{H}$ be sub-σ-algebras of $\mathcal{F}$. Then the σ-algebras $\mathcal{G}_1$ and $\mathcal{G}_2$ are* conditionally independent *given $\mathcal{H}$ if*

$$P(A_1 \cap A_2|\mathcal{H}) = P(A_1|\mathcal{H})P(A_2|\mathcal{H}),$$

for $A_i \in \mathcal{G}_i$, $i = 1, 2$.

Exercise 1.12 (conditional Borel-Cantelli lemma)**.** *For $\mathcal{G}$ a sub-σ-algebra of $\mathcal{F}$, let $\{A_n : n \geq 1\}$ be a sequence of conditionally independent events given $\mathcal{G}$. Show that for almost all $\omega \in \Omega$,*

$$P\{A_n\ i.o.|\mathcal{G}\}(\omega) = \left\{ \begin{array}{l} 0 \\ 1 \end{array} \right.$$

according as

$$\sum_{n=1}^{\infty} P(A_n|\mathcal{G})(\omega) \left\{ \begin{array}{l} = \infty \\ < \infty, \end{array} \right.$$

Consequently,

$$P\{A_n\ i.o.\} = P\{\omega : \sum_{n=1}^{\infty} P(A_n|\mathcal{G})(\omega) = \infty\}.$$

Proposition 1.13. *Let $\mathcal{H}, \mathcal{G}$ and $\mathcal{D}$ be σ-algebras. Then*

1. *If $\mathcal{H}$ and $\mathcal{G}$ are conditionally independent given $\mathcal{D}$, then $\mathcal{H}$ and $\sigma(\mathcal{G}, \mathcal{D})$ are conditionally independent given $\mathcal{D}$.*
2. *Let $\mathcal{H}_1$ and $\mathcal{G}_1$ be sub-σ-algebras of $\mathcal{H}$ and $\mathcal{G}$, respectively. Suppose that $\mathcal{H}$ and $\mathcal{G}$ are independent. Then $\mathcal{H}$ and $\mathcal{G}$ are conditionally independent given $\sigma(\mathcal{H}_1, \mathcal{G}_1)$.*
3. *Let $\mathcal{H} \subset \mathcal{G}$. Then $\mathcal{G}$ and $\mathcal{D}$ are conditionally independent given $\mathcal{H}$ if and only if, for every $D \in \mathcal{D}$, $P(D|\mathcal{H}) = P(D|\mathcal{G})$.*

Exercise 1.14. *Prove the previous proposition. The Sierpinski class theorem will be helpful in the proof.*

1.2 Stochastic Processes, Filtrations and Stopping Times

A *stochastic process* X (or a *random process*, or simply a process) with index set Λ and a measurable state space $(S, \mathcal{B})$ defined on a probability space $(\Omega, \mathcal{F}, P)$ is a function

$$X : \Lambda \times \Omega \to S$$

such that for each $\lambda \in \Lambda$,

$$X(\lambda, \cdot) : \Omega \to S$$

is an S-valued random variable.

Note that Λ is not given the structure of a measure space. In particular, it is not necessarily the case that X is measurable. However, if Λ is countable and has the power set as its σ-algebra, then X is automatically measurable.

$X(\lambda, \cdot)$ is variously written $X(\lambda)$ or X_λ. Throughout, we shall assume that S is a metric space with metric d.

A *realization* of X or a *sample path* for X is the function

$$X(\cdot, \omega_0) : \Lambda \to S \quad \text{for some } \omega_0 \in \Omega.$$

Typically, for the processes we study Λ will be the natural numbers, and $[0, \infty)$. Occasionally, Λ will be the integers or the real numbers. In the case that Λ is a subset of a multi-dimensional vector space, we often call X a *random field*.

The distribution of a process X is generally stated via its finite dimensional distributions

$$\mu_{(\lambda_1, \ldots, \lambda_n)}(A) = P\{(X_{\lambda_1}, \ldots, X_{\lambda_n}) \in A\}.$$

If this collection probability measures satisfy the consistency consition of the Daniell-Kolmogorov extension theorem, then they determine the distribution of the process X on the space S^Λ.

We will consider the case $\Lambda = \mathbb{N}$.

Definition 1.15.

1. *A collection of σ-algebras $\{\mathcal{F}_n : n \geq 0\}$ each contained in $\mathcal{F}$ is a* filtration *if $\mathcal{F}_n \subset \mathcal{F}_{n+1}$, $n = 0, 1, \ldots$. $\mathcal{F}_n$ is meant to capture the information available to an observer at time n.*
2. *The* natural filtration *for the process X is $\mathcal{F}_n^X = \sigma\{X_k; k \leq n\}$.*
3. *This filtration is* complete *if $\mathcal{F}$ is complete and $\{A : P(A) = 0\} \subset \mathcal{F}_0$.*
4. *A process X is* adapted *to a filtration $\{\mathcal{F}_n : n \geq 0\}$ (X is $\mathcal{F}_n$-adapted.) if X_n is $\mathcal{F}_n$ measurable. In other words, X is $\mathcal{F}_n$-adapted if and only if for every $n \in \mathbb{N}$, $\mathcal{F}_n^X \subset \mathcal{F}_n$*
5. *If two processes X and $\tilde{X}$ have the same finite dimensional distributions, then X is a* version *of $\tilde{X}$.*
6. *If, for each n, $P\{X_n = \tilde{X}_n\} = 1$, then we say that X is a* modification *of $\tilde{X}$. In this case $P\{X = \tilde{X}\} = 1$. This conclusion does not necessarily hold if the index set is infinite but not countably infinite.*
7. *A non-negative integer-valued random variable τ is an $\mathcal{F}_n$-*stopping time *if $\{\tau \leq n\} \in \mathcal{F}_n$ for every $n \in \mathbb{N}$.*

Exercise 1.16. *τ is an $\mathcal{F}_n$-*stopping time *if and only if $\{\tau = n\} \in \mathcal{F}_n$ for every $n \in \mathbb{N}$.*

Exercise 1.17. *Let X be $\mathcal{F}_n$-adapted and let τ be an $\mathcal{F}_n$-stopping time. For $B \in \mathcal{B}$, define*

$$\sigma(B,\tau) = \inf\{n \geq \tau : X_n \in B\} \text{ and } \sigma^+(B,\tau) = \inf\{n > \tau : X_n \in B\}.$$

Then $\sigma(B,\tau)$ and $\sigma^+(B,\tau)$ are $\mathcal{F}_n$-stopping times.

Proposition 1.18. *Let $\tau_1, \tau_2, \ldots,$ be $\mathcal{F}_n$-stopping times and let $c \geq 0$. Then*

1. *$\tau_1 + c$ and $\min\{\tau_1, c\}$ are $\mathcal{F}_n$-stopping times.*
2. *$\sup_k \tau_k$ is an $\mathcal{F}_n$-stopping time.*
3. *$\inf_k \tau_k$ is an $\mathcal{F}_n$-stopping time.*
4. *$\liminf_{k\to\infty} \tau_k$ and $\limsup_{k\to\infty} \tau_k$ are $\mathcal{F}_n$-stopping times.*

Proof.

1. $\{\tau_1 + c \leq n\} = \{\tau_1 \leq n - c\} \in \mathcal{F}_{\max\{0,n-c\}} \subset \mathcal{F}_n$.
   If $c \leq n$, $\{\min\{\tau_1, c\} \leq n\} = \Omega \in \mathcal{F}_n$, If $c > n$, $\{\min\{\tau_1, c\} \leq n\} = \{\tau_1 \leq n\} \in \mathcal{F}_n$,
2. $\{\sup_k \tau_k \leq n\} = \bigcap_k \{\tau_k \leq n\} \in \mathcal{F}_n$.
3. $\{\inf_k \tau_k \leq n\} = \bigcup_k \{\tau_k \leq n\} \in \mathcal{F}_n$. (This statement does not hold if τ_k is a continuous random variable.)
4. This follows from parts 2 and 3.

□

Definition 1.19.

1. *Let τ be an $\mathcal{F}_n$-stopping time. Then*

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau \leq n\} \in \mathcal{F}_n \text{ for all } n\}.$$

2. *Let X be a random process then X^τ, the* process X stopped at time τ, *is defined by*

$$X_n^\tau = X_{\min\{\tau,n\}}$$

The following exercise explains the intuitive idea that $\mathcal{F}_\tau$ gives the information known to observer up to time τ.

Exercise 1.20.

1. *$\mathcal{F}_\tau$ is a σ-algebra.*
2. *$\mathcal{F}_\tau^X = \sigma\{X_n^\tau : n \geq 0\}$.*

Proposition 1.21. *Let σ and τ be $\mathcal{F}_n$-stopping times and let X be an $\mathcal{F}_n$-adapted process.*

1. *τ and $\min\{\tau, \sigma\}$ are $\mathcal{F}_\tau$-measurable.*
2. *If $\tau \leq \sigma$, then $\mathcal{F}_\tau \subset \mathcal{F}_\sigma$.*
3. *X_τ is $\mathcal{F}_\tau$-measurable.*

Proof.

1. $\{\min\{\tau, \sigma\} \leq c\} \cap \{\tau \leq n\} = \{\min\{\tau, \sigma\} \leq \min\{c, n\}\} \cap \{\tau \leq n\} = (\{\tau \leq \min\{c, n\}\} \cup \{\sigma \leq \min\{c, n\}\}) \cap \{\tau \leq n\} \in \mathcal{F}_n$.

2. Let $A \in \mathcal{F}_\tau$, then
$$A \cap \{\sigma \leq n\} = (A \cap \{\tau \leq n\}) \cap \{\sigma \leq n\} \in \mathcal{F}_n.$$
Hence, $A \in \mathcal{F}_\sigma$.

3. Let $B \in \mathcal{B}$, then
$$\{X_\tau \in B\} \cap \{\tau \leq n\} = \cup_{k=1}^n (\{X_k \in B\} \cap \{\tau = k\}) \in \mathcal{F}_n.$$

□

Exercise 1.22. *Let X be an $\mathcal{F}_n$ adapted process. For $n = 0, 1, \cdots,$ define $\mathcal{F}_n^\tau = \mathcal{F}_{\min\{\tau,n\}}$. Then*

1. *$\{\mathcal{F}_n^\tau : n \geq 0\}$ is a filtration.*
2. *X^τ is both $\mathcal{F}_n^\tau$-adapted and $\mathcal{F}_n$-adapted.*

1.3 Coupling

Definition 1.23. *Let ν and $\tilde{\nu}$ be probability distributions (*of a random variable or of a stochastic process*). Then a* coupling *of ν and $\tilde{\nu}$ is a pair of random variable X and $\tilde{X}$ defined on a probability space $(\Omega, \mathcal{F}, P)$ such that the marginal distribution of X is ν and the marginal distribution of $\tilde{X}$ is $\tilde{\nu}$*

As we shall soon see, coupling are useful in comparing two distribution by constructing a coupling and comparing random variables.

Definition 1.24. *The* total variation distance *between two probability measures ν and $\tilde{\nu}$ on a measurable space $(S, \mathcal{B})$ is*
$$||\nu - \tilde{\nu}||_{TV} = \sup\{|\nu(B) - \tilde{\nu}(B)|; B \in \mathcal{B}\}. \tag{1.2}$$

Example 1.25. *Let ν be a $Ber(p)$ and let $\tilde{\nu}$ be a $Ber(\tilde{p})$ distribution, $\tilde{p} \geq p$. If X and $\tilde{X}$ are a coupling of independent random variables, then*
$$P\{X \neq \tilde{X}\} = P\{X = 0, \tilde{X} = 1\} + P\{X = 1, \tilde{X} = 0\} = (1-p)\tilde{p} + p(1-\tilde{p}) = \tilde{p} + p - 2p\tilde{p}.$$
Now couple ν and $\tilde{\nu}$ as follows. Let U be a $U(0,1)$ random variable. Set $X = 1$ if and only if $U \leq p$ and $\tilde{X} = 1$ if and only if $U \leq \tilde{p}$. Then
$$P\{X \neq \tilde{X}\} = P\{p < U \leq \tilde{p}\} = \tilde{p} - p.$$
Note that $||\nu - \tilde{\nu}||_{TV} = \tilde{p} - p$.

We now turn to the relationship between coupling and the total variation distance.

Proposition 1.26. *If S is a countable set and $\mathcal{B}$ is the power set then*
$$||\nu - \tilde{\nu}||_{TV} = \frac{1}{2} \sum_{x \in S} |\nu\{x\} - \tilde{\nu}\{x\}|. \tag{1.3}$$

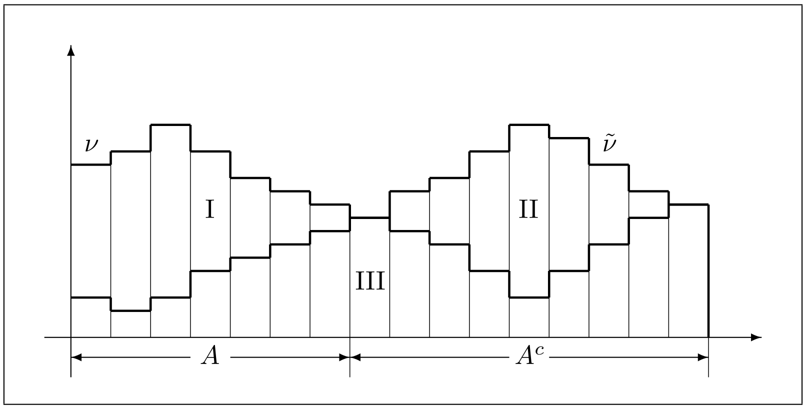

Figure 1: Distribution of ν and $\tilde{\nu}$. The event $A = \{x; \nu\{x\} > \tilde{\nu}\{x\}\}$. Regions I and II have the same area, namely $||\nu - \tilde{\nu}||_{TV} = \nu(A) - \tilde{\nu}(A) = \tilde{\nu}(A^c) - \nu(A^c)$. Thus, the total area of regions I and III and of regions II and III is 1.

Proof. (This is one half the area of region I plus region II in Figure 1.) Let $A = \{x; \nu\{x\} > \tilde{\nu}\{x\}\}$ and let $B \in \mathcal{B}$. Then

$$\nu(B) - \tilde{\nu}(B) = \sum_{x \in B}(\nu\{x\} - \tilde{\nu}\{x\}) \leq \sum_{x \in A \cap B}(\nu\{x\} - \tilde{\nu}\{x\}) = \nu(A \cap B) - \tilde{\nu}(A \cap B)$$

because the terms eliminated to create the second sum are negative. Similarly,

$$\nu(A \cap B) - \tilde{\nu}(A \cap B) = \sum_{x \in A \cap B}(\nu\{x\} - \tilde{\nu}\{x\}) \leq \sum_{x \in A}(\nu\{x\} - \tilde{\nu}\{x\}) = \nu(A) - \tilde{\nu}(A)$$

because the additional terms in the second sum are positive. Repeat this reasoning to obtain

$$\tilde{\nu}(B) - \nu(B) \leq \tilde{\nu}(A^c) - \nu(A^c).$$

However, because ν and $\tilde{\nu}$ are probability measures,

$$\tilde{\nu}(A^c) - \nu(A^c) = (1 - \tilde{\nu}(A)) - (1 - \nu(A)) = \nu(A) - \tilde{\nu}(A).$$

Consequently, the supremum in (1.2) is obtained in choosing either the event A or the event A^c and thus is equal to half the sum of the two as given in (1.3). □

Exercise 1.27. *Let* $||f||_\infty = \sup_{x \in S} |f(x)|$. *Show that*

$$||\nu - \tilde{\nu}||_{TV} = \frac{1}{2} \sup\{\sum_{x \in S} f(x)(\nu\{x\} - \tilde{\nu}\{x\}); ||f||_\infty = 1\}.$$

Theorem 1.28. *If S is a countable set and $\mathcal{A}$ is the power set, then*

$$||\nu - \tilde{\nu}||_{TV} = \inf\{P\{X \neq \tilde{X}\}; (P, X, \tilde{X}) \in \mathcal{C}(\nu, \tilde{\nu})\} \tag{1.4}$$

where the infimum is taken over all possible couplings $\mathcal{C}(\nu, \tilde{\nu})$ of ν and $\tilde{\nu}$.

Proof. For any coupling $(P, X, \tilde{X})$,

$$P\{X \in B\} = P\{X \in B, \tilde{X} \in B\} + P\{X \in B, \tilde{X} \in B^c\} \geq P\{\tilde{X} \in B\} + P\{X \in B, \tilde{X} \in B^c\}$$

Thus,

$$P\{X \neq \tilde{X}\} \geq P\{X \in B, \tilde{X} \in B^c\} \geq P\{X \in B\} - P\{\tilde{X} \in B\} = \nu(B) - \tilde{\nu}(B).$$

Thus, $P\{X \neq \tilde{X}\}$ is at least big as the total variation distance.

Next, we construct a coupling that achieve the infimum. First, set

$$p = \sum_{x \in S} \min\{\nu\{x\}, \tilde{\nu}\{x\}\} = \sum_{x \in S} \frac{1}{2}(\nu\{x\} + \tilde{\nu}\{x\} - |\nu\{x\} - \tilde{\nu}\{x\}|) = 1 - ||\nu - \tilde{\nu}||_{TV},$$

the area of region III in Figure 1.

If $p = 0$, then ν and $\tilde{\nu}$ takes values on disjoint set and the total variation distance between the two is 1. For $p > 0$, define a random variable Y, $P\{Y = x\} = \min\{\nu\{x\}, \tilde{\nu}\{x\}\}/p$. In addition, define the probability distributions,

$$\mu\{x\} = \frac{\nu\{x\} - \tilde{\nu}\{x\}}{1-p} I_{\{\nu\{x\} > \tilde{\nu}\{x\}\}}, \quad \text{and} \quad \tilde{\mu}\{\tilde{x}\} = \frac{\tilde{\nu}\{\tilde{x}\} - \nu\{\tilde{x}\}}{1-p} I_{\{\tilde{\nu}\{\tilde{x}\} > \nu\{\tilde{x}\}\}}.$$

Now, flip a coin that lands heads with probability p.

If the coin lands heads, set $X = \tilde{X} = Y$.

If the coin lands tails, then let X and $\tilde{X}$ be independent, X has distribution μ and $\tilde{X}$ has distribution $\tilde{\mu}$.

Checking, we see that

$$\begin{aligned} P\{X = x\} &= P\{X = x|\text{coin lands heads}\}P\{\text{coin lands heads}\} + P\{X = x|\text{coin lands tails}\}P\{\text{coin lands tails}\} \\ &= \frac{\min\{\nu\{x\}, \tilde{\nu}\{x\}\}}{p} p + \frac{\nu\{x\} - \tilde{\nu}\{x\}}{1-p} I_{\{\nu\{x\} > \tilde{\nu}\{x\}\}}(1-p) = \nu\{x\} \end{aligned}$$

Similarly, $P\{\tilde{X} = \tilde{x}\} = \tilde{\nu}\{\tilde{x}\}$.

Finally,

$$\begin{aligned} P\{X = \tilde{X}\} &= P\{X = \tilde{X}|\text{coin lands heads}\}P\{\text{coin lands heads}\} + P\{X = \tilde{X}|\text{coin lands tails}\}P\{\text{coin lands tails}\} \\ &= p + \sum_{x \in S} P\{X = x, \tilde{X} = x|\text{coin lands tails}\}(1-p) \\ &= p + \sum_{x \in S} P\{X = x|\text{coin lands tails}\}P\{\tilde{X} = x|\text{coin lands tails}\}(1-p) \\ &= p + \sum_{x \in S} \mu\{x\}\tilde{\mu}\{x\}(1-p) = p. \end{aligned}$$

Note that each term in sum is zero. Either the first term, the second term or both terms are zero based on the sign of $\nu\{x\} - \tilde{\nu}\{x\}$. Thus,

$$P\{X \neq \tilde{X}\} = 1 - p = ||\nu - \tilde{\nu}||_{TV}.$$

□

We will typically work to find successful couplings of stochastic processes. They are defined as follows:

Definition 1.29. *A coupling* $(P, X, \tilde{X})$ *is called a* successful coupling *of two adapted stochastic processes, if the stopping time*

$$\tau = \inf\{t \geq 0; X_t = \tilde{X}_t\}$$

is almost surely finite and if $X_t = \tilde{X}_t$ *for all* $t \geq \tau$. *Then, we say that* τ *is the* coupling time *and that* X and $\tilde{X}$ are coupled at time τ .

If ν_t and $\tilde{\nu}_t$ are the distributions of the processes at time t, then $(P, X_t, \tilde{X}_t)$ is a coupling of ν_t and $\tilde{\nu}_t$ and, for a successful coupling,

$$||\nu_t^X - \nu_t^Y||_{TV} \leq P\{X_t \neq Y_t\} = P\{\tau > t\}.$$

Thus, the earlier the coupling time, the better the estimate of the total variation distance of the time t distributions.

2 Random Walk

Definition 2.1. *Let X be a sequence of independent and identically distributed $\mathbb{R}^d$-valued random variables and define*

$$S_0 = 0, \quad S_n = \sum_{k=1}^{n} X_k.$$

Then the sequence S is called a random walk *with steps $X_1, X_2, \ldots$.*

If the steps take on only the values $\pm \mathbf{e}_j$, the standard unit basis vectors in $\mathbb{R}^d$, then the random sequence S is called a simple random walk. *If each of these $2d$ steps are equally probable, the S is called a* symmetric simple random walk.

Theorem 2.2. *Let $X_1, X_2, \ldots$ be independent with distribution ν and let τ be a stoppping time. Then, conditioned on $\{\tau < \infty\}$, $\{X_{\tau+n}, n \geq 1\}$ is independent of $\mathcal{F}_\tau^X$ and has the same distribution as the original sequence.*

Proof. Choose $A \in \mathcal{F}_\tau$ and $B_j \in \mathcal{B}(\mathbb{R}^d)$, then because the finite dimensional distributions determine the process, it suffices to show that

$$P(A \cap \{\tau < \infty, X_{\tau+j} \in B_j, 1 \leq j \leq k\}) = P(A \cap \{\tau < \infty\}) \prod_{1 \leq j \leq k} \nu(B_k).$$

To do this note that $A \cap \{\tau = n\} \in \mathcal{F}_n^X$ and thus is independent of $X_{n+1}, X_{n+2}, \ldots$. Therefore,

$$\begin{aligned} P(A \cap \{\tau = n, X_{\tau+j} \in B_j, 1 \leq j \leq k\}) &= P(A \cap \{\tau = n, X_{n+j} \in B_j, 1 \leq j \leq k\}) \\ &= P(A \cap \{\tau = n\}) P\{X_{n+j} \in B_j, 1 \leq j \leq k\}) \\ &= P(A \cap \{\tau = n\}) \prod_{1 \leq j \leq k} \nu(B_j). \end{aligned}$$

Now, sum over n and note that each of the events are disjoint for differing values of n, □

Exercise 2.3.

1. *Assume that the steps of a real valued random walk S have finite mean and let τ be a stopping with finite mean. Show that*
$$ES_\tau = \mu E\tau.$$
2. *In the exercise, assume that $\mu = 0$. For $c > 0$ and $\tau = \min\{n \geq 0; S_n \geq c\}$, show that $E\tau$ is infinite.*

2.1 Recurrence and Transience

The *strong law of large numbers* states that if the step distribution has finite mean μ, then

$$P\{\lim_{n \to \infty} \frac{1}{n} S_n = \mu\} = 1.$$

So, if the mean exists and is not zero, the walk will drift in the direction of the mean.

If the mean does not exist or if it is zero, we have the possibility that the walk visits sites repeatedly. Noting that the walk could only return to possible sites motivates the following definitions.

Definition 2.4. *Let S be a random walk. A point $x \in \mathbb{R}^d$ is called a* recurrent state *if for every neighborhood U of x,*

$$P\{S_n \in U \text{ i. o.}\} = 1.$$

Definition 2.5. *If, for some random real valued random variable X and some $\ell > 0$,*

$$\sum_{n=-\infty}^{\infty} P\{X = n\ell\} = 1,$$

then X is said to be distributed on the lattice $L_\ell = \{n\ell : n \in \mathbb{Z}\}$ *provided that the equation hold for no larger ℓ. If this holds for no $\ell > 0$, then X is called* non-lattice *and is said to be distributed on $L_0 = \mathbb{R}$.*

Exercise 2.6. *Let the steps of a random walk be distributed on a lattice L_ℓ, $\ell > 0$. Let τ_n be the n-th return of the walk to zero. Then*

$$P\{\tau_n < \infty\} = P\{\tau_1 < \infty\}^n.$$

Theorem 2.7. *If the step distribution is on L_ℓ, then either every state in L_ℓ is recurrent or no states are recurrent.*

Proof. Let G be the set of recurrent points.

Claim I. The set G is closed.

Choose a sequence $\{x_k : k \geq 0\} \subset G$ with limit x and choose U a neighborhood of x. Then, there exists an integer K so that $x_k \in U$ whenever $k \geq K$ and hence U is a neighborhood of x_k. Because $x_k \in G$, $P\{S_n \in U \text{ i. o.}\} = 1$ and $x \in G$.

Call y be *a possible state* ($y \in C$) if for every neighborhood U of y, there exists k so that $P\{S_k \in U\} > 0$.

Claim II. If $x \in G$ and $y \in C$, then $x - y \in G$.

Let $U = (y - \epsilon, y + \epsilon)$ and pick k so that $P\{|S_k - y| < \epsilon\} > 0$. Use f. o., finitely often, to indicate the complement of i. o., infinitely often. Then

$$P\{|S_n - x| < \epsilon \text{ f. o.}\} \geq P\{|S_{k+n} - x| < \epsilon \text{ f. o.}, |S_k - y| < \epsilon\}.$$

If S_k is within ϵ of y and S_{k+n} is within ϵ of x, then $S_{k+n} - S_k$ is within 2ϵ of $x - y$. Therefore, for each k,

$$\{|S_{k+n} - x| < \epsilon, |S_k - y| < \epsilon\} \subset \{|(S_{k+n} - S_k) - (x - y)| < 2\epsilon, |S_k - y| < \epsilon\}$$

and

$$\{|S_{k+n} - x| < \epsilon \text{ f. o.}, |S_k - y| < \epsilon\} \supset \{|(S_{k+n} - S_k) - (x - y)| < 2\epsilon \text{ f. o.}, |S_k - y| < \epsilon\}.$$

Use the fact that S_k and $S_{k+n} - S_k$ are independent and that the distribution of S_n and $S_{k+n} - S_k$ are equal.

$$\begin{aligned} P\{|S_n - x| < \epsilon \text{ f. o.}\} &\geq P\{|(S_{n+k} - S_k) - (x - y)| < 2\epsilon \text{ f. o.}, |S_k - y| < \epsilon\} \\ &= P\{|(S_{n+k} - S_k) - (x - y)| < 2\epsilon \text{ f. o.}\}P\{|S_k - y| < \epsilon\} \\ &= P\{|S_n - (x - y)| < 2\epsilon \text{ f. o.}\}P\{|S_k - y| < \epsilon\}. \end{aligned}$$

Because $x \in G$, $0 = P\{|S_n - x| < \epsilon \text{ f. o.}\}$. Recalling that $P\{|S_k - y| < \epsilon\} > 0$, this forces

$$P\{|S_n - (x - y)| < 2\epsilon \text{ f. o.}\} = 0,$$

$$P\{|S_n - (x-y)| < 2\epsilon \text{ i. o.}\} = 1.$$

Finally, choose U, a neighborhood of $x - y$ and ϵ so that $(x - y - 2\epsilon, x - y + 2\epsilon) \in U$, then

$$P\{S_n \in U \text{ i. o.}\} \geq P\{|S_n - (x-y)| < 2\epsilon \text{ i. o.}\} = 1,$$

and $x - y \in G$.

Claim III. G is a group.

If $x \in G$ and $y \in G$, then $y \in C$ and $x - y \in G$.

Claim IV. If $G \neq \emptyset$, and if $\ell > 0$, then $G = L_\ell$.

Clearly $L_\ell \subset G$.

Because $G \neq \emptyset$, $0 \in G$. Set $S_X = \{x : P\{X_1 = x\} > 0\}$. If $x \in S_X$, then $x \in C$ and $-x = 0 - x \in G$ and $x \in G$. By the definition of ℓ, all integer multiples of ℓ can be written as positive linear sums from $x, -x$, $x \in S_X$ and $G \subset L_\ell$

Claim V. If $\ell = 0$, and $G \neq \emptyset$, then $G = L_0$.

$0 = \inf\{y > 0 : y \in G\}$. Otherwise, we are in the situation of a lattice valued random walk.

Let $x \in \mathbb{R}$ and let U be a neighborhood of x. Now, choose $\epsilon > 0$ so that $(x - \epsilon, x + \epsilon) \subset U$. By the definition of infimum, there exists $\tilde{\ell} < \epsilon$ so that $\tilde{\ell} \in G$. Thus, $L_{\tilde{\ell}} \subset G$ and $L_{\tilde{\ell}} \cap U \neq \emptyset$. Consequently,

$$P\{S_n \in U \text{ i. o.}\} = 1,$$

and $x \in G$. □

Corollary 2.8. *Let S be a random walk in $\mathbb{R}^d$. Then, either the set of recurrent states forms a closed subgroup of $\mathbb{R}^d$ or no states are recurrent. This subgroup is the closed subgroup generated by the step distribution.*

Proof. For claims I, II, and III, let $|\cdot|$ stand for some norm in $\mathbb{R}^d$ and replace open intervals by open balls in that norm. □

Theorem 2.9. *Let S be a random walk in $\mathbb{R}^d$.*

1. *If there exists an open set U containing 0 so that*

$$\sum_{n=0}^{\infty} P\{S_n \in U\} < \infty,$$

then no states are recurrent.

2. *If there is a bounded open set U such that*

$$\sum_{n=0}^{\infty} P\{S_n \in U\} = \infty,$$

then 0 is recurrent.

Proof. If $\sum_{n=0}^{\infty} P\{S_n \in U\} < \infty$, then by the Borel-Cantelli, $P\{S_n \in U \text{ i. o.}\} = 0$. Thus 0 is not recurrent. Because the set of recurrent states is a group, then no states are recurrent.

For the second half, let $\epsilon > 0$ and choose a finite set F so that

$$U \subset \bigcup_{x \in F} B(x, \epsilon).$$

Then,

$$\{S_n \in U\} \subset \bigcup_{x \in F} \{S_n \in B(x, \epsilon)\}, \quad P\{S_n \in U\} \leq \sum_{x \in F} P\{S_n \in B(x, \epsilon)\}.$$

Consequently, for some x,

$$\sum_{n=0}^{\infty} P\{S_n \in B(x, \epsilon)\} = \infty.$$

Define the pairwise disjoint sets, A_k, the last visit to $B(x, \epsilon)$ occurred for S_k.

$$A_k = \begin{cases} \{S_n \notin B(x, \epsilon), n = 1, 2, \ldots\} & \text{for} \quad k = 0 \\ \{S_k \in B(x, \epsilon), S_{n+k} \notin B(x, \epsilon), n = 1, 2, \ldots\} & \text{for} \quad k > 0 \end{cases}$$

Then

$$\{S_n \in B(x, \epsilon) \text{ f. o.}\} = \bigcup_{k=0}^{\infty} A_k, \quad \text{and} \quad P\{S_n \in B(x, \epsilon) \text{ f. o.}\} = \sum_{k=0}^{\infty} P(A_k).$$

For $k > 0$, if $S_k \in B(x, \epsilon)$ and $|S_{n+k} - S_k| > 2\epsilon$, then $S_{n+k} \notin B(x, \epsilon)$. Consequently,

$$\begin{aligned} P(A_k) &\geq P\{S_k \in B(x, \epsilon), |S_{n+k} - S_k| > 2\epsilon, n = 1, 2, \ldots\} \\ &= P\{S_k \in B(x, \epsilon)\} P\{|S_{n+k} - S_k| > 2\epsilon, n = 1, 2, \ldots\} \\ &= P\{S_k \in B(x, \epsilon)\} P\{|S_n| > 2\epsilon, n = 1, 2, \ldots\} \end{aligned}$$

Summing over k, we find that

$$P\{S_n \in B(x, \epsilon) \text{ f. o.}\} \geq \left(\sum_{k=0}^{\infty} P\{S_k \in B(x, \epsilon)\} \right) P\{|S_n| > 2\epsilon, n = 1, 2, \ldots\}.$$

The number on the left is at most one, the sum is infinite and consequently for any $\epsilon > 0$,

$$P\{|S_n| > 2\epsilon, n = 1, 2, \ldots\} = 0.$$

Now, define the sets A_k as before with $x = 0$. By the argument above $P(A_0) = 0$. For $k \geq 1$, define a strictly increasing sequence $\{\epsilon_i : i \geq 1\}$ with limit ϵ. By the continuity property of a probability, we have,

$$P(A_k) = \lim_{i \to \infty} P\{S_k \in B(0, \epsilon_i), S_{n+k} \notin B(0, \epsilon), n = 1, 2, \ldots\}.$$

Pick a term in this sequence, then

$$\begin{aligned} & P\{S_k \in B(0, \epsilon_i), S_{n+k} \notin B(0, \epsilon), n = 1, 2, \ldots\} \\ \leq\ & P\{S_k \in B(0, \epsilon_i), |S_{n+k} - S_k| \geq \epsilon - \epsilon_i, n = 1, 2, \ldots\} \\ =\ & P\{S_k \in B(0, \epsilon_i)\} P\{|S_{n+k} - S_k| \geq \epsilon - \epsilon_i, n = 1, 2, \ldots\} \\ =\ & P\{S_k \in B(0, \epsilon_i)\} P\{|S_n| \geq \epsilon - \epsilon_i, n = 1, 2, \ldots\} = 0. \end{aligned}$$

Therefore, $P(A_k) = 0$ for all $k \geq 0$.

$$P\{S_n \in B(0,\epsilon) \text{ f. o.}\} = \sum_{K=0}^{\infty} P(A_k) = 0.$$

Because every neighborhood of 0 contains a set of the form $B(0,\epsilon)$ for some $\epsilon > 0$, we have that 0 is recurrent. □

2.2 The Role of Dimension

Lemma 2.10. *Let $\epsilon > 0, m \geq 2$, and let $|\cdot|$ be the norm $|x| = \max_{1 \leq i \leq d} |x_i|$. Then*

$$\sum_{n=0}^{\infty} P\{|S_n| < m\epsilon\} \leq (2m)^d \sum_{n=0}^{\infty} P\{|S_n| < \epsilon\}.$$

Proof. Note that

$$P\{|S_n| < m\epsilon\} \geq \sum_{\alpha \in \{-m,\ldots,m-1\}^d} P\{S_n \in \alpha\epsilon + [0,\epsilon)^d\}$$

and consider the stopping times

$$\tau_\alpha = \min\{n \geq 0 : S_n \in \alpha\epsilon + [0,\epsilon)^d\}.$$

Consequently,

$$\begin{aligned}
\sum_{n=0}^{\infty} P\{S_n \in \alpha\epsilon + [0,\epsilon)^d\} &= \sum_{k=0}^{\infty}\sum_{n=0}^{\infty} P\{S_n \in \alpha\epsilon + [0,\epsilon)^d, \tau_\alpha = k\} \\
&\leq \sum_{k=0}^{\infty}\sum_{n=0}^{\infty} P\{|S_n - S_k| < \epsilon, \tau_\alpha = k\} \\
&= \sum_{k=0}^{\infty}\sum_{n=k}^{\infty} P\{|S_n - S_k| < \epsilon\} P\{\tau_\alpha = k\} \\
&\leq \sum_{k=0}^{\infty}\left(\sum_{n=0}^{\infty} P\{|S_n| < \epsilon\}\right) P\{\tau_\alpha = k\} \leq \sum_{n=0}^{\infty} P\{|S_n| < \epsilon\}
\end{aligned}$$

because the sum on k is at most 1. The proof ends upon noting that the sum on α consists of $(2m)^d$ terms. □

Theorem 2.11 (Chung-Fuchs)**.** *Let $d = 1$. If the weak law of large numbers*

$$\frac{1}{n} S_n \to^P 0$$

holds, then S_n is recurrent.

Proof. Let $A > 0$. Using the lemma above with $d = 1$ and $\epsilon = 1$, we have that

$$\sum_{n=0}^{\infty} P\{|S_n| < 1\} \geq \frac{1}{2m}\sum_{n=0}^{\infty} P\{|S_n| < m\} \geq \frac{1}{2m}\sum_{n=0}^{Am} P\left\{|S_n| < \frac{n}{A}\right\}.$$

Note that, for any A, we have by the weak law of large numbers,

$$\lim_{n\to\infty} P\left\{|S_n| < \frac{n}{A}\right\} = 1.$$

If a sequence has a limit, then its Cesaro sum has the same limit and therefore,

$$\lim_{m\to\infty} \frac{1}{2m}\sum_{n=0}^{Am} P\left\{|S_n| < \frac{n}{A}\right\} = \frac{A}{2}.$$

Because A is arbitrary, the first sum above is infinite and the walk is recurrent. □

Theorem 2.12. *Let S be an random walk in $\mathbb{R}^2$ and assume that the central limit theorem holds in that*

$$\lim_{n\to\infty} P\left\{\frac{1}{\sqrt{n}}S_n \in B\right\} = \int_B n(x)\ dx$$

where n is a normal density whose support is $\mathbb{R}^2$, then S is recurrent.

Proof. By the lemma, with $\epsilon = 1$ and $d = 2$.

$$\sum_{n=0}^{\infty} P\{|S_n| < 1\} \geq \frac{1}{4m^2}\sum_{n=0}^{\infty} P\{|S_n| < m\}.$$

Let $c > 0$, then

$$n \to \infty \text{ and } \frac{m}{\sqrt{n}} \to c \quad \text{implies} \quad P\{|S_n| < m\} = P\left\{\frac{1}{\sqrt{n}}|S_n| < \frac{m}{\sqrt{n}}\right\} \to \int_{[-c.c]^2} n(x)\ dx.$$

Call this function $\rho(c)$, then for $\theta > 0$,

$$\lim_{m\to\infty} P\{|S_{[\theta m^2]}| < m\} = \rho(\theta^{-1/2}).$$

and we can write

$$\frac{1}{m^2}\sum_{n=0}^{\infty} P\{|S_n| < m\} = \int_0^{\infty} P\{|S_{[\theta m^2]}| < m\}\ d\theta.$$

Let $m \to \infty$ and use Fatou's lemma.

$$\liminf_{m\to\infty} \frac{1}{4m^2}\sum_{n=0}^{\infty} P\{|S_n| < m\} \geq \frac{1}{4}\int_0^{\infty} \rho(\theta^{-1/2})\ d\theta.$$

The proof is complete if we show the following.

Claim. $\int_0^\infty \rho(\theta^{-1/2})\ d\theta$ diverges.

Choose ϵ so that $n(x) \geq n(0)/2 > 0$ if $|x| < \epsilon$, then for $\theta > 1/\epsilon^2$

$$\rho(\theta^{-1/2}) = \int_{[-\theta^{-1/2},\theta^{-1/2}]} n(x)\ dx \geq \frac{1}{2}n(0)\frac{4}{\theta},$$

showing that the integral diverges. □

2.3 Simple Random Walks

To begin on $\mathbb{R}^1$, let the steps X_k take on integer values. Define the stopping time $\tau_0 = \min\{n > 0; S_n = 0\}$ and set the *probability of return to zero*

$$p_n = P\{S_n = 0\}$$

and *the first passage probabilities*

$$f_n = P\{\tau_0 = n\}.$$

Now define their generating functions

$$G_p(z) = \sum_{n=0}^{\infty} p_n z^n \quad \text{and} \quad G_f(z) = E[z^{\tau_0}] = \sum_{n=0}^{\infty} f_n z^n.$$

Note that τ_0 may be infinite. In this circumstance, $P\{\tau_0 < \infty\} = G_f(1) < 1$.

Proposition 2.13. *Using the notation above, we have*

$$G_p(z) = 1 + G_p(z)G_f(z).$$

Proof. Note that $p_0 = 1$ and $f_0 = 0$.

For $n \geq 1$,

$$\begin{aligned} p_n = P\{S_n = 0\} &= \sum_{k=0}^{n} P\{S_n = 0, \tau_0 = k\} = \sum_{k=0}^{n} P\{S_n = 0 | \tau_0 = k\}P\{\tau_0 = k\} \\ &= \sum_{k=0}^{n} P\{S_{n-k} = 0\}P\{\tau_0 = k\} = \sum_{k=0}^{n} p_{n-k} f_k. \end{aligned}$$

Now multiply both sides by z^n, sum on n, and use the property of multiplication of power series. □

Proposition 2.14. *For a simple random walk, with the steps X_k taking the value $+1$ with probability p and -1 with value $q = 1 - p$.*

1. $G_p(z) = (1 - 4pqz^2)^{-1/2}$.
2. $G_f(z) = 1 - (1 - 4pqz^2)^{1/2}$.

Proof.

1. Note that $p_n = 0$ if n is odd. For n even, we must have $n/2$ steps to the left and $n/2$ steps to the right. Thus,

$$p_n = \binom{n}{\frac{1}{2}n}(pq)^{n/2}.$$

Now, set $x = \sqrt{pq}z$ and use the binomial power series for $(1 - x)^{-1/2}$.

2. Apply the formula in 1 with the identity in the previous proposition.

□

Corollary 2.15. $P\{\tau_0 < \infty\} = 1 - |p - q|$

Proof. $G_f(1) = 1 - \sqrt{1 - 4pq}$. Now a little algebra gives the result. □

Exercise 2.16. *Show that* $E\tau_0 = G_f'(1) = \infty$.

Exercise 2.17. *Let* S *be a simple symmetric random walk on* $\mathbb{Z}$. *Show that*

$$\lim_{n\to\infty} \sqrt{\pi n} P\{S_{2n} = 0\} = 1.$$

and find c_n *so that*

$$\lim_{n\to\infty} c_n P\{\tau_0 = 2n\} = 1.$$

Hint: *Use the Stirling formula.*

Exercise 2.18. *Let* S *be a random walk in* $\mathbb{R}^d$ *having the property that its components form* d *independent simple symmetric random walks on* $\mathbb{Z}$. *Show that* S *is recurrent if* $d = 1$ *or* 2, *but is not recurrent if* $d \geq 3$.

Exercise 2.19. *Let* S *be the simple symmetric random walk in* $\mathbb{R}^3$. *Show that*

$$P\{S_{2n} = 0\} = \frac{1}{6^{2n}} \binom{2n}{n} \sum_{j=0}^{n} \sum_{k=1}^{n-j} \binom{n}{j, k, (n-j-k)}^2 \quad n = 0, 1, \ldots.$$

Exercise 2.20. *The simple symmetric random walk in* $\mathbb{R}^3$ *is not recurrent.* Hint: *For a nonnegative sequence* $a, \ldots, a_m$, $\sum_{\ell=1}^{n} a_\ell^2 \leq (\max_{1\leq \ell \leq m} a_\ell) \sum_{\ell=1}^{n} a_\ell$.

For S, a simple symmetric random walk in $\mathbb{R}^d$, $d > 3$. Let $\tilde{S}$ be the projection onto the first three coordinates. Define the stopping times

$$\tau_0 = 0, \quad \tau_{k+1} = \inf\{n > \tau_k : \tilde{S}_n \neq \tilde{S}_{\tau_k}\}.$$

Then $\{\tilde{S}_{\tau_k} : k \geq 0\}$ is a simple symmetric random walk in $\mathbb{R}^3$ and consequently return infinitely often to zero with probability 0. Consequently, S is not recurrent.

3 Renewal Sequences

Definition 3.1. *1. A random $\{0,1\}$-valued sequence X is a* renewal sequence *if*

$$X_0 = 1, \quad P\{X_n = \epsilon_n, 0 < n \leq r+s\} = P\{X_n = \epsilon_n, 0 < n \leq r\}P\{X_{n-r} = \epsilon_n, r \leq n \leq r+s\}$$

for all positive integers r and s and sequences $(\epsilon_1, \ldots, \epsilon_{r+s}) \in \{0,1\}^{(r+s)}$ such that $\epsilon_r = 1$.

2. *The random set $\Sigma_X = \{n; X_n = 1\}$ is called the* regenerative set *for the renewal sequence X.*

3. *The stopping times*

$$T_0 = 0, \quad T_m = \inf\{n > T_{m-1}; X_n = 1\}$$

are called the renewal times*, their differences*

$$W_j = T_j - T_{j-1}, \quad j = 1, 2, \ldots.$$

are called the waiting *or* sojourn times *and*

$$T_m = \sum_{j=1}^{n} W_j.$$

A renewal sequence is a random sequence whose distribution beginning at a renewal time is the same as the distribution of the original renewal sequence. We can characterize the renewal sequence in any one of four equivalent ways, through

1. the renewal sequence X,
2. the regenerative set Σ_X,
3. the sequence of renewal times T, or
4. the sequence of sojourn times W.

Check that you can move easily from one description to another.

Definition 3.2. *1. For a sequence, $\{a_n : n \geq 1\}$, define the* generating function

$$G_a(z) = \sum_{k=0}^{\infty} a_k z^k.$$

2. *For pair of sequences, $\{a_n : n \geq 1\}$ and $\{b_n : n \geq 1\}$, define the* convolution sequence

$$(a * b)_n = \sum_{k=0}^{n} a_k b_{n-k}.$$

3. *For a pair of measures A and B on $\mathbb{N}$, define the* convolution measure

$$(A * B)(K) = \sum_{j+k \in K} A\{j\}B\{k\}.$$

By the uniqueness theorem for power series, a sequence is uniquely determined by its generating function. Convolution is associative and commutative. We shall write $a^{*2} = a * a$ and a^{*m} for the m-fold convolution.

Exercise 3.3. *1.* $G_{a*b}(z) = G_a(z)G_b(z)$.

2. If X and Y are independent $\mathbb{N}$-valued random variables, then the mass function for $X + Y$ is the convolution of the mass functions for X and the mass function for Y.

Exercise 3.4. *The sequence $\{W_j : j \geq 1\}$ are independent and identically distributed.*

Thus, T is a random walk on $\mathbb{Z}^+$ having on strictly positive integral steps. If we let R be the waiting time distribution for T, then

$$R^{*m}(K) = P\{T_m \in K\}.$$

We can allow for

$$P\{X_n = 0 \text{ for all } n \geq 1\} > 0.$$

In this case the step size will be infinite with positive probability, $R\{\infty\}$. This can be determined from the generating function because $R\{\infty\} = 1 - G_R(1)$.

There are many regenerative sets in a random walk.

Exercise 3.5. *Let S be a random walk on $\mathbb{R}$, then the following are regenerative sets.*

1. $\{n; S_n = 0\}$
2. $\{n; S_n > S_k$ *for* $0 \leq k < n\}$ *(strict ascending ladder times)*
3. $\{n; S_n < S_k$ *for* $0 \leq k < n\}$ *(strict descending ladder times)*
4. $\{n; S_n \geq S_k$ *for* $0 \leq k < n\}$ *(weak ascending ladder times)*
5. $\{n; S_n \leq S_k$ *for* $0 \leq k < n\}$ *(weak decending ladder times)*

   If S is a random walk in Z, then the following are regenerative sets.
6. $\{S_n; S_n > S_k$ *for* $0 \leq k < n\}$
7. $\{-S_n; S_n > S_k$ *for* $0 \leq k < n\}$

Theorem 3.6. *Let X and Y be independent renewal sequencces, then so is $\{X_0Y_0, X_1Y_1, \ldots\}$. In other words, the coincident times of independent renewal sequences is itself a renewal sequence.*

Proof. Choose natural numbers r and s and fix a sequence $(\epsilon_0, \ldots, \epsilon_{r+s}) \in \{0,1\}^{(r+s+1)}$ with $\epsilon_r = 1$. Let A be the sets of pairs of sequences $(\epsilon_0^X, \ldots, \epsilon_{r+s}^X)$ and $(\epsilon_0^Y, \ldots, \epsilon_{r+s}^Y)$ in $\{0,1\}^{(r+s+1)}$ whose term by term product is $(\epsilon_0, \ldots, \epsilon_{r+s})$. Note that $\epsilon_r^X = \epsilon_r^Y = 1$.

$$
\begin{aligned}
& P\{X_nY_n = \epsilon_n, n = 1, \ldots, r+s\} \\
= & \sum_A P\{X_n = \epsilon_n^X, Y_n = \epsilon_n^Y, n = 1, \ldots, r+s\} \\
= & \sum_A P\{X_n = \epsilon_n^X, n = 1, \ldots, r+s\} P\{Y_n = \epsilon_n^Y, n = 1, \ldots, r+s\} \\
= & \sum_A P\{X_n = \epsilon_n^X, n = 1, \ldots, r\} P\{X_{n-r} = \epsilon_n^X, n = r+1, \ldots, r+s\} \\
& \quad \times P\{Y_n = \epsilon_n^Y, n = 1, \ldots, r\} P\{Y_{n-r} = \epsilon_n^Y, n = r+1, \ldots, r+s\} \\
= & \sum_A P\{X_n = \epsilon_n^X, Y_n = \epsilon_n^Y, n = 1, \ldots, r\} P\{X_{n-r} = \epsilon_n^X, Y_n = \epsilon_n^Y, n = r+1, \ldots, r+s\} \\
= & P\{X_nY_n = \epsilon_n, n = 1, \ldots, r\} P\{X_{n-r}Y_{n-r} = \epsilon_n, n = r+1, \ldots, r+s\}
\end{aligned}
$$

□

3.1 Waiting Time Distributions and Potential Sequences

Definition 3.7. *1. For a renewal sequence X define the* renewal measure

$$N(B) = \sum_{n \in B} X_n = \sum_{n=0}^{\infty} I_B(T_n).$$

The σ-finite random measure N gives the number of renewels during a time set B.

2. $U(B) = EN(B) = \sum_{n=0}^{\infty} P\{T_n \in B\} = \sum_{n=0}^{\infty} R^{*n}(B)$ *is the* potential measure*.*
   Use the monotone convergence theorem to show that it is a measure.

3. *The sequence $\{u_k : k \geq 0\}$ defined by*

$$u_k = P\{X_k = 1\} = U\{k\}$$

is called the potential sequence *for X.*

Note that

$$u_k = \sum_{n=0}^{\infty} P\{T_n = k\}.$$

Exercise 3.8. *1.*

$$G_U = 1/(1 - G_R) \tag{3.1}$$

2. *$U = (U * R) + \delta_0$ where δ_0 denotes the delta distribution.*

Consequently, the waiting time distribution, the potential measure and the generating function for the waiting time distribution each uniquely determine the distribution of the renewal sequence.

Theorem 3.9. *For a renewal process, let R denote the waiting time distribution, $\{u_n; n \geq 0\}$ the potential sequence, and N the renewal measure. For non-negative integers k and ℓ,*

$$\begin{aligned} P\{N[k+1,k+\ell] > 0\} &= \sum_{n=0}^{k} R[k+1-n, k+\ell-n]u_n \\ &= 1 - \sum_{n=0}^{k} R[k+\ell+1-n,\infty]u_n = \sum_{n=k+1}^{k+\ell} R[k+\ell+1-n,\infty]u_n. \end{aligned}$$

Proof. If we decompose the event according to the last renewal in $[k+1, k+\ell]$, we obtain

$$\begin{aligned} P\{N[k+1,k+\ell] > 0\} &= \sum_{j=0}^{\infty} P\{T_j \leq k, k+1 \leq T_{j+1} \leq k+\ell\} \\ &= \sum_{j=0}^{\infty}\sum_{n=0}^{k} P\{T_j = n, k+1 \leq T_{j+1} \leq k+\ell\} \\ &= \sum_{j=0}^{\infty}\sum_{n=0}^{k} P\{T_j = n, k+1-n \leq W_{j+1} \leq k+\ell-n\} \\ &= \sum_{j=0}^{\infty}\sum_{n=0}^{k} P\{T_j = n\}P\{k+1-n \leq W_{j+1} \leq k+\ell-n\} \\ &= \sum_{n=0}^{k} R[k+1-n, k+\ell-n]u_n. \end{aligned}$$

On the other hand,

$$\begin{aligned} P\{N[k+1,k+\ell] = 0\} &= \sum_{j=0}^{\infty} P\{T_j \leq k, T_{j+1} > k+\ell\} = \sum_{j=0}^{\infty}\sum_{n=0}^{k} P\{T_j = n, T_{j+1} > k+\ell\} \\ &= \sum_{j=0}^{\infty}\sum_{n=0}^{k} P\{T_j = n, W_{j+1} > k+\ell-n\} = \sum_{j=0}^{\infty}\sum_{n=0}^{k} P\{T_j = n\}P\{W_{j+1} > k+\ell-n\} \\ &= \sum_{n=0}^{k} R[k+\ell+1-n,\infty]u_n. \end{aligned}$$

Now, take complements.

Finally, if we decompose the event according to the first renewal in $[k+1, k+\ell]$, we obtain

$$\begin{aligned} P\{N[k+1,k+\ell] > 0\} &= \sum_{j=0}^{\infty} P\{k+1 \le T_j \le k+\ell, T_{j+1} > k+\ell\} \\ &= \sum_{j=0}^{\infty}\sum_{n=k+1}^{k+\ell} P\{T_j = n, T_{j+1} > k+\ell\} = \sum_{j=0}^{\infty}\sum_{n=k+1}^{k+\ell} P\{T_j = n, W_{j+1} > k+\ell-n\} \\ &= \sum_{j=0}^{\infty}\sum_{n=k+1}^{k+\ell} P\{T_j = n\}P\{W_{j+1} > k+\ell-n\} = \sum_{n=k+1}^{k+\ell} R[k+\ell+1-n,\infty]u_n. \end{aligned}$$

□

Taking $\ell = 0$ for the second statement in the theorem above gives:

Corollary 3.10. *For each $k \ge 0$,*

$$1 = \sum_{n=0}^{k} R[k+1-n,\infty]u_n = \sum_{n=0}^{k} R[n+1,\infty]u_{k-n}. \tag{3.2}$$

Exercise 3.11. *Let X be a renewal sequence with potential measure U and waiting time distribution R. If $U(\mathbb{Z}^+) < \infty$, then $N(\mathbb{Z}^+)$ is a geometric random variable with mean $U(\mathbb{Z}^+)$. If $U(\mathbb{Z}^+) = \infty$, then $N(\mathbb{Z}^+) = \infty$ a.s. In either case,*

$$U(\mathbb{Z}^+) = \frac{1}{R\{\infty\}}.$$

Definition 3.12. *A renewal sequence and the corresponding regenerative set are* transient *if the corresponding renewal measure is finite almost surely and they are* recurrent *otherwise.*

Theorem 3.13 (Strong Law for Renewal Sequences)**.** *Given a renewal sequence X, let N denote the renewal measure, U, the potential measure and $\mu \in [1,\infty]$ denote the mean waiting time. Then,*

$$\lim_{n\to\infty} \frac{1}{n}U[0,n] = \frac{1}{\mu}, \quad \lim_{n\to\infty} \frac{1}{n}N[0,n] = \frac{1}{\mu} \quad a.s. \tag{3.3}$$

Proof. Because $N[0,n] \le n+1$, the first equality follows from the second equality and the bounded convergence theorem.

To establish this equality, note that by the strong law of large numbers,

$$\frac{T_m}{m} = \frac{1}{m}\sum_{j=1}^{m} W_j = \mu \quad \text{a.s.}.$$

If $T_{m-1} \le n < T_m$, then $N[0,n] = m$, and

$$\frac{m}{T_m} \ge \frac{1}{n}N[0,n] = \frac{m}{n} \ge \frac{m}{T_{m-1}} = \frac{m}{m-1}\frac{m-1}{T_{m-1}}$$

and the equality follows by taking limits as $n \to \infty$. □

Definition 3.14. *Call a recurrent renewal sequence* positive recurrent *if the mean waiting time is finite and* null recurrent *if the mean waiting time is infinite.*

Definition 3.15. *Let X be a random sequence $X_n \in \{0,1\}$ for all n and let $\tau = \min\{n \geq 0; X_n = 1\}$. Then X is a* delayed renewal seqence *if either $P\{\tau = \infty\} = 1$ or, conditioned on the event $\{\tau < \infty\}$,*

$$\{X_{\tau+n}; n \geq 0\}$$

is a renewal sequence. τ is called the delay time.

Exercise 3.16. *Let S be a random walk on $\mathbb{Z}$ and let $X_n = I_{\{x\}}(S_n)$. Then X is a delayed renewal sequence.*

Exercise 3.17. *Let X be a renewal sequence. Then a $\{0,1\}$-valued sequence, $\tilde{X}$, is a delayed renewal sequence with the same waiting time distribution as X if and only if*

$$P\{\tilde{X}_n = \epsilon_n \text{ for } 0 \leq n \leq r+s\} = P\{\tilde{X}_n = \epsilon_n \text{ for } 0 \leq n \leq r\}P\{X_{n-r} = \epsilon_n \text{ for } r < n \leq r+s\} \tag{3.4}$$

for all $r, s \in N$ and $(\epsilon_0, \ldots, \epsilon_{r+s}) \in \{0,1\}^{(r+s+1)}$ satisfying $\epsilon_r = 1$.

Theorem 3.18. *Let X and $\tilde{Y}$ be independent delayed renewal sequences, both with waiting time distribution R. Define the $\mathcal{F}^{(X,Y)}$-stopping time $\sigma = \min\{n \geq 0 : X_n = \tilde{Y}_n = 1\}$. Define*

$$\tilde{X}_n = \begin{cases} X_n & \text{if } n \leq \sigma, \\ \tilde{Y}_n & \text{if } n > \sigma \end{cases}$$

Then X and $\tilde{X}$ have the same distribution. In other words, $\tilde{X}$ and $\tilde{Y}$ is a coupling of the renewal processes having marginal distributions X and Y with coupling time σ.

Proof. By the Daniell-Kolmogorov extension theorem, it is enough to show that they have the same finite dimensional distributions.

With this in mind, fix n and $(\epsilon_0, \ldots, \epsilon_n) \in \{0,1\}^{(n+1)}$. Define

$$\tilde{\sigma} = \min\{n, \min\{k, \tilde{Y}_k = \epsilon_k = 1\}\}.$$

Note that on the set $\{\tilde{X}_0 = \epsilon_0, \ldots, \tilde{X}_n = \epsilon_n\}$, $\tilde{\sigma} = \min\{n, \sigma\}$. Noting that X and $(\tilde{Y}, \tilde{\sigma})$ are independent, we have

$$\begin{aligned} P\{\tilde{X}_0 = \epsilon_0 \ldots, \tilde{X}_n = \epsilon_n\} &= \sum_{k=0}^{n} P\{\tilde{\sigma} = k, \tilde{X}_0 = \epsilon_0, \ldots, \tilde{X}_n = \epsilon_n\} \\ &= \sum_{k=0}^{n} P\{\tilde{\sigma} = k, X_0 = \epsilon_0, \ldots, X_k = \epsilon_k, Y_{k+1} = \epsilon_{k+1}, \ldots, Y_n = \epsilon_n\} \\ &= \sum_{k=0}^{n} P\{X_0 = \epsilon_0, \ldots, X_k = \epsilon_k\}P\{\tilde{\sigma} = k, \tilde{Y}_{k+1} = \epsilon_{k+1}, \ldots, \tilde{Y}_n = \epsilon_n\} \end{aligned}$$

Claim. Let Y be a non-delayed renewal sequence with waiting distribution R. Then for $k \leq n$

$$P\{\tilde{\sigma} = k, \tilde{Y}_{k+1} = \epsilon_{k+1}, \ldots, \tilde{Y}_n = \epsilon_n\} = P\{\tilde{\sigma} = k\}P\{Y_{k+1} = \epsilon_{k+1}, \ldots, Y_n = \epsilon_n\}.$$

If $k = n$, then because $k+1 > n$, the given event is impossible and both sides of the equation equal zero.

For $k < n$ and $\epsilon_k = 0$, then $\{\tilde\sigma = k\} = \emptyset$ and again the equation is trivial.

For $k < n$ and $\epsilon_k = 1$, then $\{\tilde\sigma = k\}$ is the finite union of events of the form $\{\tilde Y_0 = \tilde\epsilon_0, \ldots, \tilde Y_k = 1\}$ and the claim follows from the identity (3.4) above on delayed renewal sequences.

Use this identity once more to obtain

$$\begin{aligned} P\{\tilde X_0 = \epsilon_0 \ldots, \tilde X_n = \epsilon_n\} &= \sum_{k=0}^{n} (P\{\tilde\sigma = k\}P\{X_0 = \epsilon_0, \ldots, X_k = \epsilon_k\}P\{Y_{k+1} = \epsilon_{k+1}, \ldots, Y_n = \epsilon_n\}) \\ &= \sum_{k=0}^{n} P\{\tilde\sigma = k\}P\{X_0 = \epsilon_0, \ldots, X_n = \epsilon_n\} = P\{X_0 = \epsilon_0, \ldots, X_n = \epsilon_n\}. \end{aligned}$$

□

Exercise 3.19. *Let W be an $\mathbb{N}$-valued random variable with mean μ, then*

$$\sum_{n=1}^{\infty} P\{W \geq n\} = \mu.$$

More generally, set

$$f(n) = \sum_{k=1}^{n} a_n,$$

then

$$Ef(W) = \sum_{n=1}^{\infty} a_n P\{W \geq n\}.$$

Theorem 3.20. *Let R be a waiting distribution with finite mean μ and let $\tilde Y$ be a delayed renewal sequence having waiting time distribution R and delay distribution*

$$D\{n\} = \frac{1}{\mu}R[n+1, \infty), \quad n \in \mathbb{Z}^+.$$

Then $P\{\tilde Y_k = 1\} = 1/\mu$.

Proof. By the exercise above, D is a distribution. Let T_0, the waiting time, be a random variable having this distribution. Let $\{u_n; n \geq 1\}$ be the potential sequence corresponding to R. Then $P\{X_n = 1|T_0 = k\} = u_{n-k}$ provided $k \leq n$ and 0 otherwise. Hence,

$$P\{\tilde Y_k = 1\} = \sum_{n=0}^{k} P\{\tilde Y_k = 1|T_0 = n\}P\{T_0 = n\} = \sum_{n=0}^{k} u_{k-n}\frac{1}{\mu}R[n+1, \infty] = \frac{1}{\mu}$$

by the identity (3.2). □

3.2 Renewal Theorem

Definition 3.21. *The* period *of a renewal sequence is*

$$gcd\{n > 0; u_n > 0\}, \quad gcd(\emptyset) = \infty.$$

A renewal process is aperiodic *if its period is one.*

Note that if B is a set of positive integers, $\gcd(B) = 1$, then there exists K such that every integer $k > K$ can be written as a positive linear combination of elements in B.

Theorem 3.22 (Renewal)**.** *Let $\mu \in [1, \infty]$ and let X be an aperiodic renewal sequence, then*

$$\lim_{k\to\infty} P\{X_k = 1\} = \frac{1}{\mu}.$$

Proof. (transient case) In this case $R\{\infty\} > 0$ and thus $\mu = \infty$. Recall that $u_k = P\{X_k = 1\}$. Because

$$\sum_{k=0}^{\infty} u_k < \infty,$$

$\lim_{k\to\infty} u_k = 0 = 1/\mu$.

(positive recurrent) Let $\tilde{Y}$ be a delayed renewal sequence, independent of X, having the same waiting time distribution as X and satisfying

$$P\{\tilde{Y}_k = 1\} = \frac{1}{\mu}.$$

Define the stopping time

$$\sigma = \min\{k \geq 0; X_k = \tilde{Y}_k = 1\}.$$

and the process

$$\tilde{X}_k = \begin{cases} X_k & \text{if } k \leq \sigma, \\ \tilde{Y}_k & \text{if } k > \sigma. \end{cases}$$

Then X and $\tilde{X}$ have the same distribution. If $P\{\sigma < \infty\} = 1$, then

$$\begin{aligned} & |P\{X_k = 1\} - \frac{1}{\mu}| \\ = & |P\{\tilde{X}_k = 1\} - P\{\tilde{Y}_k = 1\}| \\ = & |(P\{\tilde{X}_k = 1, \sigma \leq k\} - P\{\tilde{Y}_k = 1, \sigma \leq k\}) + (P\{\tilde{X}_k = 1, \sigma > k\} - P\{\tilde{Y}_k = 1, \sigma > n\})| \\ = & |P\{\tilde{X}_k = 1, \sigma > k\} - P\{\tilde{Y}_k = 1, \sigma > k\}| \leq \max\{P\{\tilde{X}_k = 1, \sigma > k\}, P\{\tilde{Y}_k = 1, \sigma > k\}\}P\{\sigma > k\}. \end{aligned}$$

Thus, the renewal theorem holds in the positive recurrent case if σ is finite almost surely. With this in mind, define the $\mathcal{F}_t^Y$-stopping time

$$\tau = \min\{k; \tilde{Y}_k = 1 \text{ and } u_k > 0\}.$$

Because $u_k > 0$ for sufficiently large k, and $\tilde{Y}$ has a finite waiting, then $\tilde{Y}$ has a finite delay, τ is finite with probability 1. Note that X, τ and $\{\tilde{Y}_{\tau+k}; k = 0, 1, \ldots\}$ are independent. Consequently $\{X_k\tilde{Y}_{\tau+k}; k = 0, 1, \ldots\}$ is a renewal sequence independent of τ with potential sequence

$$P\{X_k\tilde{Y}_{\tau+k} = 1\} = P\{X_k = 1\}P\{\tilde{Y}_{\tau+k} = 1\} = u_k^2.$$

Note that $\sum_{k=1}^{\infty} u_k^2 = \infty$. Otherwise, $\lim_{k\to\infty} u_k = 0$ contradicting from (3.3), the strong law for renewal sequences, that $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} u_k = 1/\mu$. Thus, the renewal process is recurrent. i.e.,

$$P\{X_k = \tilde{Y}_{\tau+k} = 1 \text{ i.o}\} = 1.$$

Let $\{\sigma_m; m \geq 1\}$ denote the times that this holds.

Because this sequence is strictly increasing, $\sigma_{2mj} + m < \sigma_{2k(m+1)}$. Consequently, the random variables

$$\{X_{\sigma_{2km}+m}; j \geq 1\}$$

are independent and take the value one with probability u_m. If $u_m > 0$, then

$$\sum_{k=1}^{\infty} P\{X_{\sigma_{2km}+m} = 1\} = \sum_{k=1}^{\infty} u_m = \infty.$$

Thus, by the second Borel-Cantelli lemma,

$$P\{X_{\sigma_{2mj+m}} = 1 \text{ i.o}\} = 1.$$

Recall that τ is independent of X and $\{\tilde{Y}_{\tau+k}; k \geq 1\}$ and so

$$P\{X_{\sigma_{2\tau j+\tau}} = 1 \text{ i.o}\} = E[P\{X_{\sigma_{2\tau j+\tau}} = 1 \text{ i.o.}|\tau\}] = \sum_{m=1}^{\infty} P\{X_{\sigma_{2mj+m}} = 1 \text{ i.o}\}P\{\tau = m\} = 1.$$

Each one of these occurences is an example of $\sigma < \infty$.

(null recurrent) By way of contradiction, choose $\epsilon > 0$, so that

$$\liminf_{k\to\infty} u_k > \epsilon.$$

Now pick an integer q so that

$$\sum_{n=1}^{q} R[n, \infty) > \frac{2}{\epsilon}.$$

Introduce the independent delayed renewal processes $\tilde{Y}^0, \ldots, \tilde{Y}^q$ with waiting distribution R and fixed delay r for $\tilde{Y}^r$. Thus X and $\tilde{Y}^0$ have the same distribution.

Guided by the proof of the positive recurrent case, define the stopping time

$$\sigma \ = \min\{k \geq 1 : \tilde{Y}_k^r = 1, \text{ for all } r = 0, \ldots, q\},$$

and set

$$\tilde{X}_k^r = \begin{cases} \tilde{Y}_k^r & \text{if } k \leq \sigma, \\ X_k & \text{if } k > \sigma. \end{cases}$$

We have shown that $\tilde{X}^r$ and $\tilde{Y}^r$ have the same distribution.

Because $u_k > \epsilon$ infinitely often, $\sum_{k=1}^{\infty} u_k^{q+1} = \infty$. Modify the argument about to see that $\sigma < \infty$ with probability one. Consequenty, there exists an integer $k > q$ so that

$$P\{\sigma < k\} \geq 1/2 \quad \text{and} \quad u_k = P\{X_k = 1\} > \epsilon.$$

Thus,

$$u_{k-r} = P\{\tilde{X}_k^r = 1\} \geq P\{\tilde{X}_k^r = 1, \sigma < k\} = P\{\tilde{X}_k^r = 1|\sigma < k\}P\{\sigma \leq k\} \geq \frac{\epsilon}{2}.$$

Consequently

$$\sum_{n=1}^{q+1} u_{k+1-n}R[n,\infty) > 1.$$

However identity (3.2) states that this sum is at most one. This contradiction proves the null recurrent case. □

Putting this altogether, we see that a renewal sequence is

recurrent	if and only if	$\sum_{n=0}^{\infty} u_n = \infty$,
null recurrent	if and only if	it is recurrent and $\lim_{n\to\infty} u_n = 0$.

Exercise 3.23. *Fill in the details for the proof of the renewal theorem in the null recurrent case.*

Corollary 3.24. *Let $\mu \in [1,\infty]$ and let X be an renewal sequence with period γ, then*

$$\lim_{k\to\infty} P\{X_{\gamma k} = 1\} = \frac{\gamma}{\mu}.$$

Proof. The process Y defined by $Y_k = X_{\gamma k}$ is an aperiodic renewal process with renewal measure $R_Y(B) = R_X(\gamma B)$ and mean μ/γ. □

3.3 Applications to Random Walks

We now apply this to random walks.

Lemma 3.25. *Let $n \geq 0$. For a real-valued random walk, S, the probability that n is a strict ascending ladder time equals the probability that there is not positive ladder time less than or equal to n*

Proof. Let $\{X_k : k \geq 1\}$ be the steps in S. Then

$$\{n \text{ is a strict ascending ladder time for } S\} = \left\{\sum_{k=m}^{n} X_k > 0; \text{ for } m = 1,2,\ldots,n\right\}.$$

and

$$\{1,2,\ldots,n \text{ is not a weak descending ladder time for } S\} = \left\{\sum_{k=1}^{m} X_k > 0; \text{ for } m = 1,2,\ldots,n\right\}.$$

By replacing X_k with X_{n+1-k}, we see that these two events have the same probability. □

Theorem 3.26. *Let G_{++} and G_- be the generating functions for the waiting time distributions for strict ascending ladder times and weak descending latter times for a real-valued random walk. Then*

$$(1 - G_{++}(z))\,(1 - G_-(z)) = (1 - z)\,.$$

Proof. Let u_n^{++} be the probability that n is a strict ascending ladder time and r_k^- be the probability that k is the smallest positive weak ascending ladder time. Then the lemma above states that

$$u_n^{++} = 1 - \sum_{k=0}^{n} r_k^- .$$

By the generation function identity (3.1), $(1 - G_{++}(z))^{-1}$ is the generating function of the sequence $\{u_n^{++} : n \geq 1\}$, we have for $z \in [0, 1)$,

$$\begin{aligned} \frac{1}{1 - G_{++}(z)} &= \sum_{n=0}^{\infty}(1 - \sum_{k=0}^{n} r_k^-)z^n = \frac{1}{1-z} - \sum_{k=0}^{\infty}\sum_{n=k}^{\infty} r_k^- z^n \\ &= \frac{1}{1-z} - \sum_{k=0}^{\infty} r_k^- \frac{z^k}{1-z} = \frac{1 - G_-(z)}{1-z}. \end{aligned}$$

The case $z = 1$ now follows from the monotone convergence theorem. □

Exercise 3.27. *Either the set of strict ascending ladder times and the set of weak descending ladder times are both null recurrent or one of these sets is transient and the other is positive recurrent.*

Remark 3.28. *If the step distribution in a random walk assigns zero probability to each one-point set, the weak and strict ladder times are almost surely equal. In this case, we have, with the obvious notational addition,*

$$(1 - G_{++}(z))\,(1 - G_{--}(z)) = (1 - z)\,.$$

If the step distribution is symmetric about zero, then

$$G_{++}(z) = G_{--}(z) = 1 - \sqrt{1 - z}$$

and a Taylor's series expansion gives the waiting time distribution of ladder times.

4 Martingales

Definition 4.1. *A real-valued random sequence X with $E|X_n| < \infty$ for all $n \geq 0$ and adapted to a filtration $\{\mathcal{F}_n; n \geq 0\}$ is an $\mathcal{F}_n$-*martingale *if*

$$E[X_{n+1}|\mathcal{F}_n] = X_n \quad \text{for } n = 0, 1, \ldots,$$

*an $\mathcal{F}_n$-*submartingale *if*

$$E[X_{n+1}|\mathcal{F}_n] \geq X_n \quad \text{for } n = 0, 1, \ldots,$$

*and an $\mathcal{F}_n$-*supermartingale *if the inequality above is reversed.*

Thus, X is a supermartingale if and only if $-X$ is a submartingale and X is a martingale if and only if X and $-X$ are submartingales. If X is an $\mathcal{F}_n^X$-*martingale*, then we simply say that X is a martingale.

Exercise 4.2. *1. Show that X is an $\mathcal{F}_n$-martingale if and only if for every $j < k$,*

$$E[X_k; A] = E[X_j; A], \quad \text{for all } A \in \mathcal{F}_j.$$

The case $A = \Omega$ shows that EX_k is constant.

2. Give a similar characterization for submartingales.

3. Show that if X is an $\mathcal{F}_n$-martingale, then X is a martingale.

Proposition 4.3. *1. Suppose X is an $\mathcal{F}_n$-martingale, ϕ is convex and $E|\phi(X_n)| < \infty$ for all $n \geq 1$. Then $\phi \circ X$ is an $\mathcal{F}_n$-submartingale.*

2. Suppose X is an $\mathcal{F}_n$-submartingale, ϕ is convex and non-decreasing, $E|\phi(X_n)| < \infty$ for all $n \geq 1$. Then $\phi \circ X$ is an $\mathcal{F}_n$-submartingale.

Proof. By Jensen's inequality,

$$E[\phi(X_{n+1})|\mathcal{F}_n] \geq \phi(E[X_{n+1}|\mathcal{F}_n]) \geq \phi(X_n).$$

For part 1, the last inequality is an equality. For part 2, the last inequality follows from the assumption that ϕ is nondecreasing. □

Exercise 4.4. *Let X and Y are $\mathcal{F}_n$-submartingales, then $\{\max\{X_n, Y_n\}; n \geq 0\}$ is an $\mathcal{F}_n$-submartingale.*

Exercise 4.5. *Let X and $\tilde{X}$ are $\mathcal{F}_n$-supermartingales, and let τ be a finite $\mathcal{F}_n$-stopping time. Assume that $X_\tau \geq \tilde{X}_\tau$ almost surely. Show that*

$$Y_n = \begin{cases} X_n & \text{if } n \leq \tau \\ \tilde{X}_n & \text{if } n > \tau \end{cases}$$

is an $\mathcal{F}_n$-supermartingale.

4.1 Examples

1. Let S be a random walk whose step sizes have mean μ. Then S is $\mathcal{F}_n^X$-submartingale, martingale, or supermartingale according to the property $\mu > 0$, $\mu = 0$, and $\mu < 0$.

   To see this, note that
   $$E|S_n| \leq \sum_{k=1}^{n} E|X_k| = nE|X_1|$$
   and
   $$E[S_{n+1}|\mathcal{F}_n^X] = E[S_n + X_{n+1}|\mathcal{F}_n^X] = E[S_n|\mathcal{F}_n^X] + E[X_{n+1}|\mathcal{F}_n^X] = S_n + \mu.$$

2. In addition,
   $$E[S_{n+1} - (n+1)\mu|\mathcal{F}_n^X] = E[S_{n+1}|\mathcal{F}_n^X] - (n+1)\mu = S_n + \mu - (n+1)\mu = S_n - n\mu.$$
   and $\{S_n - n\mu : n \geq 0\}$ is a $\mathcal{F}_n^X$-martingale.

3. Now, assume $\mu = 0$ and that $\mathrm{Var}(X_1) = \sigma^2$ let $Y_n = S_n^2 - n\sigma^2$. Then
   $$\begin{aligned} E[Y_{n+1}|\mathcal{F}_n^X] &= E[(S_n + X_{n+1})^2|\mathcal{F}_n^X] - (n+1)\sigma^2 \\ &= E[S_n^2 + 2S_nX_{n+1} + X_{n+1}^2|\mathcal{F}_n^X] - (n+1)\sigma^2 \\ &= +S_n^2 + 2S_nE[X_{n+1}|\mathcal{F}_n^X] + E[X_{n+1}^2|\mathcal{F}_n^X] - (n+1)\sigma^2 \\ &= +S_n^2 + 2 \times S_n \times 0 + E[X_{n+1}^2] - (n+1)\sigma^2 \\ &= S_n^2 + \sigma^2 - (n+1)\sigma^2 = Y_n \end{aligned}$$

4. (Wald's martingale) Let $\phi(t) = E[\exp(itX_1)]$ be the characteristic function for the step distribution for the random walk S. Fix $t \in \mathbb{R}$ and define $Y_n = \exp(itS_n)/\phi(t)^n$. Then
   $$\begin{aligned} E[Y_{n+1}|\mathcal{F}_n^X] &= E[\exp(it(S_n + X_{n+1})|\mathcal{F}_n^X]/\phi(s)^{n+1} \\ &= \exp(isS_n)E[\exp(itX_{n+1})|\mathcal{F}_n^X]/\phi(t)^{n+1} \\ &= \exp(itS_n)/\phi(t)^n = Y_n. \end{aligned}$$
   A similar martingale can be constructed using the Laplace transform
   $$L(\alpha) = E[\exp(-\alpha X_1)].$$

5. (Likelihood ratios) Let $X_0, X_1, \ldots$ be independent and let f_0 and f_1 be probability density functions with respect to a σ-finite measure μ on some state space S. A stochastic process of fundamental importance in the theory of statistical inference is the sequence of likelihood ratios
   $$L_n = \frac{f_1(X_0)f_1(X_1)\cdots f_1(X_n)}{f_0(X_0)f_0(X_1)\cdots f_0(X_n)}, \quad n = 0, 1, \ldots.$$
   To assure that L_n is defined assume that $f_0(x) > 0$ for all $x \in S$. Then
   $$E[L_{n+1}|\mathcal{F}_n^X] = E\left[L_n \frac{f_1(X_{n+1})}{f_0(X_{n+1})}\Big|\mathcal{F}_n^X\right] = L_nE\left[\frac{f_1(X_{n+1})}{f_0(X_{n+1})}\right].$$

If f_0 is the probability density function for X_k, then

$$E\left[\frac{f_1(X_{n+1})}{f_0(X_{n+1})}\right] = \int_S \frac{f_1(x)}{f_0(x)} f_0(x)\ \mu(dx) = \int_S f_1(x)\ \mu(dx) = 1,$$

and L is a martingale.

6. Let $\{X_n; n \geq 0\}$ be independent mean 1 random variables. Then $Z_n = \prod_{k=1}^n X_k$ is an $\mathcal{F}_n^X$-martingale.

7. (Doob's martingale) Let X be a random variable, $E|X| < \infty$ and define $X_n = E[X|\mathcal{F}_n]$, then by the tower property,
$$E[X_{n+1}|\mathcal{F}_n] = E[E[X|\mathcal{F}_{n+1}]|\mathcal{F}_n] = E[X|\mathcal{F}_n] = X_n.$$

8. (Polya urn) An urn has initial number of balls, colored green and blue with at least one of each color. Draw a ball at random from the urn and return it and c others of the same color. Repeat this procedure, letting X_n be the number of blue and Y_n be the number of green after n iterations. Let R_n be the fraction of blue balls.
$$R_n = \frac{X_n}{X_n + Y_n}. \tag{4.1}$$
Then
$$\begin{aligned} E[R_{n+1}|\mathcal{F}_n^{(X,Y)}] &= \frac{X_n + c}{X_n + Y_n + c}\frac{X_n}{X_n + Y_n} + \frac{X_n}{X_n + Y_n + c}\frac{Y_n}{X_n + Y_n} \\ &= \frac{(X_n + c)X_n + X_nY_n}{(X_n + Y_n + c)(X_n + Y_n)} = R_n \end{aligned}$$

9. (martingale transform) For a filtration $\{\mathcal{F}_n : n \geq 0\}$, Call a random sequence H *predictable* if H_n is $\mathcal{F}_{n-1}$-measurable. Define $(H \cdot X)_0 = 0$ and
$$(H \cdot X)_n = \sum_{k=1}^n H_k(X_k - X_{k-1}).$$
Assume that H_n is non-negative and bounded. Then
$$\begin{aligned} E[(H \cdot X)_{n+1}|\mathcal{F}_n] &= (H \cdot X)_n + E[H_{n+1}(X_{n+1} - X_n)|\mathcal{F}_n] \\ &= (H \cdot X)_n + H_{n+1}E[(X_{n+1} - X_n)|\mathcal{F}_n]. \end{aligned}$$
Thus, $(H \cdot X)$ is a martingale, submartingale, or supermartingale according to what property X has.

Exercise 4.6. *1. Let $f : [0,1) \to \mathbb{R}$ be measurable with $\int_0^1 |f(x)|\, dx < \infty$. Let $\mathcal{F}_n = \sigma\{[(k-1)/2^n, k/2^n)$, where $k = 1, 2, \ldots, 2^n$. Let $Z_n = E[f|\mathcal{F}_n]$. Describe Z_n.*

2. Let τ be an $\mathcal{F}_n$-stopping time and define $H_n = I_{\{n \leq \tau\}}$. Show that H is predictable and describe $(H \cdot X)_n$.

4.2 Doob Decomposition

Theorem 4.7 (Doob Decomposition)**.** *Let X be an $\mathcal{F}_n$-submartingale. There exists unique random sequences M and V such that*

1. $X_n = M_n + V_n$ *for all* $n \geq 0$.
2. M *is an* $\mathcal{F}_n$*-martingale.*
3. $0 = V_0 \leq V_1 \leq \ldots$.
4. V *is* $\mathcal{F}_n$*-predictable.*

Proof. Define $U_k = E[X_k - X_{k-1}|\mathcal{F}_{k-1}]$ and

$$V_n = \sum_{k=1}^{n} U_k.$$

then 3 follows from the submartingale property of X and 4 follows from the basic properties of conditional expectation. Define M so that 1 is satisfied, then

$$\begin{aligned} E[M_n - M_{n-1}|\mathcal{F}_{n-1}] &= E[X_n - X_{n-1}|\mathcal{F}_{n-1}] - E[V_n - V_{n-1}|\mathcal{F}_{n-1}] \\ &= E[X_n - X_{n-1}|\mathcal{F}_{n-1}] - E[U_n|\mathcal{F}_{n-1}] \\ &= E[X_n - X_{n-1}|\mathcal{F}_{n-1}] - U_n = 0. \end{aligned}$$

Thus, there exists processes M and V that satisfy 1-4. To show uniqueness, choose at second pair $\tilde{M}$ and $\tilde{V}$. Then

$$M_n - \tilde{M}_n = \tilde{V}_n - V_n.$$

is $\mathcal{F}_{n-1}$-measurable. Therefore,

$$M_n - \tilde{M}_n = E[M_n - \tilde{M}_n|\mathcal{F}_{n-1}] = M_{n-1} - \tilde{M}_{n-1}$$

by the martingale property for M and $\tilde{M}$. Thus, the difference between M_n and $\tilde{M}_n$ is constant, independent of n. This difference

$$M_0 - \tilde{M}_0 = \tilde{V}_0 - V_0 = 0$$

by property 3. □

Exercise 4.8. *Let $B_m \in \mathcal{F}_m$ and define the process*

$$X_n = \sum_{m=0}^{n} I_{B_m}.$$

Find the Doob decomposition of X.

Theorem 4.9. *Each of the components in the Doob decomposition of a uniformly integrable submartingale is uniformly integrable.*

Proof. Use the Doob decomposition to write

$$X = M + V,$$

where M is an $\mathcal{F}_n$-martingale and V is an increasing $\mathcal{F}_n$-predictable process. Call

$$V_\infty = \lim_{n\to\infty} V_n.$$

Then, by the monotone convergence theorem,

$$EV_\infty = \lim_{n\to\infty} EV_n = \lim_{n\to\infty} E[X_n - M_n] = \lim_{n\to\infty} E[X_n - M_0] \le \sup_n E|X_n| + E|M_0| < \infty \tag{4.2}$$

the uniform integrability of X. Because the sequence V_n is dominated by an integrable random variable V_∞, it is uniformly integrable. $M = X - V$ is uniformly integrable because the difference of uniformly integrable sequences is uniformly integrable. □

4.3 Optional Sampling Theorem

Definition 4.10. *Let X be a random sequence of real valued random variables with finite absolute mean. A $\mathbb{N} \cup \{\infty\}$-valued random variable τ that is a.s. finite satisfies the* sampling integrability conditions *for X if the following conditions hold:*

1. *$E|X_\tau| < \infty$;*
2. *$\liminf_{n\to\infty} E[|X_n|; \{\tau > n\}] = 0$.*

Theorem 4.11 (Optional Sampling)**.** *Let $\tau_0 \le \tau_1 \le \tau_2 \cdots$ be an increasing sequence of $\mathcal{F}_n$-stopping times and let X be an $\mathcal{F}_n$-submartingale. Assume that for each n, τ_n is an almost surely finite random variable that satisfies the sampling integrability conditions for X. Then $\{X_{\tau_k}; k \ge 0\}$ is an $\mathcal{F}_{\tau_k}$-submartingale.*

Proof. $\{X_{\tau_n}; n \ge 0\}$ is $\mathcal{F}_{\tau_n}$-adapted and each random variable in the sequence has finite absolute mean. The theorem thus requires that we show that for every $A \in \mathcal{F}_{\tau_n}$

$$E[X_{\tau_{n+1}}; A] \ge E[X_{\tau_n}; A].$$

With this in mind, set

$$B_m = A \cap \{\tau_n = m\}.$$

These sets are pairwise disjoint and because τ_n is almost surely finite, their union is almost surely A. Thus, the theorem will follow from the dominated convergence theorem if we show

$$E[X_{\tau_{n+1}}; B_m] \ge E[X_{\tau_n}; B_m] = E[X_m; B_m].$$

Choose $p > m$. Then,

$$E[X_{\tau_{n+1}}; B_m] = E[X_m; B_m] + \sum_{\ell=m}^{p-1} E[(X_{\ell+1} - X_\ell); B_m \cap \{\tau_{n+1} > \ell\}] + E[(X_{\tau_{n+1}} - X_p); B_m \cap \{\tau_{n+1} > p\}].$$

Note that for each ℓ,

$$\{\tau_{n+1} > \ell\} = \{\tau_{n+1} \le \ell\}^c \in \mathcal{F}_\ell, \quad \text{and} \quad B_m \in \mathcal{F}_m \subset \mathcal{F}_\ell.$$

Consequently, by the submartingale property for X, each of the terms in the sum is non-negative.

To check that the last term can be made arbitrarily close to zero, Note that $\{\tau_{n+1} > p\} \to \emptyset$ a. s. as $p \to \infty$ and

$$|X_{\tau_{n+1}} I_{B_m \cap \{\tau_{n+1} > p\}}| \leq |X_{\tau_{n+1}}|,$$

which is integrable by the first property of the sampling integability condition. Thus by the dominated convergence theorem,

$$\lim_{p \to \infty} E[|X_{\tau_{n+1}}|; B_m \cap \{\tau_{n+1} > p\}] = 0.$$

For the remaining term, by property 2 of the sampling integribility condition, we are assured of a subsequence $\{p(k); k \geq 0\}$ so that

$$\lim_{k \to \infty} E[|X_{p(k)}|; \{\tau_{n+1} > p(k)\}] = 0.$$

□

A reveiw of the proof shows that we can reverse the inequality in the case of supermartingales and replace it with an equality in the case of martingales.

The most frequent use of this theorem is:

Corollary 4.12. *Let τ be a $\mathcal{F}_n$-stopping time and let X be an $\mathcal{F}_n$-(sub)martingale. Assume that τ is an almost surely finite random variable that satisfies the sampling integrability conditions for X. Then*

$$E[X_\tau] = (\geq) E[X_0].$$

Example 4.13. *Let X be a random sequence satisfying $X_0 = 1$ and*

$$P\{X_{n+1} = x | \mathcal{F}_n^X\} = \begin{cases} \frac{1}{2} & \text{if } x = 0, \\ \frac{1}{2} & \text{if } x = 2X_n, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, after each bet the fortune either doubles or vanishes. Let τ be the time that the fortune vanishes, i.e.

$$\tau = \min\{n > 0 : X_n = 0\}.$$

Note that $P\{\tau > k\} = 2^{-k}$. However,

$$EX_\tau = 0 < 1 = EX_0$$

and τ cannot satisfy the sampling integrability conditions.

Because the sampling integrability conditions are a technical relationship between X and τ, we look to find easy to verify sufficient conditions. The first of three is the case in which the stopping τ is best behaved - it is bounded.

Proposition 4.14. *If τ is bounded, then τ satifies the sampling integrability conditions for any sequence X of integrable random variables.*

Proof. Let b be a bound for τ. Then

1.
$$|X_\tau| = |\sum_{k=0}^{b} X_k I_{\{\tau=k\}}| \leq \sum_{k=0}^{b} |X_k|,$$
which is integrable and property 1 follows.

2. For $n > b$, $\{\tau > n\} = \emptyset$ and property 2 follows.

□

In the second case we look for a trade-off in the growth of the absolute value of conditional increments of the submartingale verses moments for the stopping time.

Exercise 4.15. *Show that if $0 \leq X_0 \leq X_1 \leq \cdots$, then 1 implies 2 in the sampling integrability conditions.*

Proposition 4.16. *Let X be an $\mathcal{F}_n$-adapted sequence of real-valued random variables and let τ be an almost surely finite $\mathcal{F}_n$-stopping time. Suppose for each $n > 0$, there exists m_n so that*
$$E[|X_k - X_{k-1}|\ |\mathcal{F}_{k-1}] \leq m_k \quad \textit{a.s. on the set } \{\tau \geq k\}.$$
Set $f(n) = \sum_{k=1}^{n} m_k$. If $Ef(\tau) < \infty$, then τ satisfies the sampling integrability conditions for X.

So, large values for the conditional mean of the absolute value of the step size forces the stopping time to have more stringent integrability conditions in order to satisfy the sufficient condition above.

Proof. For $n \geq 0$, define
$$Y_n = |X_0| + \sum_{k=1}^{n} |X_k - X_{k-1}|.$$
Then $Y_n \geq |X_n|$ and so it suffices to show that τ satisfies the sampling integrability conditions for Y. Because Y is an increasing sequence, we can use the exercise above to note that the theorem follows if we can prove that $EY_\tau < \infty$.

By the hypotheses, and the tower property, noting that $\{\tau \geq k\} = \{\tau \leq k-1\}^c$, we have
$$\begin{aligned} E[|X_k - X_{k-1}|; \{\tau \geq k\}] &= E[E[|X_k - X_{k-1}| I_{\{\tau \geq k\}} | \mathcal{F}_{k-1}]] \\ &= E[E[|X_k - X_{k-1}||\mathcal{F}_{k-1}] I_{\{\tau \geq k\}}] \leq m_k P\{\tau \geq k\} < \infty. \end{aligned}$$
Therefore,
$$EY_\tau = E|X_0| + \sum_{k=1}^{\infty} E[|X_k - X_{k-1}|; \{\tau \geq k\}] \leq \sum_{k=1}^{\infty} m_k P\{\tau \geq k\} = Ef(\tau).$$

□

Corollary 4.17. *1. If $E\tau < \infty$, and $|X_k - X_{k-1}| \leq c$, then τ satisfies the sampling integrability conditions for X.*

2. If $E\tau < \infty$, and S is a random walk whose steps have finite mean, then τ satisfies the sampling integrability conditions for S.

The last of the three cases show that if we have good behavior for the submartingale, then we only require a bounded stopping time.

Theorem 4.18. *Let X be a uniformly integrable $\mathcal{F}_n$-submartingale and τ an almost surely finite $\mathcal{F}_n$-stopping time. Then τ satisfies the sampling integrability conditions for X.*

Proof. Because $\{\tau > n\} \to \emptyset$, condition 2 follows from the fact that X is uniformly integrable.

By the optional sampling theorem, the stopped sequence $X_n^\tau = X_{\min\{n,\tau\}}$ is an $\mathcal{F}_n^\tau$-submartingale. If we show that X^τ is uniformly integrable, then, because $X_n^\tau \to X_\tau$ a.s., we will have condition 1 of the sampling integrability condition. Write $X = M + V$, the Doob decomposition for X, with M a martingale and V a predictable increasing process. Then both M and V are uniformly integrable sequences.

As we learned in (4.2), $V_\infty = \lim_{n\to\infty} V_n$ exists and is integrable. Because

$$V_n^\tau \leq V_\infty,$$

V^τ is a uniformly integrable sequence. To check that M^τ and consequently X^τ are uniformly integrable sequences, note that $\{|M_n|; n \geq 0\}$ is a submartingale and that $\{M_n^\tau \geq c\} = \{M_{\min\{\tau,n\}} \geq c\} \in \mathcal{F}_{\min\{\tau,n\}}$. Therefore, by the optional sampling theorem applied to $\{|M_n|; n \geq 0\}$,

$$E[|M_n^\tau|; \{|M_n^\tau| \geq c\}] \leq E[|M_n|; \{|M_{\min\{\tau,n\}}| \geq c\}]$$

and using the fact that M is uniformly integrable

$$P\{|M_{\min,\{\tau,n\}}| \geq c\} \leq \frac{E|M_{\min,\{\tau,n\}}|}{c} \leq \frac{E|M_n|}{c} \leq \frac{K}{c}$$

for some constant K. Therefore, by the uniform integrability of the sequence M,

$$\lim_{c\to\infty} \sup_n E[|M_n^\tau|; \{|M_n^\tau \geq c\}] \leq \lim_{c\to\infty} \sup_n E[|M_n|; \{|M_{\min,\{\tau,n\}} \geq c\}] = 0,$$

and therefore M^τ is a uniformly integrability sequence. □

Theorem 4.19. *(First Wald identity) Let S be a real valued random walk whose steps have finite mean μ. If τ is a stopping time that satisfies the sampling integrability conditions for S then*

$$ES_\tau = \mu E\tau.$$

Proof. The random process $Y_n = S_n - n\mu$ is a martingale. Thus, by the optional sampling theorem, applied using the bounded stopping time $\min\{\tau, n\}$,

$$0 = EY_{\min\{\tau,n\}} = ES_{\min\{\tau,n\}} - \mu E[\min\{\tau, n\}].$$

Note that,

$$\liminf_{n\to\infty} E[|S_{\min\{\tau,n\}} - S_\tau|] \leq \liminf_{n\to\infty} E[|S_n - S_\tau|; \{\tau > n\}] \leq \liminf_{n\to\infty} E[|S_n|; \{\tau > n\}] + \liminf_{n\to\infty} E[|S_\tau|; \{\tau > n\}].$$

The first term has liminf zero by the second sampling integrability condition. The second has limit zero. To see this, use the dominated convergence theorem, noting that the first sampling integrability condition states that $E|S_\tau| < \infty$. Consequently $S_{\min\{\tau,n\}} \to^{L^1} S_\tau$ as $n \to \infty$.

Let $\{n(k) : k \geq 1\}$ be a subsequence in which this limit is obtained. Then, by the monotone convergence theorem,

$$\mu E\tau = \mu \lim_{k\to\infty} E[\min\{\tau, n(k)\}] = \lim_{k\to\infty} ES_{\min\{\tau,n(k)\}} = ES_\tau.$$

□

Theorem 4.20. *(Second Wald Identity) Let S be a real-valued random walk whose step sizes have mean 0 and variance σ^2. If τ is a stopping time that satisfies the sampling integrability conditions for $\{S_n^2; n \geq 0\}$ then*

$$Var(S_\tau) = \sigma^2 E\tau.$$

Proof. Note that τ satisfies the sampling integrability conditions for S. Thus, by the first Wald idenity, $ES_\tau = 0$ and hence $\mathrm{Var}(S_\tau) = ES_\tau^2$. Consider $Z_n = S_n^2 - n\sigma^2$, an $\mathcal{F}_n^S$-martingale, and follow the steps in the proof of the first Wald identity. □

Example 4.21. *Let S be the asymmetric simple random work on $\mathbb{Z}$. Let p be the probability of a forward step. For any $a \in \mathbb{Z}$, set*

$$\tau_a = \min\{n \geq 0 : S_n = a\}.$$

For $a < 0 < b$, set $\tau = \min\{\tau_a, \tau_b\}$. The probability of $b - a + 1$ consecutive foward steps is p^{b-a+1}. Let

$$A_n = \{b{-}a{+}1 \textit{ consecutive foward steps beginning at time } 2n(b-a)+1\} \in \sigma\{X_{2n(b-a)+1}, \ldots X_{(2n+1)(b-a)+1}\}.$$

Because the A_n depend on disjoint steps, they are independent. In addition, $\sum_{n=1}^\infty P(A_n) = \infty$. Thus, by the second Borel-Cantelli lemma, $P\{A_n \text{ i.o.}\} = 1$. If A_n happens, then $\{\tau < 2(n+1)(b-a)\}$. In particular, τ is finite with probability one.

Consider Wald's martingale $Y_n = \exp(-\alpha S_n)L(\alpha)^{-n}$. Here,

$$L(\alpha) = e^{-\alpha}p + e^{\alpha}(1-p).$$

A simple martingale occurs when $L(\alpha) = 1$. Solving, we have the choices $\exp(-\alpha) = 1$ or $\exp(-\alpha) = (1-p)/p$. The first choice gives the constant martingale $Y_n = 1$, the second choice yields

$$Y_n = \left(\frac{1-p}{p}\right)^{S_n} = \psi(S_n).$$

Use the fact that

$$0 < Y_n \leq \max\left\{\left(\frac{1-p}{p}\right)^a, \left(\frac{1-p}{p}\right)^b\right\}$$

on the event $\{\tau > n\}$ to see that τ satisfies the sampling integrability conditions for Y. Therefore,

$$1 = EY_0 = EY_\tau = \psi(b)P\{S_\tau = b\} + \psi(a)P\{S_\tau = a\}$$

or

$$\psi(0) = \psi(b)(1 - P\{\tau_a < \tau_b\}) + \psi(a)P\{\tau_a < \tau_b\},$$
$$P\{\tau_a < \tau_b\} = \frac{\psi(b) - \psi(0)}{\psi(b) - \psi(a)}.$$

b

Suppose $p > 1/2$ and let $b \to \infty$ to obtain

$$P\{\min_n S_n \leq a\} = P\{\tau_a < \infty\} = \left(\frac{1-p}{p}\right)^{-a}. \tag{4.3}$$

In particular, $\min_n S_n$ *is integrable*

Let $a \to -\infty$ *to obtain*

$$P\{\tau_b < \infty\} = 1.$$

By Wald's first martingale identity,

$$ES_{\min\{\tau_b,n\}} = (2p-1)E[\min\{\tau_b,n\}].$$

Note that

$$|S_{\min\{\tau_b,n\}}| \leq \max\{b, |\min_m S_m|\},$$

is by (4.3) an integrable random variable. Thus, by the dominated convergence theorem,

$$\lim_{n\to\infty} ES_{\min\{\tau_b,n\}} = ES_{\tau_b} = b.$$

By the monotone convergence theorem,

$$\lim_{n\to\infty} E[\min\{\tau_b,n\}] = E\tau_b.$$

Therefore,

$$E\tau_b = \frac{b}{2p-1}.$$

Exercise 4.22. 1. *Let* S *denote a real-valued random walk whose steps have mean zero. Let*

$$\tau = \min\{n \geq 0 : S_n > 0\}.$$

Show that $E\tau = \infty$.

2. *For the simple symmetric random walk on* $\mathbb{Z}$*, show that*

$$P\{\tau_a < \tau_b\} = \frac{b}{b-a}. \tag{4.4}$$

3. *For the simple symmetric random walk on* $\mathbb{Z}$*, use the fact that* $S_n^2 - n$ *is a martingale to show that*

$$E[\min\{\tau_a,\tau_b\}] = -ab.$$

4. *For the simple symmetric random walk on* $\mathbb{Z}$*. Set*

$$(e^{-\alpha}p + e^{\alpha}(1-p)) = \frac{1}{z}.$$

to obtain

$$Ez^{\tau_1} = \frac{1-\sqrt{1-4p(1-p)z^2}}{2(1-p)z}.$$

4.4 Inequalities and Convergence

Lemma 4.23. *Let X be a submartingale. Then for each $x > 0$*

$$P\{\max_{0\le k\le n} X_k \ge x\} \le \frac{1}{x}E[X_n^+;\{\max_{0\le k\le n} X_n \ge x\}] \le \frac{1}{x}EX_n^+. \tag{4.5}$$

Proof. Let $\tau = \min\{k \ge 0 : X_k \ge x\}$ and note that X^+ is a submartingale. Then,

$$A = \{\max_{0\le k\le n} X_n \ge x\} = \{X_{\min\{\tau,n\}} \ge x\} \in \mathcal{F}_{\min\{\tau,n\}}.$$

Because $\min\{\tau, n\} \le n$, the optional sampling theorem gives, $X^+_{\min\{\tau,n\}} \le E[X_n^+|\mathcal{F}_{\min\{\tau,n\}}]$ and, therefore,

$$xP(A) \le E[X_{\min\{\tau,n\}}; A] \le E[X^+_{\min\{\tau,n\}}; A] \le E[X_n^+; A] \le E[X_n^+].$$

Now, divide by x to obtain (4.5). □

Corollary 4.24. *Let X be a martingale with $E|X_n|^p < \infty$, $p \ge 1$. Then for each $x > 0$,*

$$P\{\max_{0\le k\le n} |X_n| \ge x\} \le \frac{1}{x^p}E|X_n|^p.$$

The case that X is a random walk whose step size has mean zero and the choice $p = 2$ is known as *Kolmogorov's inequality.*

Theorem 4.25 (L^p maximal inequality)**.** *Let X be a nonnegative submartingale. Then for $p > 1$,*

$$E[\max_{1\le k\le n} X_k^p] \le \left(\frac{p}{p-1}\right)^p E(X_n)^p. \tag{4.6}$$

Proof. For $M > 0$, write $Z = \max\{X_k; 1 \le k \le n\}$ and $Z_M = \min\{Z, M\}$. Then

$$\begin{aligned} E[Z_M^p] &= \int_0^\infty \zeta^p \, dF_{Z_M}(\zeta) = \int_0^M p\zeta^{p-1}P\{Z > \zeta\}\, d\zeta \\ &= \int_0^M p\zeta^{p-1}P\{\max_{1\le k\le n} X_k > \zeta\}\, d\zeta \le \int_0^M p\zeta^{p-1}\frac{1}{\zeta}E[X_n;\{Z > \zeta\}]\, d\zeta \\ &= E[X_n \int_0^M p\zeta^{-2} I_{\{Z>\zeta\}}\, d\zeta] = E[X_n\beta(Z_M)] \end{aligned}$$

where

$$\beta(z) = \int_0^z p\zeta^{p-2}\, d\zeta = \frac{p}{p-1}z^{p-1}.$$

Note that (4.6) was used in giving the inequality. Now, use Hölder's inequality,

$$E[Z_M^p] \le \frac{p}{p-1}E[X_n^p]^{1/p}E[Z_M^p]^{(p-1)/p}.$$

and hence

$$E[Z_M^p]^{1/p} \le \frac{p}{p-1}E[X_n^p]^{1/p}.$$

Take p-th powers, let $M \to \infty$ and use the monotone convergence theorem. □

Example 4.26. *To show that no similar inequality holds for $p = 1$, let S be the simple symmetric random walk starting at 1 and let $\tau = \min\{n > 0 : S_n = 0\}$. Then $Y_n = S_{\min\{\tau,n\}}$ is a martingale.*

Let M be an integer greater than 1 the by equation (4.4),

$$P\{\max_n Y_n < M\} = P\{\tau_0 < \tau_M\} = \frac{M-1}{M}.$$

and

$$E[\max_n Y_n] = \sum_{M=1}^{\infty} P\{\max_n Y_n \geq M\} = \sum_{M=1}^{\infty} \frac{1}{M} = \infty.$$

Exercise 4.27. *For $p = 1$, we have the following maximum inequality. Let X be a nonnegative submartingale. Then*

$$E[\max_{1\leq k\leq n} X_k] \leq \left(\frac{e}{e-1}\right) E[X_n(\log X_n)^+].$$

Definition 4.28. *Let X be a real-valued process. For $a < b$, set*

$$\tau_1 = \min\{m \geq 0 : X_m \leq a\}.$$

For $k = 1, 2, \ldots,$ define

$$\sigma_k = \min\{m > \tau_k : X_k \geq b\}$$

and

$$\tau_{k+1} = \min\{m > \sigma_k ; X_k \leq a\}.$$

Define the number of upcrossings *of the interval $[a, b]$ up to time n by*

$$U_n(a, b) = \max\{k \leq n; \sigma_k \leq n\}.$$

Lemma 4.29 (Doob's Upcrossing Inequality)**.** *Let X be a submartingale. Fix $a < b$. Then for all $n > 0$,*

$$EU_n(a, b) \leq \frac{E(X_n - a)^+}{b - a}.$$

Proof. Define

$$C_m = \sum_{k=1}^{\infty} I_{\{\sigma_k < m \leq \tau_{k+1}\}}.$$

In other words, $C_m = 1$ between potential upcrossings and 0 otherwise. Note that $\{\sigma_k < m \leq \tau_{k+1}\} = \{\sigma_k \leq m-1\} \cap \{\tau_{k+1} \leq m-1\}^c \in \mathcal{F}_{m-1}$ and therefore C is a predictable process. Thus, $C \cdot X$ is a submartingale.

$$\begin{aligned}
0 \leq E(C\cdot X)_n &= E[\sum_{k=1}^{U_n(a,b)} (X_{\min\{\tau_{k+1},n\}} - X_{\min\{\sigma_k,n\}})] \\
&= E[-\sum_{k=2}^{U_n(a,b)} (X_{\min\{\sigma_k,n\}} - X_{\min\{\tau_k,n\}}) + X_{\min\{\tau_{U_n(a,b)+1},n\}} - (X_{\min\{\sigma_1,n\}} - a) - a] \\
&\leq E[-(b-a)U_n(a,b) + X_{\min\{\tau_{U_n(a,b)+1},n\}} - a] \\
&\leq E[-(b-a)U_n(a,b) + (X_n - a)^+]
\end{aligned}$$

which give the desired inequality. □

Corollary 4.30. *Let X be a submartingale. If*

$$\sup_n EX_n^+ < \infty,$$

then the total number of upcrossings of the interval $[a, b]$

$$U_n(a, b) = \lim_{n\to\infty} U_n(a, b)$$

has finite mean.

Proof. Use the monotone convergence theorem to conclude that

$$EU(a, b) \le \frac{|a| + EX_n^+}{b - a}.$$

□

Remark 4.31. $\lim_{n\to\infty} x_n$ *fails to exist in $[-\infty, \infty]$ if and only if for some $a, b \in \mathbb{Q}$*

$$\liminf_{n\to\infty} x_n \le a < b \le \limsup_{n\to\infty} x_n.$$

Theorem 4.32 (Submartingale convergence theorem)**.** *Let X be a submartingale. If*

$$\sup_n EX_n^+ < \infty,$$

then

$$\lim_{n\to\infty} X_n = X_\infty$$

exists almost surely with $E|X_\infty| < \infty$.

Proof. Choose $a < b$, then

$$\{U(a, b) = \infty\} = \{\liminf_{n\to\infty} X_n < a < b < \limsup_{n\to\infty} X_n\}.$$

Because $U(a, b)$ has finite mean, this event has probability 0 and thus

$$P\left(\bigcup_{a,b\in\mathbb{Q}} \{\liminf_{n\to\infty} X_n \le a < b \le \limsup_{n\to\infty} X_n\}\right) = 0.$$

Hence,

$$X_\infty = \lim_{n\to\infty} X_n \quad \text{exists a.s.}$$

This limit might be infinite. However, by Fatou's lemma,

$$EX_\infty^+ \le \liminf_{n\to\infty} EX_n^+ < \infty$$

eliminating the possibility of a limit of $+\infty$. Observe that because X is a submartingale

$$EX_n^- = EX_n^+ - EX_n \le EX_n^+ - EX_0.$$

Using Fatou's lemma a second time yields

$$EX_\infty^- \le \liminf_{n\to\infty} EX_n^+ - EX_0 \le \sup_n EX_n^+ - EX_0 < \infty$$

eliminating the possibility of a limit of $-\infty$. In addition, $E|X_\infty| = EX_\infty^+ + EX_\infty^- < \infty$. □

Corollary 4.33. *A nonnegative supermartingale converges almost surely.*

Proof. $Y_n = -X_n$ is a submartingale with $EY_n^+ = 0$. □

Note that $EX_0 \geq EX_n$ and so $EX_\infty < \infty$ by Fatou's lemma.

Theorem 4.34. *Let X be an $\mathcal{F}_n$-submartingale. Then X is uniformly integrability if and only if there exists a random variable X_∞ such that*

$$X_n \to^{L^1} X_\infty.$$

Furthermore, when this holds

$$X_n \to X_\infty \quad a.s.$$

Proof. If X is uniformly integrable, then $\sup_n E|X_n| < \infty$, and by the submartingale convergence theorem, there exists a random variable X_∞ with finite mean so that $X_n \to X_\infty$ a.s. Again, because X is uniformly integrable $X_n \to^{L^1} X_\infty$.

If $X_n \to^{L^1} X_\infty$, then X is uniformly integrable. Moreover, $\sup_n E|X_n| < \infty$, then by the submartingale convergence theorem

$$\lim_{n\to\infty} X_n$$

exists almost surely and must be X_∞. □

Theorem 4.35. *A sequence X is a uniformly integrable $\mathcal{F}_n$-martingale if and only if there exist a random variable X_∞ such that*

$$X_n = E[X_\infty|\mathcal{F}_n].$$

Proof. $X_n \to X_\infty$ a.s. and in L^1 by the theorem above. Let $A \in \mathcal{F}_n$ Then for $m > n$,

$$E[X_m; A] = E[X_n; A].$$

Letting $m \to \infty$, L^1-convergence implies that

$$E[X_\infty; A] = E[X_n; A].$$

Conversely, if such an $X_\infty \in L^1$ exists, then X is a martingale and the collection $\{E[X_\infty|\mathcal{F}_n] : n \geq 0\}$ is uniformly integrable. □

Exercise 4.36. *Let $\{Y_n; n \geq 1\}$ be a sequence of independent mean zero random varibles. Show that*

$$\sum_{n=1}^{\infty} EY_n^2 < \infty \quad \textit{implies} \quad \sum_{n=1}^{N} X_n \textit{ converges almost surely}$$

as $N \to \infty$.

Exercise 4.37 (Lebesgue points)**.** *Let m be Lebesgue measure on $[0,1)^n$. For $n \geq 0$ and $x \in [0,1)^n$, let $I(n,x)$ be the interval of the form $[(k_1-1)2^{-n}, k_12^{-n}) \times \cdots \times [(k_n-1)2^{-n}, k_n2^{-n})$ that contains x. Prove that for any measurable function $f : [0,1)^n \to \mathbb{R}$,*

$$\lim_{n\to\infty} \frac{1}{m(I(x,n))} \int_{I(x,n)} f \; dm = f(x)$$

for m-almost all x.

Example 4.38 (Polya's urn)**.** *Because positive martingales have a limit, we have for the martingale R defined by equation (4.1)*

$$R_\infty = \lim_{n\to\infty} R_n \quad a.s.$$

Because the sequence is bounded between 0 and 1, it is uniformly integrable and so the convergence is is also in L^1.

Start with b blue and g green, the probability that the first m draws are blue and the next $n-m$ draws are green is

$$\left(\frac{b}{b+g}\right)\left(\frac{b+c}{b+g+c}\right)\cdots\left(\frac{b+(m-1)c}{b+g+(m-1)c}\right)$$
$$\times\left(\frac{g}{b+g+mc}\right)\left(\frac{g+c}{b+g+(m+1)c}\right)\cdots\left(\frac{r+(m-n-1)c}{b+g+(n-1)c}\right).$$

This is the same probability for any other outcome that chooses m blue balls in the first n draws. Consequently, writing $a^{\bar{k}} = a(a+1)\cdots(a+k-1)$ for a to the *k* rising.

$$P\{X_n = b+mc\} = \binom{n}{m}\frac{(b/c)^{\bar{m}}(g/c)^{n\bar{-}m}}{((b+g)/c)^{\bar{n}}}.$$

For the special case $c=b=g=1$,

$$P\{X_n = 1+m\} = \frac{1}{n+1}$$

and R_∞ has a uniform distribution on $[0,1]$.

In general, R has a Beta$(g/c, b/c)$ distribution. The density for the distribution is

$$\frac{\Gamma((r+b)/c)}{\Gamma(b/c)\Gamma(g/c)}(1-r)^{b/c-1}r^{g/c-1}.$$

Exercise 4.39. *Show the beta distribution limit for Polya's urn.*

Example 4.40 (Radon-Nikodym theorem)**.** *Let ν be a finite measure and let μ be a probability measure on $(\Omega,\mathcal{F})$ with $\nu << \mu$. Choose a filtration $\{\mathcal{F}_n : n\ge 0\}$ so that $\mathcal{F} = \sigma\{\mathcal{F}_n : n \ge 0\}$. Write $\mathcal{F}_n = \sigma\{A_{n,1},\ldots,A_{n,k(n)}\}$ where $\{A_{n,1},\ldots,A_{n,k(n)}\}$ is a partition of Ω and define for $x\in A_{n,m}$,*

$$f_n(x) = \begin{cases} \nu(A_{n,m})/\mu(A_{n,m}) & \text{if } \mu(A_{n,m})>0 \\ 0 & \text{if } \mu(A_{n,m})=0 \end{cases}$$

Claim. *$\{f_n : n\ge 0\}$ is an $\mathcal{F}_n$-martingale on $(\Omega,\mathcal{F},\mu)$.*

Any $\tilde{A}\in\mathcal{F}_n$ is the finite disjoint union of set chosen from $\{A_{n,1},\ldots,A_{n,k(n)}\}$. Thus, we must show that

$$\int_{A_{n,j}} f_{n+1}\,d\mu = \int_{A_{n,j}} f_n\,d\mu.$$

This equality is immediate if $\mu(A_{n,k})=0$. Otherwise,

$$\int_{A_{n,j}} f_n\,d\mu = \frac{\nu(A_{n,j})}{\mu(A_{n,j})}\mu(A_{n,j}) = \nu(A_{n,j}).$$

Write

$$A_{n,j} = \bigcup_{k=1}^{\ell} \tilde{A}_k, \quad \tilde{A}_1, \ldots, \tilde{A}_\ell \in \mathcal{F}_{n+1}.$$

Then,

$$\int_{A_{n,j}} f_{n+1}\, d\mu = \sum_{k;\mu(\tilde{A}_k)>0} \frac{\nu(\tilde{A}_k)}{\mu(\tilde{A}_k)}\mu(\tilde{A}_k) = \sum_{k;\mu(\tilde{A}_k)>0)} \nu(\tilde{A}_k) = \nu(A_{n,j}).$$

Because f_n is a positive martingale, the martingale convergence theorem applies and

$$f(x) = \lim_{n\to\infty} f_n(x)$$

exists μ almost surely.

To check uniform integrability, note that $\nu << \mu$ implies, for every $\epsilon > 0$, there exists $\delta > 0$ so that

$$\mu(A) < \delta \quad \textit{implies} \quad \nu(A) < \epsilon.$$

Note that

$$\mu\{f_n > c\} \le \frac{\int f_n\, d\mu}{c} = \frac{\nu(\Omega)}{c} < \delta,$$

for sufficiently large c, and, therefore,

$$\int_{\{f_n > c\}} f_n\, d\mu = \nu\{f_n > c\} < \epsilon.$$

Therefore, the limit above is also in L^1 and

$$\nu(A) = \int_A f\, d\mu$$

for every $A \in \mathcal{S} = \cup_{n=1}^{\infty}\mathcal{F}_n$, and an algebra of sets with $\sigma(\mathcal{S}) = \mathcal{F}$. By the Sierpinski class theorem, the identity holds for all $A \in \mathcal{F}$. Uniqueness of f is an exercise.

Exercise 4.41. *Show that if $\tilde{f}$ satisfies*

$$\nu(A) = \int_A \tilde{f}\, d\mu$$

then $f = \tilde{f}$, μ-a.s.

Exercise 4.42. *If we drop the condition $\mu << \nu$, then show that f_n is an $\mathcal{F}_n$ supermartingale.*

Example 4.43 (Lévy's 0-1 law)**.** *Let $\mathcal{F}_\infty$ be the smallest σ-algebra that contains a filtration $\{\mathcal{F}_n; n \ge 0\}$, then the martingale convergence theorem gives us Lévy's 0-1 law:*

For any $A \in \mathcal{F}_\infty$,

$$P(A|\mathcal{F}_n) \to I_A \quad \textit{a.s. and in } L^1 \textit{ as } n \to \infty.$$

For a sequence X of random varibles, define the tail σ-field

$$\mathcal{T} = \bigcap_{n=0}^{\infty} \sigma(X_n, X_{n+1}, \ldots).$$

Elements in $\mathcal{T}$ are called tail events. *Some examples are*

$$\{\limsup_{n\to\infty} X_n > 0\}, \quad \text{and} \quad \{\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} X_n = \mu\}.$$

If the sequence is composed of independent random variables, and if $A \in \mathcal{T}$, then A is independent of $\mathcal{F}_n^X$. Consequently,

$$P(A|\mathcal{F}_n) = P(A)$$

and

$$P(A) = I_A \quad a.s.$$

Thus, $P(A) \in \{0, 1\}$, the Kolmogorov 0-1 law.

Exercise 4.44. *Call Y a* tail random variable *if it is measurable with respect to a the tail σ-algebra. If the tail σ-algebra is composed on independent random variable, then Y is almost surely a constant.*

Exercise 4.45. *Let M_n be a martingale with $EM_n^2 < \infty$ for all n. Then,*

$$E[(M_n - M_m)^2|\mathcal{F}_m] = E[M_n^2|\mathcal{F}_m] - M_m^2. \tag{4.7}$$

Example 4.46 (Branching Processes)**.** *Let $\{X_{n,k}; n \geq 1, k \geq 1\}$ be a doubly indexed set of independent and identically distributed $\mathbb{N}$-valued random variables. From this we can define a* branching *or* Galton-Watson-Bienaymé process *by $Z_0 = 1$ and*

$$Z_{n+1} = \begin{cases} X_{n,1} + \cdots + X_{m,Z_n}, & \text{if } Z_n > 0, \\ 0, & \text{if } Z_n = 0. \end{cases}$$

In words, each individual independently of the others, gives rise to offspring with a common offspring distribution,

$$\nu\{j\} = P\{X_{1,1} = j\}.$$

Z_n gives the population size of the n-th generation.

The mean μ of the offspring distribution is called the Malthusian parameter*. Define the filtration $\mathcal{F}_n = \sigma\{X_{m,k}; n \geq m \geq 1, k \geq 1\}$. Then*

$$\begin{aligned} E[Z_{n+1}|\mathcal{F}_n] &= \sum_{k=1}^{\infty} E[Z_{n+1}I_{\{Z_n=k\}}|\mathcal{F}_n] = \sum_{k=1}^{\infty} E[(X_{n+1,1} + \cdots + X_{n+1,k})I_{\{Z_n=k\}}|\mathcal{F}_n] \\ &= \sum_{k=1}^{\infty} I_{\{Z_n=k\}} E[(X_{n+1,1} + \cdots + X_{n+1,k})|\mathcal{F}_n] = \sum_{k=1}^{\infty} I_{\{Z_n=k\}} E[(X_{n+1,1} + \cdots + X_{n+1,k})] \\ &= \sum_{k=1}^{\infty} I_{\{Z_n=k\}} k\mu = \mu Z_n. \end{aligned}$$

Consequently, $M_n = \mu^{-n} Z_n$ is a martingale. Because it is nonnegative, it has a limit M_∞ almost surely as $n \to \infty$.

We will now look at three cases for branching process based on the value of the Malthusian parameter.

If $\mu < 1$, the the process is called subcritical. *In this case,*

$$P\{Z_n \neq 0\} = P\{Z_n \geq 1\} \leq E[Z_n; \{Z_n \geq 1\}] = EZ_n = \mu^n \to 0$$

as $n \to \infty$. Thus, $Z_n = 0$ for sufficiently large n. For such n, $M_n = 0$ and consequently $M_\infty = 0$ Thus, the martingale does not converge in L^1 and therefore cannot be uniformly integrable.

If $\mu = 1$, the the process is called critical. *In this case, Z_n itself is a martingale. Because Z_n is $\mathbb{N}$-valued and has a limit, it must be constant from some point on. If the process is not trivial ($\nu\{1\} < 1$), then the process continues to fluctuate unless the population is 0. Thus, the limit must be zero. $EZ_0 = 1 \neq 0 = EZ_\infty$ and again the martingale is not uniformly integrable.*

The case $\mu > 1$ is called the supercritical case. *We shall use G the generating function for the offspring distribution to analyze this case.*

Then G is the generating function for Z_1, $G_2 = G \circ G$ is the generating function for Z_2 and $G_n = G \circ G_{n-1}$ is the generating function for Z_n.

This is helpful in studying extinctions because

$$G_n(0) = P\{Z_n = 0\}.$$

Note that this sequence is monotonically increasing and set the probability of eventual extinction by

$$\rho = \lim_{n\to\infty} P\{Z_n = 0\} = \lim_{n\to\infty} G_n(0) = G\left(\lim_{n\to\infty} G_{n-1}(0)\right) = G(\rho).$$

Note that G is strictly convex and increasing. Consider the function $H(x) = G(x) - x$. Then,

1. *H is strictly convex.*
2. *$H(0) = G(0) - 0 \geq 0$ and is equal to zero if and only if $\nu\{0\} = 0$.*
3. *$H(1) = G(1) - 1 = 0$.*
4. *$H'(1) = G'(1) - 1 = \mu - 1 > 0$.*

Consequently, H is negative in a neighborhood of 1. H is non-negative at 0. If $\nu\{0\} \neq 0$, by the intermediate value theorem, H has a zero in $(0,1)$. By the strict convexity of H, this zero is unique and must be $\rho > 0$.

In the supercritical case, now add the assumption that the offspring distribution has variance σ^2. We will now try to develop a iterative formula for $a_n = E[M_n^2]$. To begin, use identity (4.7) above to write

$$E[M_n^2|\mathcal{F}_n] = M_{n-1}^2 + E[(M_n - M_{n-1})^2|\mathcal{F}_{n-1}].$$

For the second term,

$$\begin{aligned} E[(M_n - M_{n-1})^2|\mathcal{F}_{n-1}] &= E[(Z_n/\mu^n - Z_{n-1}/\mu^{n-1})^2|\mathcal{F}_{n-1}] = \mu^{-2n}E[(Z_n - \mu Z_{n-1})^2|\mathcal{F}_{n-1}] \\ &= \mu^{-2n}\sum_{k=1}^{\infty} E[(Z_n - \mu Z_{n-1})^2 I_{\{Z_{n-1}=k\}}|\mathcal{F}_{n-1}] \\ &= \mu^{-2n}\sum_{k=1}^{\infty} I_{\{Z_{n-1}=k\}} E[(\sum_{j=1}^{k} X_{n,j} - \mu k)^2|\mathcal{F}_{n-1}] \\ &= \mu^{-2n}\sum_{k=1}^{\infty} I_{\{Z_{n-1}=k\}} k\sigma^2 = \frac{\sigma^2}{\mu^{2n}} Z_{n-1} \end{aligned}$$

Consequently, by taking expectations, we find that

$$EM_n^2 = EM_{n-1}^2 + \frac{\sigma^2}{\mu^{2n}} EZ_{n-1} = EM_{n-1}^2 + \frac{\sigma^2}{\mu^{n+1}}$$

or

$$a_n = a_{n-1} + \frac{\sigma^2}{\mu^{n+1}}$$

with $a_0 = EM_0 = 1$. *Thus,*

$$a_n = 1 + \sigma^2 \sum_{k=1}^{n} \mu^{-(k+1)} = 1 + \sigma^2 \frac{\mu^n - 1}{\mu^{n+1}(\mu - 1)}.$$

In particular, $\sup_n EM_n^2 < \infty$. *Consequently,* M *is a uniformly integrable martingale and the convergence is also in* L^1. *Thus, there exists a mean one random variable* M_∞ *so that*

$$Z_n \approx \mu^n M_\infty$$

and $M_\infty = 0$ *with probability* ρ.

Exercise 4.47. *Let the offspring distribution of a mature individual have generating function* G_m *and let* p *be the probability that a immature individual survives to maturity.*

1. *Given* k immature *individuals, find the probability generating function for the number of number of* immature *individuals in the next generation of a branching process.*
2. *Given* k mature *individuals, find the probability generating function for the number of number of* mature *individuals in the next generation of a branching process.*

Exercise 4.48. *For a supercritical branching process,* Z_n *with extinction probability* ρ, ρ^{Z_n} *is a martingale.*

Exercise 4.49. *Let* Z *be a branching process whose offspring distribution has mean* μ. *Show that for* $n > m$, $E[Z_n Z_m] = \mu^{n-m} EZ_m^2$. *Use this to compute the correlation of* Z_m *and* Z_n

Exercise 4.50. *Consider a branching process whose offspring distribution satisfies* $\nu\{j\}j = 0$ *for* $j \geq 3$. *Find the probability of eventual extinction.*

Exercise 4.51. *Let the offspring distribution satisfy* $\nu\{j\} = (1-p)p^{j-1}$. *Find the generating function* Ez^{Z_n}. Hint: *Consider the case* $p = 1/2$ *and* $p \neq 1/2$ *separately. Use this to compute* $P\{Z_n = 0\}$ *and verify the limit satisfies* $G(\rho) = \rho$.

4.5 Backward Martingales

Definition 4.52. *Let $\mathcal{F}_0 \supset \mathcal{F}_1 \supset \cdots$ be a decreasing sequence of σ-algebras. Call a random sequence X a* backward *or* reverse martingale *if for each $n = 0, 1, \ldots,$*

1. *X_n is $\mathcal{F}_n$-measurable.*
2. *$E|X_n| < \infty$*
3. *$E[X_n|\mathcal{F}_{n+1}] = X_{n+1}$*

Exercise 4.53. *Let X be a backward martingale with respect to $\mathcal{F}_0 \supset \mathcal{F}_1 \supset \cdots$. Then for every n*

$$X_n = E[X_0|\mathcal{F}_n].$$

Example 4.54 (Random walks on $\mathbb{R}$)**.** *Let S be a random walk with steps $\{Y_n; n \geq 1\}$. Define the σ-algebras $\mathcal{F}_n = \sigma\{S_k; k \geq n\}$. Then*

$$S_{n+1} = E[S_{n+1}|S_{n+1}] = E[Y_1|S_{n+1}] + E[Y_2|S_{n+1}] + \cdots + E[Y_{n+1}|S_{n+1}].$$

By symmetry,

$$E[Y_k|S_{n+1}] = \frac{1}{n+1} S_{n+1}.$$

Note that $\sigma\{S_k; k \geq n\} = \sigma\{S_n, Y_k; k \geq n+1\}$. Consequently, because S_{n+1} and $\{Y_k; k \geq n+2\}$ are independent,

$$E[S_n|\mathcal{F}_{n+1}] = E[S_n|S_{n+1}, Y_k; k \geq n+2] = E[S_n|S_{n+1}] = \frac{n}{n+1} S_{n+1}$$

and

$$X_n = \frac{1}{n} S_n$$

is a backward martingale.

Theorem 4.55 (backward martingale convergence theorem)**.** *Let X be a backward martingale with respect to $\mathcal{F}_1 \supset \mathcal{F}_2 \supset \cdots$. Then there exists a random variable X_∞ such that*

$$X_n \to X_\infty$$

as $n \to \infty$ almost surely and in L^1.

Proof. Note that for any n, $X_n, X_{n-1}, \ldots, X_0$ is a martingale. Thus, the upcrossing inequality applies and we obtain

$$E[U_n(a,b)] \leq \frac{E[(X_0 - a)^+]}{b-a}.$$

Arguing as before, we can define

$$U(a,b) = \lim_{n\to\infty} U_n(a,b),$$

a random variable with finite expectation. Because it is finite almost surely, we can show that there exists a random variable X_∞ so that

$$X_n \to X_\infty$$

almost surely as $n \to \infty$.

Because $X_n = E[X_0|\mathcal{F}_n]$, we have that X is a uniformly integrable sequence and consequently, this limit is also in L^1. □

Corollary 4.56 (Strong Law of Large Numbers)**.** *Let $\{Y_j; j \geq 1\}$ be an sequence of independent and identically distributed random variables. Denote by μ, their common mean. Then,*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{\infty} Y_j = \mu \quad a.s.$$

Proof. We have that

$$X_n = \frac{1}{n}\sum_{j=1}^{n} Y_j$$

is a backward martingale. Thus, there there exists a random variable X_∞ so that $X_n \to X_\infty$ almost surely as $n \to \infty$.

Because X_∞ is a tail random variable from an independent set of random variables. Consequently, it is almost surely a constant. Because $X_n \to X_\infty$ in L^1 and $EX_n = \mu$, this constant must be μ. □

Example 4.57 (The Ballot Problem)**.** *Let $\{Y_k; k \geq 1\}$ be independent identically distributed $\mathbb{N}$-valued random variables. Set*

$$S_n = Y_1 + \cdots + Y_n.$$

Then,

$$P\{S_j < j \text{ for } 1 \leq j \leq n | S_n\} = \left(1 - \frac{S_n}{n}\right)^+. \tag{4.8}$$

To explain the name, consider the case in which the Y_k's take on the values 0 or 2 each with probability 1/2. In an election with two candidates A and B, interpret a 0 as a vote for A and a 2 as a vote for B. Then

$$\{A \text{ leads } B \text{ throughout the counting}\} = \{S_j < j \text{ for } 1 \leq j \leq n\} = G.$$

and the result above says that

$$P\{A \text{ leads } B \text{ throughout the counting} | A \text{ gets } r \text{ votes}\} = \left(1 - \frac{2r}{n}\right)^+.$$

As before, $X_n = S_n/n$ is a backward martingale with respect to $\mathcal{F}_n = \sigma\{S_k; k \geq n\}$.

To prove (4.8), first define

$$\tau = \begin{cases} \max\{k : S_k \geq k\} & \text{if this exists} \\ 1 & \text{otherwise} \end{cases}$$

Note that for $k > 1$

$$\{\tau = k\} = \{S_k \geq k, S_j < j \text{ for } j > k\} \in \mathcal{F}_k.$$

Similarly $\{\tau = 1\} \in \mathcal{F}_1$ and τ is a stopping time.

The result is immediate if $S_n \geq n$, so we shall assume that $S_n < n$.

Claim. $X_\tau = I_{G^c}$.

On G^c, $S_j \geq j$ for some j, $1 \leq j \leq n$. Consequently,

$$S_\tau \geq \tau, \quad \textit{and} \quad S_{\tau+1} < \tau + 1$$

and

$$0 \leq Y_{\tau+1} = S_{\tau+1} - S_\tau < (\tau + 1) - \tau = 1.$$

Thus, $Y_{\tau+1} = 0$, and

$$X_\tau = \frac{S_\tau}{\tau} = 1.$$

On G, $\tau = 1$ and $S_1 < 1$. Thus, $S_1 = 0$ and $X_\tau = 0$. .

Finally, by the optional sampling theorem,

$$P(G^c|S_n) = E[X_\tau|\mathcal{F}_n] = X_n = \frac{S_n}{n}$$

and

$$P(G|S_n) = 1 - \frac{S_n}{n}.$$

5 Markov Chains

5.1 Definition and Basic Properties

Definition 5.1. *A process X is called a* Markov chain *with values in a state space S if*

$$P\{X_{n+m} \in A|\mathcal{F}_n^X\} = P\{X_{n+m} \in A|X_n\}$$

for all $m, n \geq 0$ and $A \in \mathcal{B}(S)$.

In words, given $\mathcal{F}_n^X$, the entire history of the process up to time n, the only part that is useful in predicting the future is X_n, the position of the process at time n. In the case that S is a metric space with Borel σ-algebra $\mathcal{B}(S)$, this gives a mapping

$$S \times \mathbb{N} \times \mathcal{B}(S) \to [0, 1]$$

that gives the probability of making a transition from the state X_n at time n into a collection of states A at time $n + m$. If this mapping does not depend on the time n, call the Markov chain *time homogeneous*. We shall consider primarily time homogeneous chains.

Consider the case with $m = 1$ and time homogeneous transitions probabilities. We now ask that this mapping have reasonable measurability properties.

Definition 5.2. *1. A function*

$$\mu : S \times \mathcal{B}(S) \to [0, 1]$$

is a transition function *if*

(a) For every $x \in S$, $\mu(x, \cdot)$ is a probability measure, and

(b) For every $A \in \mathcal{B}(S)$, $\mu(\cdot, A)$ is a measurable function.

2. A transition function μ is called a transition function for a time homogeneous Markov chain X *if for every $n \geq 0$ and $A \in \mathcal{B}(S)$,*

$$P\{X_{n+1} \in A|\mathcal{F}_n^X\} = \mu(X_n, A).$$

Naturally associated to a measure is an operator that has this measure as its kernel.

Definition 5.3. *The* transition operator T *corresponding to μ is*

$$Tf(x) = \int_S f(y)\ \mu(x, dy)$$

for f, a bounded real-valued measurable function.

Exercise 5.4. *Use the standard machine to show that if T is the transition operator corresponding to the transition function μ for a Markov chain X, then for any bounded measurable function f,*

$$E[f(X_{n+1})|\mathcal{F}_n^X] = (Tf)(X_n) = \int_S f(y)\ \mu(X_n, dy).$$

The three expressions above give us three equivalent ways to think about Markov chain transitions.

With dynamical systems, the state of the next time step is determined by the position at the present time. No knowledge of previous positions of the system is needed. For Markov chains, we have a similar statement with state replaced by *distribution*. As with dynamical systems, along with the dynamics we need a statement about the initial position.

Definition 5.5. *The probability*

$$\alpha(A) = P\{X_0 \in A\}$$

is called the initial distribution *for the Markov chain.*

To see how multistep transitions can be computed, we use the tower property to compute

$$\begin{aligned} E[f_2(X_{n+2})f_1(X_{n+1})|\mathcal{F}_n^X] &= E[E[f_2(X_{n+2})f_1(X_{n+1})|\mathcal{F}_{n+1}^X]|\mathcal{F}_n^X] = E[E[f_2(X_{n+2})|\mathcal{F}_{n+1}^X]f_1(X_{n+1})|\mathcal{F}_n^X] \\ &= E[Tf_2(X_{n+1})f_1(X_{n+1})|\mathcal{F}_n^X] = (T((Tf_2)f_1))(X_n) \\ &= \int_S Tf_2(y)f_1(y)\ \mu(X_n, dy) = \int_S f_1(y)\mu(X_n, dy)\int_S f_2(z)\ \mu(y, dz). \end{aligned}$$

Continue this strategy to obtain

$$E[f_m(X_{n+m})\cdots f_1(X_{n+1})|\mathcal{F}_n^X] = (T(f_m\cdots T(f_2(Tf_1))))(X_n).$$

In particular, by taking $f_m = f$ and $f_j = 1$ for $j < m$, we obtain,

$$E[f(X_{n+m})|\mathcal{F}_n^X] = (T^m f)(X_n).$$

Now, consider the case $n = 0$ and $f_j = I_{A_j}$, then

$$P\{X_m \in A_m, \ldots, X_1 \in A_1|\mathcal{F}_0^X\} = \int_{A_1} \mu(X_0, dx_1)\int_{A_2} \mu(x_1, dx_2)\ldots\int_{A_m} \mu(x_{m-1}, dx_m),$$

and

$$P\{X_m \in A_m, \ldots, X_1 \in A_1, X_0 \in A_0\} = \int_{A_0} P\{X_m \in A_m, \ldots, X_1 \in A_1|X_0 = x_0\}\ \alpha(dx_0).$$

Therefore, the intial distribution and the transition function determine the finite dimensional distributions and consequently, by the Daniell-Kolmogorov extension theorem, the distribution of the process on the sequence space $S^{\mathbb{N}}$.

Typically we shall indicate the intial distribution for the Markov by writing P_α or E_α. If $\alpha = \delta_{x_0}$, then we shall write P_{x_0} or E_{x_0}. Thus, for example,

$$E_\alpha[f(X_n)] = \int_S T^n f(x)\ \alpha(dx).$$

If the state space S is countable, then we have

$$P\{X_{n+1} = y|X_n = x\} = \mu(x, \{y\}),$$

and

$$Tf(x) = \sum_{y \in S} f(y)\mu(x, \{y\}).$$

Consequently, the transition operator T can be viewed as a *transition matrix* with $T(x,y) = \mu(x,\{y\})$. The only requirements on such a matrix is that its entries are non-negative and for each $x \in S$ the row sum

$$\sum_{y \in S} T(x,y) = 1.$$

In this case the operator identity $T^{m+n} = T^m T^n$ gives the *Chapman-Kolmogorov equations*:

$$T^{m+n}(x,z) = \sum_{y \in S} T^m(x,y)T^n(y,z)$$

Exercise 5.6. *Suppose* $S = \{0,1\}$ *and*

$$T = \begin{pmatrix} 1 - t_{01} & t_{01} \\ t_{10} & 1 - t_{10} \end{pmatrix}.$$

Show that

$$P_\alpha\{X_n = 0\} = \frac{t_{10}}{t_{01} + t_{10}} + (1 - t_{01} - t_{10})^n \left(\alpha\{0\} - \frac{t_{10}}{t_{01} + t_{10}} \right).$$

5.2 Examples

1. (Independent Identically Distributed Sequences) Let ν be a probability measure and set $\mu(x,A) = \nu(A)$. Then,
$$P\{X_{n+1} \in A | \mathcal{F}_n^X\} = \nu(A).$$

2. (Random Walk on a Group) Let ν be a probability measure on a group G and define the transition function
$$\mu(x,A) = \nu(x^{-1}A).$$
In this case, $\{Y_k; k \geq 1\}$ are independent G-valued random variables with distribution ν. Set
$$X_{n+1} = Y_1 \cdots Y_{n+1} = X_n Y_{n+1}.$$
Then
$$P\{Y_{n+1} \in B | \mathcal{F}_n^X\} = \nu(B),$$
and
$$P\{X_{n+1} \in A | \mathcal{F}_n^X\} = P\{X_n^{-1} X_{n+1} \in X_n^{-1} A | \mathcal{F}_n^X\} = P\{Y_n \in X_n^{-1} A | \mathcal{F}_n^X\} = \nu(X_n^{-1} A).$$
In the case of $G = \mathbb{R}^n$ under addition, X is a random walk.

3. (Shuffling) If the G is S_N, the symmetric group on n letters, then ν is a shuffle.

4. (Simple Random Walk) $S = \mathbb{Z}$, $T(x,x+1) = p, T(x,x-1) = q, p + q = 1$.

5. (Simple Random Walk with Absorption) $S = \{x \in \mathbb{Z} : -M_1 \leq j \leq M_2\}$ For $-M_1 < j < M_2$, the transition probabilites are the same as in the simple random walk. At the endpoints,
$$T(-M_1, -M_1) = T(M_2, M_2) = 1.$$

6. (Simple Random Walk with Reflection) This is the same as the simple random walk with absorption except for the endpoints. In this case,
$$T(-M_1, -M_1 + 1) = T(M_2, M_2 - 1) = 1.$$

7. (Renewal sequences) $S = \mathbb{N} \cup \{0, \infty\}$. Let R be a distribution on $S \backslash \{0\}$. Consider a Markov chain Y with transition function.
$$\mu(y, \cdot) = \begin{cases} R & \text{if } y = 1 \\ \delta_{y-1} & \text{if } 1 < y < \infty \\ \delta_\infty & \text{if } y = \infty \end{cases}$$
Then $X_n = I_{\{Y_n = 1\}}$ is a renewal sequence with waiting distribution R.

8. (Random Walk on a Graph) A *graph* $G = (V, E)$ consist of a *vertex set* V and an *edge set* $E \subset \{\{x, y\}; x, y \in V, x \neq y\}\}$, the unordered pairs of vertices. The *degree* of vertex element x, $\deg(x)$, is the number of neighbors of x, i. e., the vertices y satisfying $\{x, y\} \in E$. The transition function, $\mu(x, \cdot)$ is a probability measure supported on the neighbors of x. If the degree of every vertex is finite and
$$\mu(x, \{y\}) = \begin{cases} \frac{1}{\deg(x)} & \text{if } x \text{ is a neighbor of } y, \\ 0 & \text{otherwise,} \end{cases}$$
then X is a *simple random walk on the graph*, G.

9. (Birth-Death Processes) $S = \mathbb{N}$
$$\mu(x, \{x - 1, x, x + 1\}) = 1, \ x > 0, \quad \mu(0, \{0, 1\}) = 1.$$
Sometimes we write
$$\lambda_x = T(x, x + 1) \quad \text{for the probability of a birth,}$$
and
$$\mu_x = T(x, x - 1) \quad \text{for the probability of a death.}$$

10. (Branching processes) $S = \mathbb{N}$. Let ν be a distribution on $\mathbb{N}$ and let $\{Z_i : i \geq 0\}$ be independent random variables with distribution ν. Set the transition matrix
$$T(x, y) = P\{Z_1 + \cdots + Z_x = y\} = \nu^{*x}\{y\}.$$
The probability measure ν is called the *offspring* or *branching distribution.*

11. (Polya urn model) $S = \mathbb{N} \times \mathbb{N}$. Let $c \in \mathbb{N}$.
$$T((x, y), (x + c, y)) = \frac{x}{x + y}, \qquad T((x, y), (x, y + c)) = \frac{y}{x + y}.$$

12. (Ehrenfest Urn Model) $S = \{x \in \mathbb{Z} : 0 \le j \le 2N\}$. This is a birth death model with

$$\lambda_x = \frac{2N - x}{2N} \quad \text{and} \quad \mu_x = \frac{x}{2N}.$$

Consider two urns with a total of $2N$ balls. This transitions come about by choosing a ball at random and moving it to the other urn. The Markov chain X is the number or balls in the first urn.

13. (Wright-Fisher model) $S = \{x \in \mathbb{N} : 0 \le x \le 2N\}$.

$$T(x, y) = \binom{2N}{y} \left(\frac{x}{2N}\right)^y \left(\frac{2N - x}{2N}\right)^{2N-y}.$$

Wright-Fisher model, the fundamental model of population genetics, has three features:

(a) Non-overlapping generations.
(b) Random mating.
(c) Fixed population size.

To see how this model comes about, consider a genetic locus, that is the sites on the genome that make up a gene. This locus has two alleles, A and a. Alleles are versions of genetic information that are encoded at the locus. Our population consists of N diploid individuals and consequently $2N$ copies of the locus. The state of the process x is the number of A alleles. Consider placing x of the A alleles and thus $2N - x$ of the a alleles in an urn. Now draw out $2N$ with replacement, then $T(x, y)$ is the probability that y of the A alleles are drawn.

The state of the process is the number of alleles of type A in the population.

14. (Wright-Fisher Model with Selection) Let $s > 0$, then

$$T(x, y) = \binom{2N}{y} \left(\frac{(1+s)x}{2N + sx}\right)^y \left(\frac{2N - x}{2N + sx}\right)^{2N-y}.$$

This gives $1 + s : 1$ odds that A is selected.

15. (Wright-Fisher Model with Mutation) Let $u > 0$ and $v > 0$.

$$T(x, y) = \binom{2N}{y} p^y q^{2N-y}$$

where

$$p = \frac{x}{2N}(1 - u) + (1 - \frac{x}{2N})v. \quad q = \frac{x}{2N}u + (1 - \frac{x}{2N})(1 - v).$$

This assume that mutation $A \to a$ happens with probability u and the mutation $a \to A$ happens with probability v. Thus, for example, the probability that an a allele is selected is proprotional to the expected number of a present *after mutation.*

Exercise 5.7. *Define the following* queue. *In one time unit, the individual is served with probability p. During that time unit, individuals arrive to the queue according to an arrival distribution γ. Give the transition probabilities for this Markov chain.*

Exercise 5.8. *The discrete time two allele* Moran model *is similar to the Wright-Fisher model. To describe the transitions, choose an individual at random to be removed from the population. Choose a second individual to replace that individual. Mutation and selection proceed as with the Wright-Fisher model. Give the transition probabilities for this Markov chain.*

5.3 Extensions of the Markov Property

Definition 5.9. *For the sequence space $S^{\mathbb{N}}$, define the* shift operator

$$\theta : S^{\mathbb{N}} \to S^{\mathbb{N}}$$

by

$$\theta(x_0, x_1, x_2, \ldots) = (x_1, x_2, x_3, \ldots),$$

i.e.

$$(\theta x)_k = x_{k+1}.$$

Theorem 5.10. *Let X be a time homogeneous Markov process on $\Omega = S^{\mathbb{N}}$ with the Borel σ-algebra. and let $Z : \Omega \to \mathbb{R}$ be bounded and measurable, then*

$$E_\alpha[Z \circ \theta^n | \mathcal{F}_n^X] = \phi(X_n)$$

where

$$\phi(x) = E_x Z.$$

Proof. Take $Z = \prod_{k=1}^m I_{\{x_k \in A_k\}}$ where $A_k \in \mathcal{B}(S)$. Then

$$E_\alpha[Z \circ \theta^n | \mathcal{F}_n^X] = P\{X_{n+1} \in A_1, \ldots X_{n+m} \in A_m | \mathcal{F}_n^X\}$$

and, by the time homogeneity of the process,

$$\phi(x) = P\{X_n \in A_1, \ldots X_m \in A_m | X_0 = x\} = P\{X_{n+1} \in A_1, \ldots X_{n+m} \in A_m | X_n = x\}.$$

Thus, the formula states that

$$P\{X_{n+1} \in A_1, \ldots X_{n+m} \in A_m | \mathcal{F}_n^X\} = P\{X_{n+1} \in A_1, \ldots X_{n+m} \in A_m | X_n\}$$

and, thus, by the Markov property, holds for finite products of indicator functions. Now use the Daniell-Kolmogorov extension theorem and the standard machine to complete the proof. □

Example 5.11. *If $Z = \max\{f(X_k); n_1 \leq k \leq n_2\}$ then*

$$Z \circ \theta^n = \max\{f(X_{k+n}); n_1 \leq k \leq n_2\} = \max\{f(X_k); n_1 + n \leq k \leq n + n_2\}.$$

In addition,

$$\phi(x) = E_x[\max\{f(X_k); n_1 \leq k \leq n_2\}] \text{ and } E[\max\{f(X_k); n_1 + n \leq k \leq n + n_2\} | \mathcal{F}_n^X] = \phi(X_n).$$

Exercise 5.12. *Let $A \in \sigma\{X_0, \ldots, X_n\}$ and $B \in \sigma\{X_{n+1}, X_{n+2}, \ldots\}$, then*

$$P_\alpha(A \cap B | X_n) = P_\alpha(A | X_n) P_\alpha(B | X_n).$$

In other words, the past and the future are independent, given the present.

Hint: *Write the left hand side as $E_\alpha[E_\alpha[I_A I_B | \mathcal{F}_n^X] | X_n]$.*

Theorem 5.13 (Strong Markov Property)**.** *Let Z be a bounded random sequence and let τ be a stopping time. Then*

$$E_\alpha[Z_\tau \circ \theta^\tau|\mathcal{F}_\tau] = \phi_\tau(X_\tau) \quad \textit{on } \{\tau < \infty\}$$

where

$$\phi_n(x) = E_x Z_n.$$

Proof. Let $A \in \mathcal{F}_\tau^X$. Then

$$\begin{aligned}
E_\alpha[Z_\tau \circ \theta^\tau; A \cap \{\tau < \infty\}] &= \sum_{n=0}^{\infty} E_\alpha[Z_\tau \circ \theta^\tau; A \cap \{\tau = n\}] \\
&= \sum_{n=0}^{\infty} E_\alpha[Z_n \circ \theta^n; A \cap \{\tau = n\}] \\
&= \sum_{n=0}^{\infty} E_\alpha[\phi_n(X_n); A \cap \{\tau = n\}] \\
&= E_\alpha[\phi(X_\tau); A \cap \{\tau < \infty\}].
\end{aligned}$$

In particular, for $f : S \to \mathbb{R}$, bounded and measurable, then on the set $\{\tau < \infty\}$,

$$E[f(X_{\tau+n})|\mathcal{F}_\tau^X] = \phi(X_\tau)$$

where $\phi(x) = E_x[f(X_n)]$. □

Exercise 5.14.

1. *Let X be a Markov chain with respect to the filtration $\mathcal{F}_n$ on S with transition operator T and let $f : S \to \mathbb{R}$ be bounded and measurable. Then, the sequence*

$$M_n(f) = f(X_n) - \sum_{k=0}^{n-1}(T - I)f(X_k)$$

in an $\mathcal{F}_n$-martingale.

2. *Call a measurable function $h : S \to \mathbb{R}$* (super)harmonic *(on A) if*

$$Th(x) = (\leq)h(x) \quad (\textit{for all } x \in A).$$

Show that if h is (super)harmonic then $h(X_n)$ is a (super)martingale.

Let $B \subset S$ and define

$$h_B(x) = P_x\{X_n \in B \text{ for some } n \geq 0\}, \quad h_B^+(x) = P_x\{X_n \in B \text{ for some } n > 0\},$$

If $x \in B$, then $h_B(x) = 1$. If $x \in B^c$, then $h_B(x) = h_B^+(x)$. If we set

$$Z = I_{\{X_n \in B \text{ for some } n \geq 0\}}, \quad Z^+ = I_{\{X_n \in B \text{ for some } n > 0\}}$$

then,

$$Z^+ = Z \circ \theta.$$

By the Markov property

$$h_B^+(x) = E_x Z^+ = E_x[E_x[Z^+|\mathcal{F}_1]] = E_x[E_x[Z \circ \theta|\mathcal{F}_1]] = E_x[h_B(X_1)] = Th_B(x).$$

Consequently, h_B is harmonic on B^c. Also,

$$Th_B(x) = h_B^+(x) \leq h_B(x)$$

and h_B is superharmonic. Finally, if

$$\tau_B = \min\{n \geq 0 : X_n \in B\},$$

then

$$h_B(X_{\min\{n,\tau_B\}}) = h_B^+(X_{\min\{n,\tau_B\}})$$

and $h_B(X_{\min\{n,\tau_B\}})$ is a martingale.

This leads us to a characterization of h_B.

Theorem 5.15. *Let X be a Markov sequence and let $B \in \mathcal{B}(S)$. Then h_B is the minimal bounded function among those functions h satisfying*

1. *$h = 1$ on B.*
2. *h is harmonic on B^c.*
3. *$h \geq 0$.*

Proof. h_B satisfies 1-3. Let h be a second such function. Then $Y_n = h(X_{\min\{n,\tau_B\}})$ is a nonnegative bounded martingale. Thus, for some random variable Y_∞, $Y_n \to Y_\infty$ a.s. and in L^1. If $n > \tau_B$, then $Y_n = 1$. Therefore, $Y_\infty = 1$ whenever $\tau_B < \infty$ and

$$h(x) = E_x Y_0 = E_x Y_\infty \geq P_x\{\tau_B < \infty\} = h_B(x).$$

□

Exercise 5.16.
1. *Let $B = \mathbb{Z}\backslash\{-M_1 + 1, \ldots, M_2 - 1\}$. Find the harmonic function that satisfies 1-3 for the symmetric and asymmetric random walk.*
2. *Let $B = \{0\}$ and $p > 1/2$ and find the harmonic function that satisfies 1-3 for the asymmetric random walk.*

Example 5.17. *Let G_ν denote the generating function for the offspring distribution of a branching process X and consider $h_0(x)$, the probability that the branching process goes extinct given that the initial population is x. We can view the branching process as x independent branching processes each with initial population 1. Thus, we are looking for the probability that all of these x processes die out, i.e.,*

$$h_0(x) = h_0(1)^x.$$

Because h_0 is harmonic at 1,

$$h_0(1) = Th_0(1) = \sum_y T(1,y)h_0(y) = \sum_{y=0}^{\infty} \nu\{y\}h_0(1)^y = G_Z(h_0(1)).$$

Thus, $h_0(1)$ is a solution to

$$z = G_Z(z).$$

By the theorem above, it is the minimal solution.

Note $G_\nu(1) = 1$ and therefore 1 is a solution to the equation. The case $G_\nu(z) = z$ is the case of exact replacement. As long as $\nu\{2\} > 0$, G_Z is a strictly convex increasing function and the existence of a second solution $c < 1$ will occur if and only if the mean offspring distribution $\mu_Z = EZ = G'(1) > 1$. If $G_Z(0) = \nu\{0\} = 0$, then the branching process can never go extinct. If $\nu\{0\} > 0$, second solution is greater than zero.

Note that even if $\mu = 1$, the branching process goes extinct with probability 1 as long as $\nu\{1\} \neq 1$. Check that $E[X_{n+1}|\mathcal{F}_n] = \mu X_n$. Thus, $\mu^{-n}X_n$ is a martingale. For $\mu = 1$, $X_n \to 0$ a.s. and so X cannot be uniformly integrable.

Exercise 5.18. *Give the extinction probabilities of a branching process with*

1. $\nu = p\delta_0 + (1-p)\delta_2$.
2. $\nu\{z\} = p(1-p)^z$.

5.4 Classification of States

We will focus on the situation in which the state space S is countable.

Definition 5.19. *A state y is* accessible *from a state x (written $x \to y$) if for some $n \geq 0$, $T^n(x,y) > 0$. If $x \to y$ and $y \to x$, then we write $x \leftrightarrow y$ and say that x and y* communicate.

Exercise 5.20. 1. *$\leftrightarrow$ is an equivalence relation. Thus the equivalence of communication partitions the state space by its equivalence classes.*

2. *Consider the 3 and 4 state Markov chains with transitions matrices*

$$\begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/4 & 1/4 \\ 0 & 1/3 & 2/3 \end{pmatrix}, \quad \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Give the equivalence classes under communication for the Markov chain associated to these two transtions matrices.

Definition 5.21. 1. *Call a set of states C* closed *if for all $x \in C$ and all $n \geq 0$, $P\{X_n \in C\} = 1$.*

2. *If $T(x,x) = 1$, then x is called an* absorbing state. *In particular, $\{x\}$ is a closed set of states.*
3. *A Markov chain is called* irreducible *if all states communicate.*

Proposition 5.22. *For a time homogenous Markov chain X, and for $y \in S$, Then, under the measure P_α, the process $Y_n = I_{\{y\}}(X_n)$ is a delayed renewal sequence. Under P_y, Y is a renewal sequence.*

Proof. Let r and s be positive integers and choose a sequence $(\epsilon_1, \ldots, \epsilon_{r+s}) \in \{0,1\}^{(r+s)}$ such that $\epsilon_r = 1$ Then, for $x \in S$

$$\begin{aligned} P_x\{Y_n = \epsilon_n \text{ for } 0 \le n \le r+s\} &= E_x[P\{Y_n = \epsilon_n \text{ for } 0 \le n \le r+s|\mathcal{F}_r\}] \\ &= E_x[I_{\epsilon_1}(Y_1)\cdots I_{\epsilon_r}(Y_r)P_x\{Y_n = \epsilon_n \text{ for } r+1 \le n \le r+s|\mathcal{F}_r\}] \\ &= E_x[I_{\epsilon_1}(Y_1)\cdots I_{\epsilon_r}(Y_r)]P_y\{Y_n = \epsilon_n \text{ for } 1 \le n \le s\} \\ &= P_x\{Y_n = \epsilon_n \text{ for } 1 \le n \le r\}P_y\{Y_n = \epsilon_n \text{ for } 1 \le n \le s\} \end{aligned}$$

Now, integrate with respect to the initial distribution α. □

For the renewal sequence, we have that

$$u_n = P_y\{Y_n = 1\} = P_y\{X_n = y\} = T^n(y,y)$$

is the potential sequence.

The renewal times $\tau_y^0 = 0$,

$$\tau_y^k = \min\{n > \tau_y^{k-1} : X_n = y\}$$

are the stopping times for the k-th visit to state y after time 0.

The following definitions match the classification of a state for the P_y-renewal process $Y_n = I_{\{y\}}(X_n)$.

Definition 5.23. *Set $\tau_y = \min\{n > 0 : X_n = y\}$. The state y is*

1. recurrent *if $P_y\{\tau_y < \infty\} = 1$,*
2. transient *if $P_y\{\tau_y = \infty\} > 0$,*
3. positive recurrent *if y is recurrent and $E_y\tau_y < \infty$,*
4. null recurrent *if y is recurrent and $E_x\tau = \infty$.*

   The period *of a state y, $\ell(y)$ if τ_y is distributed on the lattice $L_{\ell(y)}$ given $X_0 = y$. If $T^n(y,y) = 0$ for all n define $\ell(y) = 0$.*

   The state y is

5. periodic *if $\ell(y) > 2$,*
6. aperiodic *if $\ell(y) = 1$,*
7. ergodic *if it is recurrent and aperiodic.*

The symmetric simple random walk, all states are null recurrent with period 2. The asymmetric simple random walk, all states are transient with period 2.

Proposition 5.24. *Communicating states have the same period.*

Proof. If x and y communicate that we can pick m and n so that

$$T^n(x,y) > 0 \quad \text{and} \quad T^m(y,x) > 0,$$

and, by the Chapman-Kolmogorov equation,

$$T^{m+n}(y,y) \geq T^m(y,x)T^n(x,y) > 0$$

and $\ell(y)$ divides $m+n$. Now, for all $k \geq 0$,

$$\{X_m = x, X_{m+k} = x, X_{m+k+n} = y\} \subset \{X_{m+k+n} = y\}.$$

take the probability using initial distribution δ_y to obtain

$$T^m(y,x)T^k(x,x)T^n(x,y) \leq T^{m+k+n}(y,y). \tag{5.1}$$

If k and hence $m+k+n$ is not a multiple of $\ell(y)$, then $T^{m+k+n}(y,y) = 0$. Because $T^m(y,x)T^n(x,y) > 0$, $T^k(x,x) = 0$. Conversely, if $T^k(x,x) \neq 0$, then k must be a multiple of $\ell(y)$. Thus, $\ell(x)$ is a multiple of $\ell(y)$.

Reverse the roles of x and y to conclude that $\ell(y)$ is a multiple of $\ell(x)$. This can happen when and only when $\ell(x) = \ell(y)$. □

Denote

$$\rho_{xy} = P_x\{\tau_y < \infty\},$$

then by appealing to the results in renewal theory,

$$y \text{ is recurrent} \quad \text{if and only if} \quad \rho_{yy} = 1 \quad \text{if and only if} \quad \sum_{n=0}^{\infty} T^n(y,y) = \infty.$$

Let N_y be the number of visits to state y. Then the sum above is just $E_y N_y$. If

$$\rho_{yy} < 1 \quad \text{if and only if} \quad \sum_{n=0}^{\infty} T^n(y,y) = \frac{\rho_{yy}}{1-\rho_{yy}}$$

because N_y is a geometric random variable.

Exercise 5.25. *1. $P_x\{\tau_y^k < \infty\} = \rho_{xy}\rho_{yy}^{k-1}$.*

2. If $\rho_{yy} < 1$, then $E_x N_y = \rho_{xy}/(1-\rho_{yy})$.

Proposition 5.26. *If x is recurrent and $x \to y$, then $\rho_{yx} = 1$ and $y \to x$.*

Proof. x is recurrent, so $\rho_{xx} = 1$. Once the Markov chain reaches y, the probability of never returning to x

$$P_y\{X_{\tau_y+k} \neq x \text{ for all } n \geq 1 | X_{\tau_y} = y\} = P_y\{X_k \neq x \text{ for all } n \geq 1\} = 1 - \rho_{yx}$$

by the strong Markov property. Thus, the probability of never returning to x

$$1 - \rho_{xx} \geq \rho_{xy}(1-\rho_{yx}) \geq 0.$$

Because $1-\rho_{xx} = 0$ and $\rho_{xy} > 0$, $\rho_{yx} = 1$ and $y \to x$. □

Proposition 5.27. *From a recurrent state, only recurrent states can be reached.*

Proof. Assume x is recurrent and select y so that $x \to y$. By the previous proposition $y \to x$. As we argued to show that x and y have the same period, pick m and n so that

$$T^n(x,y) > 0 \quad \text{and} \quad T^m(y,x) > 0,$$

and so

$$T^m(y,x)T^k(x,x)T^n(x,y) \le T^{m+k+n}(y,y).$$

Therefore,

$$\sum_{k=0}^{\infty} T^k(y,y) \ge \sum_{k=0}^{\infty} T^{m+k+n}(y,y) \ge \sum_{k=0}^{\infty} T^m(y,x)T^k(x,x)T^n(x,y) \ge T^n(x,y)T^m(y,x)\sum_{k=0}^{\infty} T^k(x,x).$$

Because x is recurrent, the last sum is infinite. Because $T^n(x,y)T^m(y,x) > 0$, the first sum is infinite and y is recurrent. □

Corollary 5.28. *If for some y, $x \to y$ but $y \not\to x$ then x cannot be recurrent.*

Consequently, the state space S has a partition $\{T, R_i, i \in \mathcal{I}\}$ into T, the transient states and R_i, a closed irreducible set of recurrent states.

Remark 5.29.

1. *All states in the interior are transient for a simple random walk with absorption. The endpoints are absorbing. The same statement holds for the Wright-Fisher model.*
2. *In a branching process, if $\nu\{0\} > 0$, then $x \to 0$ but $0 \not\to x$ and so all states except 0 are transient. The state 0 is absorbing.*

Theorem 5.30. *If C is a finite close set, then C contains a recurrent state. If C is irreducible, then all states in C are recurrent.*

Proof. By the theorem, the first statement implies the second. Choose $x \in C$. If all states are transient, then $E_xN_y < \infty$. Therefore,

$$\infty > \sum_{y\in C} E_xN_y = \sum_{y\in C}\sum_{n=1}^{\infty} T^n(x,y) = \sum_{n=1}^{\infty}\sum_{y\in C} T^n(x,y) = \sum_{n=1}^{\infty} 1 = \infty.$$

This contradiction proves the theorem. □

Theorem 5.31. *If X is irreducible and recurrent, then the only bounded harmonic functions on S are constants.*

Proof. Let h be bounded and harmonic. By the martingale convergence theorem, the martingale $h(X_n)$ converges almost surely. Let $\epsilon > 0$ and choose N so that $|h(X_n) - h(X_m)| < \epsilon$ for $n, m > N$. Because X is recurrent, we can choose for any $x, y \in S$, n and m arbitrarily large with probability one, so that $X_n = x$, and $X_m = y$. Therefore,

$$|h(x) - h(y)| < \epsilon$$

and h is constant. □

Example 5.32 (Birth and death chains). *The harmonic functions h satisfy*

$$h(x) = \lambda_x h(x+1) + \mu_x h(x-1) + (1 - \lambda_x - \mu_x)h(x),$$

$$\lambda_x(h(x+1) - h(x)) = \mu_x(h(x) - h(x-1)),$$

$$h(x+1) - h(x) = \frac{\mu_x}{\lambda_x}(h(x) - h(x-1)),$$

If $\lambda_y > 0$ and $\mu_y > 0$ for all $y \in S$, then the chain is irreducible and

$$h(x+1) - h(x) = \prod_{y=1}^{x} \frac{\mu_y}{\lambda_y}(h(1) - h(0)),$$

Now, if h is harmonic, then any linear function of h is harmonic. Thus, for simplicity, take $h(0) = 0$ and $h(1) = 1$ to obtain

$$h(x) = \sum_{z=0}^{x-1} \prod_{y=1}^{z} \frac{\mu_y}{\lambda_y}.$$

Note that the terms in the sum for h are positive. Thus, the limit as $z \to \infty$ exists. If

$$h(\infty) = \sum_{z=0}^{\infty} \prod_{y=1}^{z} \frac{\mu_y}{\lambda_y} < \infty,$$

then h is a bounded harmonic function and X is transient. This will always occur, for example, if

$$\liminf_{x \to \infty} \frac{\mu_x}{\lambda_x} > 1.$$

and births are more likely than deaths.

Let $a < x < b$ and define $\tau_a = \min\{n > 0; X_n = a\}$, $\tau_b = \min\{n > 0; X_n = b\}$ and $\tau = \min\{\tau_a, \tau_b\}$. Then $\{h(X_{\min\{n,\tau\}}); n \geq 1\}$ is a bounded martingale. This martingale has a limit, which must be $h(a)$ or $h(b)$. In either case, $\tau < \infty$ a.s. and

$$h(x) = Eh(X_\tau) = h(a)P_x\{X_\tau = a\} + h(b)P_x\{X_\tau = b\} = h(a)P_x\{\tau_a < \tau_b\} + h(b)P_x\{\tau_b < \tau_a\}$$

or

$$P_x\{\tau_a < \tau_b\} = \frac{h(b) - h(x)}{h(b) - h(a)}.$$

If the chain is transient, let $b \to \infty$ to obtain

$$P_x\{\tau_a < \infty\} = \frac{h(\infty) - h(x)}{h(\infty) - h(a)}.$$

Exercise 5.33. *Apply the results above to the asymmetric random walk with $p < 1/2$.*

To obtain a sufficient condition for recurrence we have the following.

Theorem 5.34. *Suppose that S is irreducible and suppose that off a finite set F, h is a positive superharmonic function for the transition function T associated with X. If, for any $M > 0$, $h^{-1}([0, M])$ is a finite set, then X is recurrent.*

Proof. It suffices to show that one state is recurrent. Define the stopping times

$$\tau_F = \min\{n > 0 : X_n \in F\}, \quad \tau_M = \min\{n > 0 : h(X_n) > M\}, \quad M = 1, 2, \ldots.$$

Because the chain is irreducible and $h^{-1}([0, M])$ is a finite set, τ_M is finite almost surely. Note that the optional sampling theorem implies that $\{h(X_{\min\{n,\tau_F\}}); n \geq 0\}$ is a supermartingale. Consequently,

$$h(x) = E_x h(X_0) \geq E_x h(X_{\min\{\tau_M,\tau_F\}}) \geq M P_x\{\tau_M < \tau_F\}$$

because h is positive and $h(X_{\min\{\tau_M,\tau_F\}}) \geq M$ on the set $\{\tau_M < \tau_F\}$. Now, let $M \to \infty$ to see that $P_x\{\tau_F = \infty\} = 0$ for all $x \in S$. Because $P_x\{\tau_F < \infty\} = 1$, whenever the Markov chain leaves F, it returns almost surely, and thus,

$$P_x\{X_n \in F \text{ i.o.}\} = 1 \text{ for all } x \in S,$$

Because F is finite, for some $\tilde{x} \in F$

$$P_{\tilde{x}}\{X_n = \tilde{x} \text{ i.o.}\} = 1$$

i.e., $\tilde{x}$ is recurrent. □

Remark 5.35. *By adding a constant to h, the condition that h is bounded below is sufficient to prove the theorem above.*

Remark 5.36. *Returning to irreducible birth and death processes, we see that $h(\infty) = \infty$ is a necessary and sufficient condition for the process to be recurrent.*

To determine the *mean exit times*, consider the solutions to

$$Tf(x) = f(x) - 1.$$

Then

$$M_n(f) = f(X_n) - \sum_{k=0}^{n-1} (T - I) f(X_k) = f(X_n) + n$$

is a martingale provided that $f(X_n)$ is integrable. Let $\tau = \min\{n \geq 0; X_n = a \text{ or } b\}$ be a stopping time satisfying the sampling integrablility coonditions for $M(f)$, then

$$f(x) = E_x M(f)_0 = E_x M(f)_\tau = E_x f(X_\tau) + E_x \tau.$$

For the stopping time above, if we solve $Tf(x) = f(x) - 1$ with $f(a) = f(b) = 0$ then

$$f(x) = E_x \tau.$$

Exercise 5.37. 1. *For a simple random walk on $\mathbb{Z}$, let $a < x < b$ and set*

$$\tau = \min\{n \geq 0; X_n = a \text{ or } b\}.$$

Find $E_x \tau$.

2. *For a birth and death process, write*

$$\rho_n = \frac{\lambda_1 \lambda_2 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n}$$

Show that absorption to the state zero has mean time

$$\begin{array}{ll} \infty & \text{if } \sum_{n=1}^{\infty} \rho_n = \infty \\ \sum_{n=1}^{\infty} \rho_n + \sum_{n=1}^{m-1} \left(\prod_{k=1}^{n} \frac{\mu_k}{\lambda_k}\right) \sum_{j=n+1}^{\infty} \rho_j & \text{if } \sum_{n=1}^{\infty} \rho_n < \infty \end{array}$$

5.5 Recurrent Chains

5.5.1 Stationary Measures

Definition 5.38. *A measure π for a transition measure μ is called* stationary *if*

$$\int_S \mu(x, A)\pi(dx) = \pi(A).$$

Note that we are *not* requiring that π be a *probability* measure.

For a countable state space and $A = \{y\}$, this becomes

$$\pi\{y\} = \sum_{x \in S} \mu(x, \{y\})\pi\{x\} = \sum_{x \in S} \pi\{x\} T(x, y),$$

or in matrix form

$$\pi = \pi T.$$

To explain the term stationary measure, let μ is the transition operator for a Markov chain X with initial distribution α, then

$$\int_S \mu(x, A)\alpha(dx) = \int_S P_x\{X_1 \in A\}\alpha(dx) = P_\alpha\{X_1 \in A\}.$$

Thus, if π is a probability measure and is the initial distribution, then the identity above becomes

$$P_\pi\{X_1 \in A\} = P_\pi\{X_0 \in A\}.$$

Call a Markov chain *stationary* if X_n has the same distribution for all n

Example 5.39. 1. *Call a transition matrix* doubly stochastic *if*

$$\sum_{x \in S} T(x, y) = 1.$$

For this case, $\pi\{x\}$ constant is a stationary distribution. The converse is easily seen to hold.

2. *(Ehrenfest Urn Model) Recall that $S = \{x \in \mathbb{Z} : 0 \le j \le 2N\}$.*

$$T(x, x+1) = \frac{2N - x}{2N} \quad \text{and} \quad T(x, x-1) = \frac{x}{2N}.$$

For the stationary distribution,

$$\pi\{x+1\}T(x+1,x)+\pi\{x-1\}T(x-1,x)=\pi\{x\}$$

$$\pi\{x+1\}\frac{x+1}{2N}+\pi\{x-1\}\frac{2N-x-1}{2N}=\pi\{x\}$$

or

$$\pi\{x+1\}=\frac{2N}{x+1}\pi\{x\}-\frac{2N-x-1}{x+1}\pi\{x-1\}.$$

Check that this is a recursion relation for the binomial coefficients

$$\binom{2N}{x}.$$

Because the sum of the binomial coefficients is 2^{2N}, we can give a stationary probability measure

$$\pi\{x\}=2^{-2N}\binom{2N}{x}.$$

Place $2N$ coins randomly down. Then $\pi\{x\}$ is the probability that x are heads. The chain can be realized by taking a coin at random and flipping it and keeping track of the number of heads. The computation shows that π is stationary for this process.

If a Markov chain is in its stationary distribution π, then Bayes' formula allows us to look at the process in reverse time.

$$\tilde{T}(x,y)=P_\pi\{X_n=y|X_{n+1}=x\}=\frac{P_\pi\{X_{n+1}=x|X_n=y\}P_\pi\{X_n=y\}}{P_\pi\{X_{n+1}=x\}}=\frac{\pi\{y\}}{\pi\{x\}}T(y,x).$$

Definition 5.40. *The Markov chain $\tilde{X}$ in reverse time is called the* dual Markov process *and $\tilde{T}$ is called the* dual transition matrix. *If $T=\tilde{T}$, then the Markov chain is called* reversible *and the stationary distribution satisfies* detailed balance

$$\pi\{y\}T(y,x)=\pi\{x\}T(x,y). \tag{5.2}$$

Sum this equation over y to see that it is a stronger condition for π than stationarity.

Exercise 5.41. *If π satisfies detailed balance for a transition matrix T, then*

$$\pi\{y\}T^n(y,x)=\pi\{x\}T^n(x,y).$$

Example 5.42 (Birth and death chains)**.** *Here we look for a measure that satisfies detailed balance*

$$\pi\{x\}\mu_x=\pi\{x-1\}\lambda_{x-1},$$

$$\pi\{x\}=\pi\{x-1\}\frac{\lambda_{x-1}}{\mu_x}=\pi\{0\}\prod_{y=1}^{x}\frac{\lambda_{y-1}}{\mu_y}.$$

This can be normalized to give a stationary probability distribution provided that

$$\sum_{x=0}^{\infty}\prod_{y=1}^{x}\frac{\lambda_{y-1}}{\mu_y}<\infty.$$

Here we take the empty product to be equal 1.

To decide which Markov chains are reversible, we have the following cycle condition due to Kolmogorov.

Theorem 5.43. *Let T be the transition matrix for an irreducible Markov chain X. A necessary and sufficient condition for the eixstence of a reversible measure is*

1. *$T(x,y) > 0$ if and only if $T(y,x) > 0$.*
2. *For any loop $\{x_0, x_1, \dots x_n = x_0\}$ with $\prod_{i=1}^n T(x_i, x_{i-1}) > 0$,*

$$\prod_{i=1}^n \frac{T(x_{i-1}, x_i)}{T(x_i, x_{i-1})} = 1.$$

Proof. Assume that π is a reversible measure. Because $x \leftrightarrow y$, there exists n so that $T^n(x,y) > 0$ and $T^n(y,x) > 0$. Use

$$\pi\{y\}T^n(y,x) = \pi\{x\}T^n(x,y).$$

to see that any reversible measure must give positive probability to each state. Now return to the defining equation for reversible measures to se that this proves 1.

For 2, note that for a reversible measure

$$\prod_{i=1}^n \frac{T(x_{i-1}, x_i)}{T(x_i, x_{i-1})} = \prod_{i=1}^n \frac{\pi\{x_i\}}{\pi\{x_{i-1}\}}.$$

Because $\pi\{x_n\} = \pi\{x_0\}$, each factor $\pi\{x_i\}$ appears in both the numerator and the demominator, so the product is 1.

To prove sufficiently, choose $x_0 \in S$ and set $\pi\{x_0\} = \pi_0$. Because X is irreducible, for each $y \in S$, there exists a sequence $x_0, x_1, \dots, x_n = y$ so that

$$\prod_{i=1}^n T(x_{i-1}, x_i) > 0.$$

Define

$$\pi\{y\} = \prod_{i=1}^n \frac{T(x_{i-1}, x_i)}{T(x_i, x_{i-1})}\pi_0.$$

The cycle condition guarantees us that this definition is independent of the sequence. Now add the point x to the path to obtain

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

and thus π is reversible. □

Exercise 5.44. *Show that an irreducible birth-death process satisfies the Kolmogorov cycle condition.*

If we begin the Markov chain at a recurrent state x, we can divide the evolution of the chain into the *excursions.*

$$(X_1, \dots, X_{\tau_x^1}), \quad (X_{\tau_x^1+1}, \dots, X_{\tau_x^2}), \quad \cdots .$$

The average activities of these excursions ought to be to visit the states y proportional to the value of the stationary distibution. This is the content of the following theorem.

Theorem 5.45. *Let x be a recurrent state. Then*

$$\mu_x\{y\} = E_x\left[\sum_{n=0}^{\tau_x-1} I_{\{X_n=y\}}\right] = \sum_{n=0}^{\infty} P_x\{X_n = y, \tau_x > n\} \tag{5.3}$$

defines a stationary distribution.

Proof.

$$\begin{aligned}
(\mu_x T)(z) &= \sum_{y\in S} \mu_x\{y\}T(y,z) \\
&= \sum_{y\in S}\sum_{n=0}^{\infty} P_x\{X_n = y, \tau_x > n\}P_x\{X_{n+1} = z|X_n = y\} \\
&= \sum_{n=0}^{\infty}\sum_{y\in S} P_x\{X_n = y, X_{n+1} = z, \tau_x > n\}
\end{aligned}$$

If $z = x$ this last double sum becomes

$$\sum_{n=0}^{\infty} P_x\{\tau_x = n+1\} = 1 = \mu_x\{x\}.$$

because $P_x\{\tau_x = 0\} = 0$.

If $z \neq x$ this sum becomes

$$\sum_{n=0}^{\infty} P_x\{X_{n+1} = z, \tau_x > n\} = \mu_x\{z\}$$

because $P_x\{X_0 = z, \tau_x > 0\} = 0$. □

Note that the total mass of the stationary measure μ_x is

$$\sum_{y\in S} \mu_x\{y\} = \sum_{y\in S}\sum_{n=0}^{\infty} P_x\{X_n = y, \tau_x > n\} = \sum_{n=0}^{\infty} P_x\{\tau_x > n\} = E_x\tau_x.$$

Remark 5.46. *If $\mu_x\{y\} = 0$, then the individual terms in the sum (5.3), $P_x\{X_n = y, \tau_x > n\} = 0$. Thus, y is not accessible from x on the first x-excursion. By the strong Markov property, y is cannot be accessible from x. Stated in the contrapositive, $x \to y$ implies $\mu_x\{y\} > 0$*

Theorem 5.47. *Let T be the transition matrix for a irreducible and recurrent Markov chain. Then the stationary measure is unique up to a constant multiple.*

Proof. Let ν be a stationary measure and choose $x_0 \in S$. Then

$$\begin{aligned}
\nu\{z\} &= \sum_{y\in S} \nu\{y\}T(y,z) \\
&= \nu\{x_0\}T(x_0,z) + \sum_{y\neq x_0} \nu\{y\}T(y,z) \\
&= \nu\{x_0\}P_{x_0}\{X_1 = z\} + P_\nu\{X_0 \neq x_0, X_1 = z\} \\
&= \nu\{x_0\}T(x_0,z) + \sum_{y\neq x_0} \nu\{x_0\}T(x_0,y)T(y,z) + \sum_{x\neq x_0}\sum_{y\neq x_0} \nu\{x\}T(x,y)T(y,z) \\
&= \nu\{x_0\}P_{x_0}\{X_1 = x_0\} + \nu\{x_0\}P_{x_0}\{X_1 \neq x_0, X_2 = z\} + P_\nu\{X_2 = z, \tau_{x_0} > 1\} \\
&= \cdots \\
&= \nu\{x_0\}\sum_{k=1}^{n} P_{x_0}\{X_k = z, \tau_{x_0} > k-1\} + P_\nu\{X_n = z, \tau_{x_0} > n-1\}
\end{aligned}$$

Let $n \to \infty$ and note that the last term is positive to obtain

$$\nu\{z\} \geq \nu\{x_0\}\mu_{x_0}(z)$$

Noting that the sum above is from 1 to τ_{x_0} rather than 0 to $\tau_{x_0} - 1$. To obtain equality in this equation note that

$$\nu\{x_0\} = \sum_{z\in S} \nu\{z\}T^n(z,x_0) \geq \nu\{x_0\}\sum_{z\in S} \mu_{x_0}\{z\}T^n(z,x_0) = \nu\{x_0\}\mu_{x_0}\{x_0\} = \nu\{x_0\}.$$

Thus, we must have

$$\nu\{z\} = \nu\{x_0\}\mu_{x_0}(z) \text{ whenever } T^n(z,x_0) > 0.$$

However, $z \to x_0$. So this must hold for some n. □

Theorem 5.48. *If there is a stationary probability distribution, π, then all states y that have $\pi\{y\} > 0$ are recurrent.*

Proof. Recall that, if the number of number of visits to state y is N_y, then

$$E_x N_y = \sum_{n=1}^{\infty} T^n(x,y) = \frac{\rho_{xy}}{1-\rho_{yy}}.$$

Use the fact that $\pi T^n = \pi$ and Fubini's theorem to conclude that

$$\sum_{x\in S} \pi\{x\}\sum_{n=1}^{\infty} T^n(x,y) = \sum_{n=1}^{\infty} \pi\{y\} = \infty.$$

Therefore,

$$\infty = \sum_{x\in S} \pi\{x\}\frac{\rho_{xy}}{1-\rho_{yy}} \leq \frac{1}{1-\rho_{yy}}.$$

Thus, $\rho_{yy} = 1$. □

Theorem 5.49. *If T is the transition matrix for an irreducible Markov chain and if π is a stationary probability distribution, then*

$$\pi\{x\} = \frac{1}{E_x\tau_x}.$$

Proof. Because the chain is irreducible $\pi\{x\} > 0$. Because the stationary measure is unique,

$$\pi\{x\} = \frac{\mu_x\{x\}}{E_x\tau_x} = \frac{1}{E_x\tau_x}.$$

□

In summary, if a state x is positive recureent, then, $\pi = \mu_x/E_x\tau_x$ is a stationary probability measure. If $x \leftrightarrow y$ then $\pi\{y\} > 0$ and y is positive recurrent. In particular, positive recurrent states communicate with only positive recurrent states.

Example 5.50. *(Ehrenfest urn model)* $S = \{0, 1, \cdots, 2N\}$

$$E_{N+x}\tau_{N+x} = \frac{1}{\pi\{N+x\}} = \frac{2^{2N}}{\binom{2N}{N+x}}.$$

Recall Stirling's approximation

$$n! = e^{-n}n^n\sqrt{2\pi n}(1 + o(n)).$$

Then

$$\begin{aligned}\binom{2N}{N+x} &= \frac{(2N)!}{(N+x)!(N-x)!} \approx \frac{e^{-2N}(2N)^{2N}\sqrt{4\pi N}}{e^{N+x}(N+x)^{N+x}\sqrt{2\pi(N+x)}e^{N-x}(N-x)^{N-x}\sqrt{2\pi(N-x)}} \\ &\approx \frac{(2N)^{2N}\sqrt{N}}{(N+x)^{N+x}(N-x)^{N-x}\sqrt{\pi(N+x)(N-x)}}.\end{aligned}$$

Thus,

$$E_{N+x}\tau_{N+x} \approx \frac{(N+x)^{N+x}(N-x)^{N-x}\sqrt{\pi(N+x)(N-x)}}{(N)^{2N}\sqrt{N}}$$

and

$$E_N\tau_N \approx \sqrt{\pi N}.$$

The ratio,

$$\begin{aligned}\frac{E_N\tau_N}{E_{N+x}\tau_{N+x}} &\approx (1+\frac{x}{N})^{N+x+1/2}(1-\frac{x}{N})^{N-x+1/2} \\ &\approx \exp(x(1+\frac{x+1/2}{N}))\exp(-x(1-\frac{x-1/2}{N})) \\ &\approx \exp(x\frac{x+1/2}{N})\exp(x\frac{x-1/2}{N}) = \exp(\frac{2x^2}{N}).\end{aligned}$$

Now assume a small but macroscopic volume - say $N = 10^{20}$ molecules. Let $x = 10^{16}$ denote a 0.01% deviation from equilibrium. Then,

$$\exp(\frac{2x^2}{N}) = \exp(\frac{2\times 10^{32}}{10^{20}}) = \exp(2\times 10^{12}).$$

So, the two halves will pass through equibrium about $\exp(2 \times 10^{12})$ *times before the molecules in about the same time that the molecules become 0.01% removed from equilibrium once.*

5.5.2 Asymptotic Behavior

We know from the renewal theorem that for a state x with period $\ell(x)$,

$$\lim_{n\to\infty} T^{n\ell(x)}(x,x) = \frac{\ell(x)}{E_x\tau_x}.$$

Now, by the strong Markov property,

$$T^n(x,y) = \sum_{k=1}^{n} P_x\{\tau_y = k\}T^{n-k}(y,y).$$

Thus, by the dominated convergence theorem, we have for the aperiodic case

$$\lim_{n\to\infty} T^n(x,y) = \lim_{n\to\infty} \sum_{k=1}^{n} P_x\{\tau_y = k\}T^{n-k}(y,y) \sum_{k=1}^{\infty} P_x\{\tau_y = k\}\frac{\ell(y)}{E_y\tau_y} = \rho_{xy}\frac{1}{E_y\tau_y}.$$

This computation yields the following result.

Theorem 5.51. *If* T *is the transition matrix for an ergodic Markov chain with stationary distribution* π*, then*

$$\lim_{n\to\infty} T^n(x,y) = \frac{1}{E_y\tau_y} = \pi\{y\}.$$

Lemma 5.52. *Let* $Z_1, Z_2, \ldots$ *be independent and indentically distributed positive random variables with mean* μ*. Then*

$$\lim_{m\to\infty} \frac{Z_m}{m} = 0 \quad \textit{almost surely.}$$

Proof. Let $\epsilon > 0$. Then

$$\sum_{m=1}^{\infty} P\{Z_m > \epsilon m\} = \sum_{m=1}^{\infty} P\{Z_1 > \epsilon m\} \le \frac{\mu}{\epsilon} < \infty.$$

Thus, by the first Borel-Cantelli lemma,

$$P\{\frac{1}{m}Z_m > \epsilon \text{ i.o.}\} = 0.$$

□

Theorem 5.53. *(Ergodic theorem for Markov chains) Assume* X *is an ergodic Markov chain and that* $\int_S |f(y)|\ \mu_x(dy)$*, then for any initial distribution* α*,*

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \to^{a.s.} \int_S f(y)\ \pi(dy).$$

Proof. Choose m so that $\tau_x^m \leq n < \tau_x^{m+1}$ and write

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) = \frac{1}{n}\sum_{k=1}^{\tau_x^1-1} f(X_k) + \frac{1}{n}\sum_{k=\tau_x^1}^{\tau_x^m-1} f(X_k) + \frac{1}{n}\sum_{k=\tau_x^m}^{n} f(X_k)$$

In the first sum, the random variable is fixed and so as $n \to \infty$, the term has zero limit with probability one.

For $k \geq 1$, we have the independent and identically sequence of random variables

$$V(f)_k = f(X_{\tau_x^k}) + \cdots + f(X_{\tau_x^{k+1}-1}).$$

Claim I.

$$E_x V(f)_1 = \int_S f(y)\ \mu_x(dy) = (\int_S f(y)\ \pi(dy))/E_x\tau_x.$$

If $f = I_{\{z\}}$, then, by definition, $E_xV(f)_1 = \mu_x\{z\}$. Now, use the standard machine.

Claim II. The last sum has limit zero with probability one.

This sum is bounded above by $V(|f|)_m$. By the claim above, this random variable has a finite mean, namely $(\int_S |f(y)|\ \pi(dy))/E_x\tau_x$. Now, use the lemma above and the fact that $m \leq n$.

For the center sum, write

$$\frac{1}{n}\sum_{k=\tau_x^1}^{\tau_x^m-1} f(X_j) = \frac{m}{n}\frac{1}{m}\sum_{k=1}^{m-1} V(f)_k$$

.

Claim III.

$$\frac{n}{m} \to E_x\tau_x^1 \quad \text{a.s.}$$

$$\frac{n}{m} = \frac{n-\tau_x^m}{m} + \frac{\tau_x^m}{m}$$

Note that that $\{\tau_x^{m+1} - \tau_x^m : m \geq 1\}$ is an idependent and identically sequence with mean $E_x\tau_x^1$. Thus,

$$\frac{\tau_x^m}{m} \to E_x\tau_x^1 \quad \text{a.s.}$$

by the strong law of large numbers and

$$0 \leq \frac{n-\tau_x^m}{m} \leq \frac{\tau_x^{m+1}-\tau_x^m}{m}.$$

which converges to zero a.s. by the lemma.

In summary,

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) - \frac{E_x\tau_x}{m}\sum_{k=1}^{m-1} V(f)_k \to 0 \quad \text{a.s.}$$

Now use the strong law of large numbers to see that

$$\frac{1}{m}\sum_{k=1}^{m-1} V(f)_k \to^{\text{a.s.}} \frac{1}{E_x\tau_x}\int_S f(y)\ \pi(dy).$$

□

The ergodic theorem corresponds to a strong law. Here is the corresponding central limit theorem. We begin with a lemma.

Lemma 5.54. *Let U be a positive random variable $p > 0$, and $EU^p < \infty$, then*

$$\lim_{n\to\infty} n^p P\{U > n\} = 0.$$

Proof.

$$n^p P\{U > n\} = n^p \int_n^\infty dF_U(u) \le \int_n^\infty u^p\ dF_U(u).$$

Because

$$EU^p = \int_0^\infty u^p\ dF_U(u) < \infty,$$

the last term has limit 0 as $n \to \infty$. □

Theorem 5.55. *In addition to the conditions of the ergodic theorem, assume that*

$$\sum_{y\in S} f(y)\pi\{y\} = 0, \quad \textit{and} \quad E_x[V(f)_1^2] < \infty,$$

then

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n} f(X_k)$$

converges in distribution to a normal random variable, mean zero, variance.

$$\frac{1}{E_x\tau_x} Var(V(f)_1).$$

Proof. We keep the notation from the ergodic theorem and its proof. As before, write

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n} f(X_k) = \frac{1}{\sqrt{n}}\sum_{k=1}^{\tau_x^1-1} f(X_k) + \frac{1}{\sqrt{n}}\sum_{k=\tau_x^1}^{\tau_x^m-1} f(X_k) + \frac{1}{\sqrt{n}}\sum_{k=\tau_x^m}^{n} f(X_k)$$

The same argument in the proof of the ergodic theorem that shows that the first term tends almost surely to zero applies in this case.

For the last term, note that

$$\frac{1}{\sqrt{n}}|\sum_{k=\tau_x^m}^{n} f(X_k)|\frac{1}{\sqrt{n}}\sum_{k=\tau_x^m}^{\tau_{m+1}^x} |f(X_k)| = V(|f|)_k \leq \max_{1\leq j\leq n} V(|f|)_j \leq \max_{1\leq j\leq n} V(|f|)_j,$$

and that

$$P\{\frac{1}{\sqrt{n}}\max_{1\leq j\leq n} V(|f|)_j > \epsilon\} \leq \sum_{j=1}^{n} P\{\frac{1}{\epsilon^2}V(|f|)_j^2 > n\} \leq nP\{\frac{1}{\epsilon^2}V(|f|)_1^2 > n\}.$$

Now, use the lemma with $U = V(|f|)^2/\epsilon^2$ to see that

$$\frac{1}{\sqrt{n}}\max_{1\leq j\leq n} V(|f|)_j \to 0$$

as $n \to \infty$ in probability and hence in distribution. Because this bounds the third term, both the first and third terms converge in distribution to 0.

Recall that if $Z_n \to^{\mathcal{D}} Z$ and $Y_n \to^{\mathcal{D}} c$, a constant, then $Y_n + Z_n \to^{\mathcal{D}} cZ$ and $Y_n Z_n \to^{\mathcal{D}} cZ$. The first of these two conclusions show that it suffices to establish convergence in distribution of the middle term.

With this in mind, note that from the central limit theorem for independent identically distributed random variables, we have that

$$Z_n = \frac{1}{\sqrt{m}}\sum_{j=1}^{m} V(f)_j \to^{\mathcal{D}} \sigma Z.$$

Here, Z is a standard normal random variable and σ^2 is the variance of $V(f)_1$. From the proof of the ergodic, with m defined as before,

$$\frac{n}{m} \to E_x\tau_x$$

almost surely and hence in distribution.

By the continuous mapping theorem,

$$\sqrt{\frac{m}{n}} \to^{\mathcal{D}} \sqrt{\frac{1}{E_x\tau_x}}$$

and thus,

$$\frac{1}{\sqrt{n}}\sum_{k=\tau_x^1}^{\tau_x^m-1} f(X_k) = \frac{1}{\sqrt{n}}\sum_{k=1}^{m-1} V(f)_k = \sqrt{\frac{m}{n}} Z_n \to^{\mathcal{D}} \frac{1}{\sqrt{E_x\tau_x}}\sigma Z.$$

□

Example 5.56 (Markov chain Monte Carlo)**.** *If the goal is to compute an integral*

$$\int g(x)\ \pi(dx),$$

then, in circumstances in which the probability measure π *is easy to simulate, simple Monte Carlo suggests creating independent samples* $X_0(\omega), X_1(\omega), \ldots$ *having distribution* π*. Then, by the strong law of large numbers,*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} g(X_j(\omega)) = \int g(x)\ \pi(dx) \text{ with probability one.}$$

The error is determined by the central limit theorem,

$$\sqrt{n}\left(\frac{1}{n}\sum_{j=0}^{n-1} g(X_j(\omega)) - \int g(x)\ \pi(dx)\right) \to^{\mathcal{D}} \sigma Z$$

where Z *is a stardard normal random variable, and* $\sigma^2 = \int g(x)^2\ \pi(dx)) - (\int g(x)\ \pi(dx))^2$.

Markov chain Monte Carlo performs the same calculation using a irreducible Markov chain $\tilde{X}_0(\omega), \tilde{X}_1(\omega), \ldots$ *having stationary distribution* π. *The most commonly used strategy to define this sequence is the method developed by Metropolis and extended by Hastings.*

To introduce the Metropolis-Hastings algorithm, let π *be probability on a countable state space* S, $\pi \neq \delta_{x_0}$, *and let* T *be a Markov transition matrix on* S. *Further, assume that the the chain immediately enter states that have positive* π *probability. In other words, if* $T(x,y) > 0$ *then* $\pi\{y\} > 0$. *Define*

$$\alpha(x,y) = \begin{cases} \min\left\{\frac{\pi\{y\}T(y,x)}{\pi\{x\}T(x,y)}, 1\right\}, & \text{if } \pi\{x\}T(x,y) > 0, \\ 1, & \text{if } \pi\{x\}T(x,y) = 0. \end{cases}$$

If $\tilde{X}_n = x$, *generate a* candidate value y *with probability* $T(x,y)$. *With probability* $\alpha(x,y)$, *this candidate is accepted and* $\tilde{X}_{n+1} = y$. *Otherwise, the candidate is rejected and* $\tilde{X}_{n+1} = x$. *Consequently, the transition matrix for this Markov chain is*

$$\tilde{T}(x,y) = \alpha(x,y)T(x,y) + (1 - \alpha(x,y))\delta_x\{y\}.$$

Note that this algorithm only requires that we know the ratios $\pi\{y\}/\pi\{x\}$ *and thus we are not required to normalize* π.*Also, if* $\pi\{x\}T(x,y) > 0$ *and if* $\pi\{y\} = 0$, *then* $\alpha(x,y) = 0$ *and thus the chain cannot visit states with* $\pi\{y\} = 0$

Claim. $\tilde{T}$ *is the transition matrix for a reversible Markov chain with stationary distribution* π.

We must show that π *satisfies the detailed balance equation (5.2). Consequently, we can limit ourselves to the case* $x \neq y$

Case 1. $\pi\{x\}T(x,y) = 0$

In this case $\alpha(x,y) = 1$. *If* $\pi\{y\} = 0$. *Thus,* $\pi\{y\}\tilde{T}(y,x) = 0$,

$$\pi\{x\}\tilde{T}(x,y) = \pi\{x\}T(x,y) = 0,$$

and (5.2) holds.

If $\pi\{y\} > 0$ *and* $T(y,x) > 0$, *then* $\alpha(y,x) = 0$, $\tilde{T}(y,x) = 0$ *and (5.2) holds.*

If $\pi\{y\} > 0$ *and* $T(y,x) = 0$, *then* $\alpha(y,x) = 1$, $\tilde{T}(y,x) = 0$ *and (5.2) holds.*

Case 2. $\pi\{x\}T(x,y) > 0$ *and* $\alpha(x,y) = 1$

In this case,

$$\pi\{x\}\tilde{T}(x,y) = \pi\{x\}T(x,y).$$

In addition, $\alpha(y,x) \leq 1$ *and*

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

Case 3. $\pi\{x\}T(x,y) > 0$ *and* $\alpha(x,y) < 1$

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

In addition, $\alpha(y,x) = 1$ *and*

$$\pi\{y\}\tilde{T}(y,x) = \pi\{y\}T(y,x).$$

Thus, the claim holds.

Example 5.57. *The original Metropolis algorithm had* $T(x,y) = T(y,x)$ *and thus*

$$\alpha(x,y) = \min\left\{\frac{\pi\{y\}}{\pi\{x\}}, 1\right\}.$$

Example 5.58 (Independent Chains)**.** *Let* $\{X_n; n \geq 0\}$ *be independent discrete random variable with distribution function* $f(x) = P\{X_0 = x\}$*. Then*

$$\alpha(x,y) = \min\left\{\frac{w(y)}{w(x)}, 1\right\},$$

where $w(x) = f(x)/\pi\{x\}$ *is the* importance weight *function that would be used in importance sampling if the observations if observations were generated from* f

Exercise 5.59. *Take* T *to be the transition matrix for a random walk on a graph and* π *to be uniform measure. Describe* $\tilde{T}$*.*

5.5.3 Rates of Convergence to the Stationary Distribution

The coupling of two Markov process X and $\tilde{X}$ is called *successful*, if X and $\tilde{X}$ are defined on the same probability space, if

$$X_t = \tilde{X}_t$$

for all $t \geq \tau$, the coupling time, defined by

$$\tau = \inf\{t \geq 0; X_t = \tilde{X}_t\}$$

and if τ is almost surely finite. In particular, X and $\tilde{X}$ have the same transition function T

If ν_t and $\tilde{\nu}_t$ are the distributions of the processes at time t, then

$$||\nu_t^X - \nu_t^Y||_{TV} \leq P\{X_t \neq Y_t\} = P\{\tau > t\}. \tag{5.4}$$

Theorem 5.60. *Let* X *and* $\tilde{X}$ *be Markov chains with common transition matrix* T*. Let* ν_t *and* $\tilde{\nu}_t$ *denote, respectively, the distribution of* X_t *and* $\tilde{X}_t$*. Suppose that for some* $m \geq 1$*,*

$$\epsilon = \min\{T^m(x,y); x,y \in S\} > 0.$$

Then

$$||\nu_t - \tilde{\nu}_t||_{TV} \leq (1-\epsilon)^{[t/m]}.$$

Proof. Let X and $\tilde{X}$ move independently until the coupling time. Then

$$P\{\tau \le m|X_0 = x, \tilde{X}_0 = \tilde{x}\} \ge \sum_{y \in S} T^m(x,y)T^m(\tilde{x},y) \ge \epsilon \sum_{y\in S} T^m(x,y) = \epsilon,$$

and $P\{\tau > m|X_0 = x, \tilde{X}_0 = \tilde{x}\} \le 1-\epsilon$. By the Markov property, for $k = [t/m]$,

$$P\{\tau \ge t\} \le P\{\tau \ge km\} = \sum_{x,\tilde{x}\in S} P\{\tau \ge km|X_0 = x, \tilde{X}_0 = \tilde{x}\}\nu_0\{x\}\tilde{\nu}\{\tilde{x}\} \le (1-\epsilon)^k.$$

□

The rate of convergence to the stationary distribution is obtained by taking $\tilde{X}$ to start in the stationary distribution. In this case, for each t, $\tilde{\nu}_t = \pi$, the stationary distribution.

Thus, the "art" of coupling time is to arrange to have the coupling time be as small as possible.

Example 5.61 (Simple Random Walk with Reflection)**.** *Let $\{Y_n; n \ge 1\}$ be an independent sequence of $Ber(p)$ random variables. Couple X and $\tilde{X}$ as follows:*

Assume X_0 and $\tilde{X}_0$ differ by an even integer.

At time n, if $Y_n = 1$, then, if it is possible, set $X_n = X_{n-1}+1$ and $\tilde{X}_n = \tilde{X}_{n-1}+1$. This cannot occur if one of the two chains is at state M. In this circumstance, the chain must decrease by 1. Similarly, if $Y_n = -1$, then, when possible, set $X_n = X_{n-1}-1$ and $\tilde{X}_n = \tilde{X}_{n-1}-1$. This cannot occur when the chain is at state $-M$ in which case, the chain must increase by 1.

If, for example, $X_0 \le \tilde{X}_0$, then $X_n \le \tilde{X}_n$ for all $n \ge 1$. Thus, the coupling time $\tau = \min\{n \ge 0; X_n = M\}$.

Exercise 5.62. *Find $E_{(x,\tilde{x})}\tau$ in the example above and show that the coupling above is successful.*

A second, closely associated for establishing rates of convergence are *strong stationary times.*

Definition 5.63. *Let X be a time homogeneous $\mathcal{F}_t$-Markov chain on a countable state space S having unique stationary distribution π. An $\mathcal{F}_t$-stopping time τ is called a* strong stationary time *provided that*

1. *X_τ has distribution π*
2. *X_τ and τ are independent.*

Remark 5.64. *The properties of a strong stationary time imply that*

$$P\{X_t = y, \tau = s\} = P\{X_\tau = y, \tau = s\} = \pi\{y\}P\{\tau = s\}.$$

Now, sum on $s \le t$ to obtain

$$P\{X_t = y, \tau \le t\} = \sum_{s=0}^{t} P\{X_t = y, \tau = s\} = \sum_{s=0}^{t} P\{X_t = y|\tau = s\}P\{\tau = s\} = \pi\{y\}P\{\tau \le t\}$$

because

$$\begin{aligned} P\{X_t = y|\tau = s\} &= \sum_{x\in S} P\{X_t = y, X_s = x|\tau = s\} = \sum_{x\in S} P\{X_t = y|X_s = x, \tau = s\}P\{X_s = x|\tau = s\} \\ &= \sum_{x\in S} P\{X_t = y|X_s = x\}P\{X_\tau = x|\tau = s\}. = \sum_{x\in S} T^t(x,y)\pi\{x\} = \pi\{y\} \end{aligned}$$

Example 5.65 (Top to random shuffle)**.** *Let X be a Markov chain whose state space is the set of permutation on N letters, i.e, the order of the cards in a deck having a total of N cards. We write $X_n(k)$ to be the card in the k-th position after n shuffles. The transitions are to take the top card and place it at random uniformly in the deck. In the exercise to follow, you are asked to check that the uniform distribution on the $N!$ permutations is stationary. Define*

$$\tau = \inf\{n > 0; X_n(1) = X_0(N)\} + 1,$$

i.e., the first shuffle after the original bottom card has moved to the top.

Claim. *τ is a strong stationary time.*

We show that at the time in which we have k cards under the original bottom card ($X_n(N-k+1) = X_0(N)$, then all $k!$ orderings of these k cards are equally likely.

In order to establish a proof by induction on k, note that the statement above is obvious in the case $k = 1$. For the case $k = j$, we assume that all $j!$ arrangements of the j cards under $X_0(N)$ are equally likely. When one additional card is placed under $X_0(N)$, each of the $j + 1$ available positions are equally likely, and thus all $(j + 1)!$ arrangements of the $j = 1$ cards under $X_0(N)$ are equally likely.

Consequently, the first time that the original bottom card is placed into the deck all $N!$ orderings are equally likely independent of the value of τ.

Exercise 5.66. *Check that the uniform distribution is the stationary distribution for the top to random shuffle.*

We now will establish an inequality similar to (5.4) for strong stationary times. For this segment of the notes, X is an ergodic Markov chain on a countable state space S having transition matrix T and unique stationary distribution π

Definition 5.67. *Define the* separation distance

$$s_t = \sup\left\{1 - \frac{T^t(x,y)}{\pi\{y\}}; x, y \in S\right\}.$$

Proposition 5.68. *Let τ be a strong stationary time for X, then*

$$s_t \le \max_{x\in S} P_x\{\tau > t\}.$$

Proof. For any $x, y \in S$,

$$\begin{aligned} 1 - \frac{T^t(x,y)}{\pi\{y\}} &= 1 - \frac{P_x\{X_t = y\}}{\pi\{y\}} \le 1 - \frac{P_x\{X_t = y, \tau \le t\}}{\pi\{y\}} \\ &= 1 - \frac{\pi\{y\}P_x\{\tau \le t\}}{\pi\{y\}} = 1 - P_x\{\tau \le t\} = P_x\{\tau > t\}. \end{aligned}$$

□

Theorem 5.69. *For $t = 0, 1, \ldots$, write $\nu_t(A) = P\{X_t \in A\}$, then*

$$||\nu_t - \pi||_{TV} \le s_t.$$

Proof.

$$\begin{aligned}
||\nu_t - \pi||_{TV} &= \sum_{y\in S}(\pi\{y\} - \nu_t\{y\})I_{\{\pi\{y\}>\nu_t\{y\}\}} = \sum_{y\in S}\pi\{y\}\left(1 - \frac{\nu_t\{y\}}{\pi\{y\}}\right)I_{\{\pi\{y\}>\nu_t\{y\}\}} \\
&= \sum_{y\in S}\pi\{y\}\sum_{x\in S}\nu_0\{x\}\left(1 - \frac{T^t(x,y)}{\pi\{y\}}\right)I_{\{\pi\{y\}>\nu_t\{y\}\}} \\
&\le \sum_{y\in S}\pi\{y\}\sum_{x\in S}\nu_0\{x\}s_t = s_t
\end{aligned}$$

□

Combining the proposition and the corollary, we obtain the following.

Corollary 5.70. *Let τ be a strong stationary time for X, then*

$$||\nu_t - \pi||_{TV} \le \max_{x\in S} P_x\{\tau > t\}.$$

Example 5.71 (Top to random shuffle). *Note that $\tau = 1 + \sum_{k=1}^N \sigma_k$ where σ_k is the number of shuffles necessary to increase the number of cards under the original bottom card from $k-1$ to k. Then these random variables are independent and $\sigma_k - 1$ has a $Geo((k-1)/N)$ distribution.*

We have seen these distributions before, albeit in the reverse order for the coupon collectors problem. In this case, we choose from a finite set, the "coupons", uniformly with replacement and set τ_k to be the minimum time m so that the range of the first m choices is k. Then $\tau_k - \tau_{k-1} - 1$ has a $Geo(1-(k-1)/N)$ distribution.

Thus the time τ is equivalent to the number of purchases until all the coupons are collected. Define $A_{j,t}$ to be the event that the j-th coupon has not been chosen by time t. Then

$$P\{\tau > t\} = P\left(\bigcup_{j=1}^N A_{j,t}\right) \le \sum_{j=1}^N P(A_{j,t}) = \sum_{j=1}^N (1 - \frac{1}{N})^t = N(1 - \frac{1}{N})^t \le Ne^{-t/N}$$

For the choice $t = N(\log N + c)$,

$$P\{\tau > N(\log N + c)\} \le N\exp -N(\log N + c)/N = e^{-c}.$$

This shows that the shuffle of the deck has a threshold time of $N\log N$ and after that tail of the distribution of the stopping time decreases exponentially.

5.6 Transient Chains

5.6.1 Asymptotic Behavior

Now, let C be a closed set of transient states. Then we have from the renewal theorem that

$$\lim_{n\to\infty} T^n(x,y) = 0, \quad x \in S, y \in C.$$

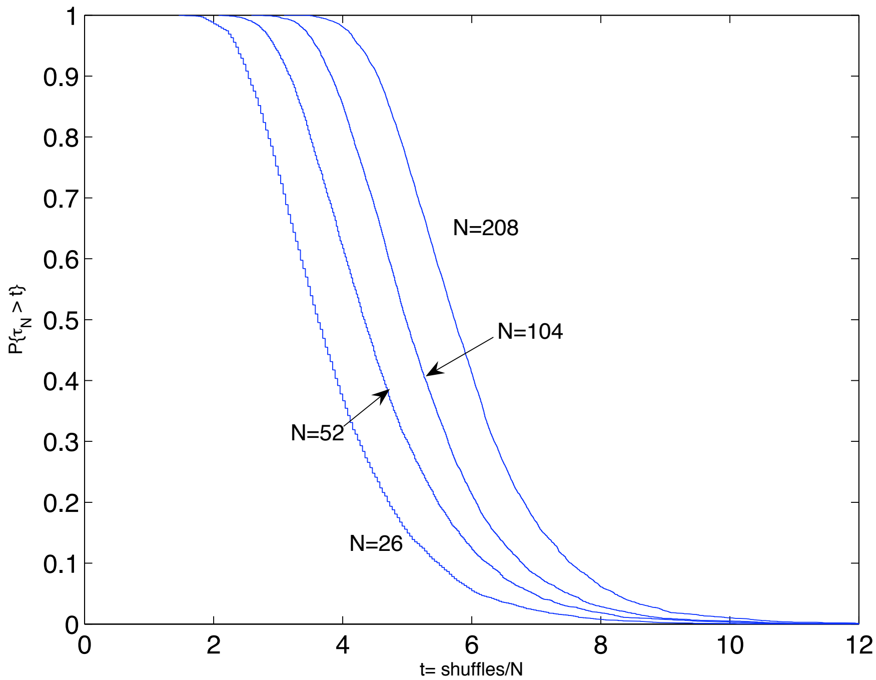

Figure 2: Reading from left to right, the plot of the bound given by the strong stationary time of the total variation distance to the uniform distribution after t shuffles for a deck of size $N = 26, 52, 104$, and 208 based on 5,000 simulations of the stopping time τ.

If we limit our scope at time n to those paths that remain in C, then we are looking at the limiting behavior of conditional probabilities

$$\tilde{T}^n(x,y) = P_x\{X_n = y | X_n \in C\} = \frac{T^n(x,y)}{\sum_{z\in C} T^n(x,z)}, \quad x,y \in C.$$

What we shall learn is that the exists a number $\rho_C > 1$ such that

$$\lim_{n\to\infty} T^n(x,y)\rho_C^n$$

exists. In this case,

$$\tilde{T}^n(x,y) = \frac{T^n(x,y)\rho_C^n}{\sum_{z\in C} T(x,z)\rho_C^n}, \quad x,y \in C.$$

The limit of this sequence will play very much the same role as that played by the invariant measure for a recurrent Markov chain.

The analysis begins with the following lemma:

Lemma 5.72. *Assume that a real-valued sequence* $\{a_n; n \geq 0\}$ *satisfies the* subadditve condition

$$a_{m+n} \leq a_m + a_n,$$

for all $m, n \in \mathbb{N}$. *Then,*

$$\lim_{n\to\infty} \frac{a_n}{n}$$

exists and equals

$$\inf_{n>0} \frac{a_n}{n} < \infty.$$

Proof. Call the infimum above a. Clearly,

$$\liminf_{n\to\infty} \frac{a_n}{n} \geq a.$$

For the reverse inequality, fix $m \geq 1$. Then write $n = km + \ell$, $0 \leq \ell < m$. Then

$$a_n \leq ka_m + a_\ell.$$

As $n \to \infty$, $n/k \to m$. Therefore,

$$\limsup_{n\to\infty} \frac{a_n}{n} \leq \frac{a_m}{m}.$$

Because this holds for all m,

$$\limsup_{n\to\infty} \frac{a_n}{n} \leq a.$$

□

Remark 5.73. *For the circumstances in the lemma, we have*

$$a_n \geq na.$$

Theorem 5.74. *Assume that* C *is a communicating class of aperiodic states. Then*

$$\ell_C = \lim_{n\to\infty} -\frac{1}{n} \log T^n(x, y)$$

exists with the limit independent of the choice of x *and* y *in* C.

Proof. Pick $x \in C$. As we have seen before, $T^n(x, x) > 0$ for sufficiently large n. Now set

$$a_n = -\log T^n(x, x).$$

Because

$$T^{m+n}(x, x) \geq T^m(x, x)T^n(x, x),$$

the sequence $\{a_n; n \geq 0\}$ satisfies the subaddtivity property in the lemma above. Thus,

$$\lim_{n\to\infty} -\frac{1}{n} \log T^n(x, x)$$

exists. Call the limit ℓ_x.

To show that the limit is independent of x, note that

$$T^n(x,x) \leq \exp(-\ell_x n).$$

As we argued previously in (5.1), choose m and n so that

$$T^n(x,y) > 0 \quad \text{and} \quad T^m(y,x) > 0.$$

Thus,

$$T^n(x,y)T^k(y,y)T^m(y,x) \leq T^{n+k+m}(x,x) \leq \exp(-\ell_x(n+k+m)).$$

Thus,

$$T^k(y,y) \leq \frac{\exp(-\ell_x(m+n))}{T^n(x,y)T^m(y,x)} \exp(-\ell_x k)$$

and

$$-\ell_y \leq \liminf_{k\to\infty} \frac{1}{k} \log T^k(y,y) \leq -\ell_x.$$

Now, reverse the roles of x and y to see that $\ell_x = \ell_y$.

Finally note that

$$T^{m+k}(y,y) \geq T^m(y,x)T^k(x,y) \quad \text{and} \quad T^{n+k}(x,y) \geq T^n(x,y)T^k(y,y).$$

Consequently,

$$\frac{1}{m+k} \log T^{m+k}(y,y) \geq \frac{1}{m+k} \log T^m(y,x) + \frac{1}{m+k} \log T^k(x,y).$$

and

$$\frac{1}{n+k} \log T^k(x,y) \geq \frac{1}{n+k} \log T^n(x,y) + \frac{1}{n+k} \log T^k(y,y)$$

Let $k \to \infty$ to conclude that

$$-\ell_y \geq \limsup_{k\to\infty} \frac{1}{m+k} \log T^k(x,y) \geq \liminf_{k\to\infty} \frac{1}{n+k} \log T^k(x,y) \geq -\ell_y$$

giving that the limit is independent of x and y. □

With some small changes, we could redo this proof with period d and obtain the following

Corollary 5.75. *Assume that C is a communicating class of states having common period d. Then there exists $\lambda_C \in [0,1]$ such that for some n depending on x and y,*

$$\lambda_C = \lim_{k\to\infty} (T^{n+kd}(x,y))^{1/(n+kd)}$$

exists with the limit independent of the choice of x and y in C.

Proof. Take $\lambda_C = \exp(-\ell_C)$. □

Return to the use of generating functions, define

$$G_{xy}(z) = \sum_{n=1}^{\infty} T^n(x,y)z^n \quad \text{and} \quad G_{\tau,xy}(z) = \sum_{n=1}^{\infty} P_x\{\tau_y = n\}z^n.$$

By the theorem, we have that G_{xy} has radius of convergence $\rho_C = \lambda_C^{-1}$.

Exercise 5.76. $G_{xx}(z) = 1 + G_{\tau,xx}(z)G_{xx}(z)$.
Hint: $T^n(x,x) = I_{\{n=0\}} + \sum_{m=0}^{n} P_x\{\tau_x = m\}T^{n-m}(x,x)$.

Lemma 5.77. *Let $x \in C$. Then*

1. *$G_{\tau,xx}(z) \leq 1$ for all $z \in [0, \rho_C]$.*
2. *$G_{\tau,xx}(\rho_C) = 1$ if and only if $G_{xx}(\rho_C) = +\infty$.*

Proof. 1. If $z \in [0, \rho_C)$, then $G_{xx}(z) < \infty$ and by the identity above, $G_{\tau,xx}(z) < 1$. Now, use the monotone convergence theorem with $z \to \rho_C-$ to obtain the result.

2. Write the idenity as

$$G_{\tau,xx}(s) = 1 - \frac{1}{G_{xx}(z)} \quad \text{and} \quad G_{xx}(z) = \frac{1}{1 - G_{\tau,xx}(z)}.$$

Again, use the monotone convergence theorem with $z \to \rho_C-$ to obtain the result.

□

Definition 5.78. *Call a state $x \in C$, ρ_C-*recurrent *if $G_{xx}(\rho_C) = +\infty$ and ρ_C-*transient *if $G_{xx}(\rho_C) < +\infty$.*

Remark 5.79.

1. *ρ_C-recurrence is equivalent to $G_{\tau,xx}(\rho_C) = 1$.*
2. *ρ_C-transience is equivalent to $G_{\tau,xx}(\rho_C) < 1$.*
3. *1-recurrent and 1-transient is simply recurrent and transient.*

Exercise 5.80. *The properties of ρ_C-recurrence and ρ_C-transience are properties of the equivalence class under communication.* Hint: $T^{m+n+\ell}(x,x)\rho_C^{m+n+\ell} \leq T^m(x,y)T^n(y,y)T^\ell(y,x)\rho_C^{m+n+\ell}$.

Example 5.81. *For a simple asymmetric random walk having p as the probability of a step right, $T^m(x,x) = 0$ if m is odd, and*

$$T^{2n}(x,x) = \binom{2n}{n} p^n(1-p)^n.$$

By Stirling's approximation, $n! \approx \sqrt{2\pi} n^{n+1/2}e^{-n}$, we have

$$\binom{2n}{n} \approx \frac{2^{2n+1/2}}{\sqrt{2\pi n}}$$

and

$$\lim_{n\to\infty} T^{2n}(x,x)^{1/2n} = \lim_{n\to\infty} 2\sqrt{p(1-p)}\left(\frac{1}{\sqrt{\pi n}}\right)^{1/2n} = 2\sqrt{p(1-p)}.$$

Thus,

$$\rho_C = \frac{1}{2\sqrt{p(1-p)}}.$$

In addition,

$$G_{xx}(z) = \sum_{n=0}^{\infty} \binom{2n}{n} (p(1-p)z^2)^n = \frac{1}{\sqrt{1 - 4p(1-p)z^2}}.$$

Consequently, $G_{xx}(\rho_C) = +\infty$ and the chain is ρ_C-recurrent.

Definition 5.82. *Let $A \subset S$. The* taboo probabilities

$$T_A^n(x,y) = \begin{cases} \delta_{xy} & \text{if } n = 0 \text{ and } x \notin A \\ 0 & \text{if } n = 0 \text{ and } x \in A \\ T(x,y) & \text{if } n = 1 \\ \sum_{z \notin A} T_A^{n-1}(x,z)T(z,y) & \text{if } n > 1 \end{cases}$$

Thus, for $n \geq 1$

$$T_A^n(x,y) = P_x\{X_n = y, X_k \notin A \text{ for } k = 1, 2, \ldots, n-1\}.$$

Denote the generating functions associated with the taboo probabilities by

$$G_{A,xy}(z) = \sum_{n=0}^{\infty} T_A^n(x,y)z^n.$$

Lemma 5.83. *$G_{x,xx}(\rho_C) \leq 1$, and $G_{x,yx}(\rho_C) < +\infty$ for all $x, y \in C$.*

Proof.

$$G_{x,xx}(\rho_C) \leq G_{\tau,xx}(\rho_C) \leq 1.$$

For $x \neq y$. because $y \to x$, we can choose $m > 0$ so that $T_x^m(y,x) > 0$. Thus, for any $n \geq 0$

$$T_x^n(x,y)T_x^m(y,x) \leq T_x^{m+n}(x,x) = P_x\{\tau_x = m+n\}.$$

Consequently,

$$G_{x,yx}(\rho_C) \leq \frac{1}{T_n^m(y,x)\rho_C^m} \sum_{n=1}^{\infty} P_x\{\tau_x = m+n\}\rho_C^{m+n} \leq \frac{G_{\tau,xx}(\rho_C)}{T_x^m(y,x)\rho_C^n} < +\infty.$$

□

5.6.2 ρ_C Invariant Measures

Lemma 5.84. *Let $r > 0$ and let m be a measure on C satisfying*

$$\sum_{x \in C} m\{x\}T(x,y)r \leq m\{y\}, \quad \text{for all } y \in C.$$

Then

1. *$\sum_{x \in C} m\{x\}T^n(x,y)r^n \leq m\{y\}$, for all $y \in C$ and $n \geq 0$.*
2. *Either $m\{x\} = 0$ for all $x \in C$ or $m\{x\} > 0$ for all $x \in C$.*
3. *If $m\{x\} > 0$ for all $x \in C$, then $r \leq \rho_C$.*

Proof. 1. Choose $x \in C$. Note that

$$\sum_{z\in C} T(x,z)T^{n-1}(z,y) = \sum_{z\in C} T(x,z)T^{n-1}(z,y) + \sum_{z\in S\setminus C} T(x,z)T^{n-1}(z,y).$$

By the Chapman-Kolmogorov equation, the first term is $T^n(x,y)$. Because C is an equivalence class under communication, we must either have $x \not\to z$ or $z \not\to y$. In either case $T(x,z)T^{n-1}(z,y) = 0$ and the second sum vanishes. Thus,

$$\begin{aligned} \sum_{x\in C} m\{x\}T^n(x,y)r^n &= \sum_{x\in C} m\{x\}\left(\sum_{z\in C} T(x,z)T^{n-1}(z,y)\right) r^n \\ &= \sum_{z\in C}\left(\sum_{x\in C} m\{x\}T(x,z)r\right) T^{n-1}(z,y)r^{n-1} \le \sum_{z\in C} m\{z\}T^{n-1}(z,y)r^{n-1}. \end{aligned}$$

and the inequality holds by induction.

2. Choose x with $m\{x\} > 0$. Given $y \in C$, choose n so that $T^n(x,y) > 0$. Then by 1,

$$m\{y\} \ge m\{x\}T^n(x,y)r^n > 0.$$

3. From 1 and 2,

$$T^{nd}(x,x)r^{nd} \le 1.$$

Thus,

$$\rho_C^{-1} = \lim_{n\to\infty} T^{nd}(x,x)^{1/nd} \le r^{-1}.$$

or $r \le \rho_C$.

□

Definition 5.85. *A measure m satisfying*

$$\sum_{x\in C} m\{x\}T(x,y)r = (\le) m\{y\}, \quad \textit{for all } y \in C$$

is called an r-(sub)invariant measure.

We shall see that there is always a ρ_C-subinvariant measure and it is unique up to constant multiples if and only if C is ρ_C-recurrent.

Theorem 5.86. *1. Choose $x \in C$ and define*

$$\tilde{m}\{y\} = G_{x,xy}(\rho_C).$$

Then $\tilde{m}$ is an ρ_C-subinvariant measure on C.

2. $\tilde{m}$ is unique up to constant multiples if and only if C is ρ_C-recurrent. In this case, $\tilde{m}$ is an ρ_C-invariant measure on C.

Proof. 1. We have, by the lemma, that $\tilde{m}\{y\} < \infty$. Also, $\tilde{m}\{x\} = G_{x,xx}(\rho_C) = G_{\tau,xx}(\rho_C) \leq 1$. Using the definition of taboo probabilities, we see that

$$\begin{aligned}
\sum_{x\in C} \tilde{m}\{x\}T(x,z)\rho_C &= \sum_{n=1}^{\infty} \rho_C^{n+1} \sum_{y\in C} T_x^n(x,y)T(y,z) \\
&= \sum_{n=1}^{\infty} \rho_C^{n+1} \left(\sum_{y\neq x} T_x^n(x,y)T(y,z) + T_x^n(x,x)T(x,z) \right) \\
&= \sum_{n=1}^{\infty} \rho_C^{n+1} \left(T_x(x,z)^{n+1} + T_x^n(x,x)T(x,z) \right) \\
&= \sum_{n=1}^{\infty} T_x^{n+1}(x,z)\rho_C^{n+1} + \left(\sum_{n=1}^{\infty} T_x^n(x,x)\rho_C^n \right) \rho_C T(x,z) \\
&= \tilde{m}\{z\} - T_x(x,z)\rho_C + \tilde{m}\{x\}\rho_C T(x,y) = \tilde{m}\{z\} - (1 - m\{x\})\rho_C T(x,y) \\
&\leq \tilde{m}\{z\}.
\end{aligned}$$

2. First, assume that the solution $\tilde{m}$ is unique up to constant multiples and define

$$n\{y\} = \begin{cases} \tilde{m}\{y\} & \text{if } y \neq x, \\ 1 & \text{if } y = x. \end{cases}$$

Then,

$$\sum_{y\in C} n\{y\}T(y,z)\rho_C = \sum_{y\in C} \tilde{m}\{y\}T(y,z)\rho_C + (1 - \tilde{m}\{x\})T(x,z)\rho_C \leq \tilde{m}\{y\} \leq n\{y\}.$$

Consequently, $\{n\{y\}; y \in C\}$ is also a subinvariant measure that differs from $\tilde{m}$ only at x. By uniqueness, we have that

$$1 = n\{x\} = G_{x,xx}(\rho_C) = G_{\tau,xx}(\rho_C).$$

Thus, x and hence every state in C is ρ_C-recurrent.

For the converse, we begin with a claim.

Claim I. Let m be a ρ_C-subinvariant measure. Then,

$$m\{z\} \geq m\{x\}\tilde{m}\{z\}.$$

We show this by establishing for every $N \in \mathbb{N}$ that

$$m\{z\} \geq m\{x\} \sum_{n=1}^{N} T_x^n(x,z)\rho_C^n.$$

The case $N = 1$ is simply the definition of ρ_C-subinvariant. Thus, suppose that the inequality holds for N. Then,

$$\begin{aligned} m\{x\}\sum_{n=1}^{N+1} T_x^n(x,z)\rho_C^n &= m\{x\}\sum_{n=0}^{N} T_x^{n+1}(x,z)\rho_C^{n+1} \\ &= m\{x\}T(x,z)\rho_C + m\{x\}\sum_{n=1}^{N}\sum_{y\neq x} T_x^n(x,y)T(y,z)\rho_C^{n+1} \\ &= m\{x\}T(x,z)\rho_C + \sum_{y\neq x}\left(m\{x\}\sum_{n=1}^{N} T_x^n(x,y)\rho_C^n\right)T(y,z)\rho_C \\ &\leq m\{x\}T(x,z)\rho_C + \sum_{y\neq x} m\{y\}T(y,z)\rho_C = \sum_{y\in C} m\{y\}T(y,z)\rho_C \leq m\{z\} \end{aligned}$$

and the claim holds.

Finally, because C is ρ_C-recurrent, $\tilde{m}\{x\} = 1$ and from the computation above, we have the equality,

$$\sum_{y\in C} \tilde{m}\{y\}T(y,z)\rho_C = \tilde{m}\{x\}$$

and thus

$$\sum_{y\in C} \tilde{m}\{y\}T^n(y,z)\rho_C = \tilde{m}\{x\}.$$

Claim II. $m\{z\} = m\{x\}\tilde{m}\{z\}$.

We have that $m\{z\} \geq m\{x\}\tilde{m}\{z\}$. If we have strict inequality for some state, then for some n,

$$m\{x\} \geq \sum_{y\in C} m\{y\}T^n(y,x)\rho_C^n > \sum_{y\in C} m\{x\}\tilde{m}\{y\}T^n(y,x)\rho_C^n = m\{x\}\tilde{m}\{x\} = m\{x\},$$

a contradiction. This establishes equality and with it that m is ρ_C-invariant.

□

Definition 5.87. *A function $h : S \to \mathbb{R}$ is called an r-(super)harmonic function for T if*

$$rTh(x)(\leq) = h(x) \quad \text{for all } x \in C.$$

Theorem 5.88. *There exists a strictly positive ρ_C-superharmonic function.*

Proof. Define

$$\tilde{T}(x,y) = \rho_C \frac{m\{y\}}{m\{x\}} T(y,x) \tag{5.5}$$

where m is a subinvariant measure for T. Thus,

$$\sum_{y\in C} \tilde{T}(x,y) \leq 1.$$

Append to C a state Δ and define

$$\tilde{T}(x,\Delta) = \begin{cases} 1 - \sum_{y\in C} T(x,y) & x \neq \Delta, \\ 1 & x = \Delta. \end{cases}$$

Note that $\tilde{T}$ is a transition function with irreducible statespace $C \cup \{\Delta\}$. Note that

$$\begin{aligned} \tilde{T}^2(x,y) &= \sum_{y\in C} \tilde{T}(x,z)\tilde{T}(z,x) \\ &= \sum_{z\in C} \tilde{T}(x,z)\tilde{T}(z,y) \\ &= \sum_{z\in C} \rho_C \frac{m\{z\}}{m\{x\}} T(z,x) \rho_C \frac{m\{y\}}{m\{z\}} T(y,z) = \rho_C^2 \frac{m\{y\}}{m\{x\}} T^2(y,x). \end{aligned}$$

By a straightforward induction argument, we have that

$$\tilde{T}^n(x,y) = \rho_C^n \frac{m\{y\}}{m\{x\}} T^n(y,x).$$

Consequently,

$$\lim_{n\to\infty} \frac{1}{n} \log \tilde{T}^n(x,y) = \log \rho_C - \log \rho_C = 0.$$

Thus, there exists a 1-subinvariant measure $\tilde{m}$ for $\tilde{T}$ on C, i.e.,

$$\sum_{x\in C} \tilde{m}\{x\}\tilde{T}(x,y) \leq \tilde{m}\{y\}.$$

or

$$\sum_{x\in C} T(y,x)\rho_C \frac{\tilde{m}\{x\}}{m\{x\}} \leq \frac{\tilde{m}\{y\}}{m\{y\}}.$$

Thus,

$$h(y) = \frac{\tilde{m}\{y\}}{m\{y\}}$$

is a strictly positive ρ_C-superharmonic function. □

Remark 5.89. *Because* $\lim_{n\to\infty} T^n(x,y)$ *exists and* $m\{x\} > 0$ *for all* x,

$$\lim_{n\to\infty} \tilde{T}^n(x,y)\rho_C^n$$

exists for all $x \in C$.

Definition 5.90. *As before, we shall make a distinction between the types of recurrence.*

1. *Call a state* ρ_C*-recurrent* ρ_C-positive recurrent *if*

$$\lim_{n\to\infty} T^n(x,y)\rho_C^n > 0.$$

*2. Call a state ρ_C-recurrent ρ_C-*null recurrent *if*

$$\lim_{n\to\infty} T^n(x,y)\rho_C^n = 0.$$

Exercise 5.91. *ρ_C-positive and null recurrence is a property of the equivalence class.*

Theorem 5.92. *Suppose that C is ρ_C-recurrent. Then m, the ρ_C-subinvariant measure and h, the ρ_C-superharmonic function are unique up to constant multiples. In fact, m is ρ_C-invariant and h is ρ_C-harmonic.*

The class is ρ_C-positive recurrent if and only if

$$\sum_{x\in C} m\{x\}f(x) < +\infty.$$

In this case,

$$\lim_{n\to\infty} \rho_C^n T^n(x,y) = \frac{m\{y\}h(x)}{\sum_{z\in C} m\{z\}h(z)}.$$

Proof. We have shown that m is ρ_C-invariant and unique up to constant multiple. In addition, using $\tilde{T}$ as defined in (5.5), we find that

$$\sum_{n=0}^{\infty} \tilde{T}^n(x,y) = \frac{m\{y\}}{m\{x\}} \sum_{n=0}^{\infty} T^n(y,x)\rho_C^n = +\infty$$

because C is ρ_C-recurrent. Consequently, the Markov chain $\tilde{X}$ associated with $\tilde{T}$ is recurrent and the measure $\tilde{m}$ and hence the function h above are unique up to constant multiples. Because $\tilde{m}$ is ρ_C-invariant, h is ρ_C-harmonic. In addition, $\tilde{T}(x,\Delta) = 0$.

To prove the second assertion, first, assume that C is ρ_C-positive recurrent. Then $\tilde{X}$ is positive recurrent and so

$$\lim_{n\to\infty} \tilde{T}^n(x,y) = \pi\{x\}$$

exists and is the unique stationary probability measure on C. In other words,

$$\sum_{x\in C} \pi\{x\}\tilde{T}(x,y) = \pi\{y\}$$

and

$$\sum_{x\in C} \frac{\pi\{x\}}{m\{x\}} T(y,x)\rho_C = \frac{\pi\{y\}}{m\{y\}}$$

and π/m is ρ_C-harmonic. Thus, for any $x \in C$,

$$\pi\{x\} = cm\{x\}h(x).$$

Because π is a probability measure,

$$\sum_{x\in C} m\{x\}h(x) = \frac{1}{c} < \infty.$$

For the converse, suppose $\sum_{x\in C} m\{x\}h(x) < \infty$. Because h is ρ_C-harmonic, $\rho_C^n T^n h = h$ or for all $x \in C$,

$$\sum_{y\in C} \rho_C^n T^n(x,y)h(y) = h(x) \quad \text{and} \quad \sum_{y\in C} m\{y\}\tilde{T}^n(y,x)h(y) = h(x)m\{x\}.$$

Let $n \to \infty$ to obtain the identity

$$\sum_{y\in C} m\{y\}\pi\{x\}h(y) = h(x)m\{x\}.$$

Consequently, $\pi\{x\} > 0$ for each $x \in C$ and C is positive recurrent for $\tilde{X}$. Thus, C is ρ_C-positive recurrent.

When this holds, the $\lim_{n\to\infty} \tilde{T}^n(y,x) = \pi\{x\}$ becomes

$$\lim_{n\to\infty} \rho_C^n T^n(x,y) = \pi\{x\}\frac{m\{y\}}{m\{x\}} = \frac{m\{y\}h(x)}{\sum_{z\in C} m\{z\}h(z)}.$$

□

Remark 5.93.

1. *If the chain is recurrent, then $\rho_C = 1$ and the only bounded harmonic functions are constant and $m\{x\} = c\pi\{x\}$.*
2. *If the chain is finite, then the condition $\sum_{x\in C} m\{x\}f(x) < +\infty$ is automatically satisfied and the class C is always ρ_C-positive recurrent for some value of ρ_C.*

6 Stationary Processes

6.1 Definitions and Examples

Definition 6.1. *A random sequence* $\{X_n; n \geq 0\}$ *is called a* stationary process *if for every* $k \in \mathbb{N}$, *the sequence*

$$X_k, X_{k+1}, \ldots$$

as the same distribution. In other words, for each $n \in \mathbb{N}$ *and each* $n+1$*-dimensional measurable set* B,

$$P\{(X_k, \ldots, X_{k+n}) \in B\} = P\{(X_0, \ldots, X_n) \in B\}.$$

Note that the Daniell-Kolmogorov extension theorem states that the $n+1$ dimensional distributions determine the distribution of the process on the probability space $S^{\mathbb{N}}$.

Exercise 6.2. *Use the Daniell-Kolmogorov extension theorem to show that any stationary sequence* $\{X_n; n \in \mathbb{N}\}$ *can be embedded in a stationary sequence* $\{\tilde{X}_n; n \in \mathbb{Z}\}$

Proposition 6.3. *Let* $\{X_n; n \geq 0\}$ *be a stationary process and let*

$$\phi : S^{\mathbb{N}} \to \tilde{S}$$

be measurable and let

$$\theta : S^{\mathbb{N}} \to S^{\mathbb{N}}$$

be the shift operator. Define the process

$$Y_k = \phi(X_k, X_{k+1}, \ldots) = \phi(\theta^k X).$$

Then $\{Y_k; k \geq 0\}$ *is stationary.*

Proof. For $x \in S^{\mathbb{N}}$, write

$$\phi_k(x) = \phi(\theta^k x).$$

Then for any $n+1$-dimensional measurable set B, set

$$A = \{x : (\phi_0(x), \phi_1(x), \ldots) \in B\}.$$

Then for any $k \in \mathbb{N}$,

$$(X_k, X_{k+1}, \ldots) \in A \text{ if and only if } (Y_k, Y_{k+1}, \ldots) \in B.$$

□

Example 6.4.

1. *Any sequence of independent and identically distributed random variables is a stationary process.*
2. *Let* $\phi : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ *be defined by*

   $$\phi(x) = \lambda_0 x_0 + \cdots + \lambda_n x_n.$$

   Then $Y_n = \phi(\theta^n X)$ *is called a* moving average process.
3. *Let* X *be a time homogeneous ergodic Markov chain. If* X_0 *has the unique stationary distribution, then* X *is a stationary process.*

4. *Let X be a Gaussian process, i.e., the finite dimensional distributions of X are multivariate random vectors. Then, if the means EX_k are constant and the covariances*

$$Cov(X_j, X_k)$$

are a function $c(k-j)$ of the difference in their indices, then X is a stationary process.

5. *Call a measurable transformation*

$$T : \Omega \to \Omega$$

measure preserving *if*

$$P(T^{-1}A) = P(A) \quad \text{for all } A \in \mathcal{F}.$$

In addition, let X_0 be measurable, then

$$X_n(\omega) = X_0(T^n\omega)$$

is a stationary process.

6. *For the example above, take $\Omega = [0, 1)$, P to be Lebesgue measure, and $\mathcal{F}$ to be the Borel σ-algebra. Define*

$$T\omega = \omega + \beta \mod 1.$$

Exercise 6.5. *Check that the examples above are indeed stationary processes.*

Definition 6.6. 1. *A subset A of $S^{\mathbb{N}}$ is called* shift invariant *if*

$$A = \theta A.$$

2. *For a random sequence X, a set B is called* strictly invariant *if*

$$B = \{X \in A\}.$$

for some measurable shift invariant set A and invariant *if*

$$B = \{(X_1, X_2, \ldots) \in A\} \quad a.s.$$

Example 6.7. *Some examples of shift invariant sets are*

1. $A_1 = \{x; x_n = 0 \text{ for almost all } n\}$.
2. $A_2 = \{x; x_n = 0 \text{ for infinitely many } n\}$.
3. $A_3 = \{x; \lim_{n\to\infty} x_n < c\}$.
4. $A_4 = \{x; \limsup_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} x_k > \epsilon\}$.

Exercise 6.8. 1. *$\mathcal{I}$, the collection of shift invariant sets is a σ-algebra.*

2. *If X_0 is integrable, then for any $k > 0$, $E[X_k|\mathcal{I}] = E[X_0|\mathcal{I}]$*

Exercise 6.9. *Let X be an i.i.d. sequence and let*

$$\mathcal{T} = \bigcap_{n=1}^{\infty} \sigma\{\theta^n X\}$$

be the tail σ-algebra. Then $\mathcal{T} \subset \mathcal{I}$.

6.2 Birkhoff's Ergodic Theorem

The proof of Birkhoff's ergodic depends on this somewhat strange lemma.

Theorem 6.10 (Maximal ergodic theorem)**.** *For a stationary process X, let*

$$S_{k,n} = X_k + \cdots + X_{k+n-1},$$

and

$$M_{k,n} = \max\{0, S_{k,1} \ldots, S_{k,n}\}.$$

Then,

$$E[X_0; \{M_{0,n} > 0\}] \geq 0.$$

Proof. (Garcia) Note that the distribution of $S_{k,n}$ and $M_{k,n}$ does not depend on k. If $j \leq n$,

$$M_{1,n} \geq S_{1,j}.$$

Consequently, adding X_0 to both sides gives

$$X_0 + M_{1,n} \geq X_0 + S_{1,j} = S_{0,j+1} \text{ for } j = 1, \ldots, n-1$$

or

$$X_0 \geq S_{0,j} - M_{1,n} \text{ for } j = 2, \ldots, n.$$

Note that for $j = 1$,

$$X_0 \geq X_0 - M_{1,n} = S_{0,1} - M_{1,n}.$$

Therefore, if $M_{0,n} > 0$,

$$X_0 \geq \max_{1 \leq j \leq n} S_{0,j} - M_{1,n} = M_{0,n} - M_{1,n}.$$

Consequently,

$$E[X_0; \{M_{0,n} > 0\}] \geq E[M_{0,n} - M_{1,n}; \{M_{0,n} > 0\}] \geq E[M_{0,n} - M_{1,n}] = 0.$$

The last inequality follows from the fact that on the set $\{M_{0,n} > 0\}^c$, $M_{0,n} = 0$ and therefore we have that $M_{0,n} - M_{1,n} \leq 0 - M_{1,n} \leq 0$. The last equality uses the stationarity of X. □

Theorem 6.11 (Birkhoff's ergodic theorem)**.** *Let X be a stationary process of integrable random variables, then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n} X_k = E[X_0|\mathcal{I}]$$

almost surely and in L^1.

Proof. By considering the stationary sequence

$$\tilde{X}_k = X_k - E[X_k|\mathcal{I}] = X_k - E[X_0|\mathcal{I}]$$

we can assume that $E[X_0|\mathcal{I}] = 0$.

Define

$$X^* = \limsup_{n\to\infty} \frac{1}{n} S_n \quad S_n = X_0 + \cdots + X_{n-1}.$$

Let $\epsilon > 0$ and define

$$D_\epsilon = \{X^* > \epsilon\}$$

and note that $D_\epsilon \in \mathcal{I}$.

Claim I. $P(D_\epsilon) = 0$.

Define

$$X_n^\epsilon = (X_n - \epsilon) I_{D_\epsilon}, \quad S_n^\epsilon = X_0^\epsilon + \cdots + X_{n-1}^\epsilon,$$

$$M_n^\epsilon = \max\{0, S_1^\epsilon, \ldots, S_n^\epsilon\}, \quad F_{\epsilon,n} = \{M_n^\epsilon > 0\}, \quad F_\epsilon = \bigcup_{n=1}^\infty F_{\epsilon,n} = \{\sup_{n\geq 1} \frac{1}{n} S_n^\epsilon > 0\}.$$

Note that $F_{\epsilon,1} \subset F_{\epsilon,2} \subset \cdots$ and that

$$F_\epsilon = \{\sup_{n\geq 1} \frac{1}{n} S_n > \epsilon\} \cap D_\epsilon = D_\epsilon.$$

The maximal ergodic theorem implies that

$$E[X_0^\epsilon; F_{\epsilon,n}] \geq 0.$$

Note that $E|X_0^\epsilon| < E|X_0| + \epsilon$. Thus, by the dominated donvergence theorem and the fact that $F_{\epsilon,n} \to F_n$, we have that

$$E[X_0^\epsilon; D_\epsilon] = E[X_0^\epsilon; F_\epsilon] = \lim_{n\to\infty} E[X_0^\epsilon; F_{\epsilon,n}] \geq 0$$

and that

$$E[X_0; F_\epsilon] = E[X_0; D_\epsilon] = E[E[X_0|\mathcal{I}]; D_\epsilon] = 0.$$

Consequently,

$$0 \leq E[X_0^\epsilon; D_\epsilon] = E[X_0; D_\epsilon] - \epsilon P(D_\epsilon) = -\epsilon P(D_\epsilon)$$

and the claim follows.

Because this holds for any $\epsilon > 0$, we have that

$$\limsup_{n\to\infty} \frac{1}{n} S_n \leq 0.$$

By considering the stationary process $-X$ we find that

$$\liminf_{n\to\infty} \frac{1}{n} S_n \geq 0.$$

This gives the desired almost sure convergence.

The convergence in L^1 follows from the following claim:

Claim II. $\{S_n/n; n \geq 1\}$ is uniformly integrable.

Note that by the stationarity of X

$$\lim_{m\to\infty}\sup_{k\geq 0} E[|X_k|;\{|X_k|\geq m\}] = \lim_{m\to\infty} E[|X_0|;\{|X_0|\geq m\}] = 0$$

by the dominated convergence theorem. Thus the sequence X is uniformly integrable.

For the sequence S_n/n,

$$E[|\frac{1}{n}S_n|] \leq \frac{1}{n}\sum_{k=0}^{n-1} E|X_k| = E|X_0| < \infty.$$

Let $\epsilon > 0$. Choose $\delta > 0$ so that

$$P(A_k) < \delta \quad \text{implies} \quad E[|X_k|;A_k] < \epsilon.$$

Now let B_n be a measurable set so that $P(B_n) < \delta$, then

$$E[|\frac{1}{n}S_n|;B_n] \leq \frac{1}{n}\sum_{k=0}^{n-1} E[|X_k|:B_n] < \frac{1}{n}n\epsilon = \epsilon.$$

The claim, and with it, the theorem, holds. □

Exercise 6.12. *Let X be stationary and let $g_n : S^{\mathbb{N}} \to \mathbb{R}$ satisfy $E[\sup_{n\geq 0}|g_n(\theta^n X)|] < \infty$. If the sequence $g_n(X) \to^{a.s.} g(X)$ as $n \to \infty$, then*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1} g_k(\theta^k X) = E[g(X)|\mathcal{I}] \quad a.s.$$

6.3 Ergodicity and Mixing

Definition 6.13. *A stationary process is called* ergodic *if the σ-algebra of invariant events $\mathcal{I}$ is trivial. In other words,*

1. *$P(B) = 0$ or 1 for all invariant sets B, or*
2. *$P\{X \in A\} = 0$ or 1 for all shift invariant sets A.*

If $\mathcal{I}$ is trivial, then $E[X_0|\mathcal{I}] = EX_0$ and the limit in Birkhoff's ergodic theorem is a constant.

Exercise 6.14. *Show that a stationary process X is ergodic if and only if for every $(k+1)$-dimensional measurable set A,*

$$\frac{1}{n}\sum_{j=0}^{n-1} I_A(X_j,\ldots,X_{j+k}) \to^{a.s.} P\{(X_0,\ldots,X_k)\in A\}.$$

Hint: *For one direction, consider shift invariant A.*

Exercise 6.15. *In example 6 above, take β to be rational. Show that T is not ergodic.*

Exercise 6.16. *The following steps can be used to show that the stationary process in example 6 is ergodic if β is irrational.*

1. *The values* $T^n\omega$ *are distinct.*
2. *For any* $n \in \mathbb{N}$, $|T^j\omega - T^k\omega| \le 1/n \mod 1$ *for some* $j, k \le n$.
3. $\{T^n\omega; n \ge 1\}$ *is dense in* $[0, 1)$.
4. *If* A *is any Borel set with positive Lebesgue measure, then for any* $\delta > 0$, *there exists an interval* I *so that* $P(A \cap I) > (1 - \delta)P(I)$
5. *If* A *is invariant, then* $P(A) = 1$.

Example 6.17. *(Weyl's equidistribution theorem) For the transformation above,*

$$\frac{1}{n}\sum_{k=0}^{n-1} I_A(T^n\omega) = P(A)$$

where P *is Lebesgue measure.*

Exercise 6.18. *Let* $\{X_n; n \ge 0\}$ *be a stationary and ergodic process and let* $\phi : S^{\mathbb{N}} \to \tilde{S}$ *be measurable. Then the process*

$$Y_k = \phi(X_k, X_{k+1}, \ldots) = \phi(\theta^k X).$$

is stationary and ergodic.

Definition 6.19. *A stationary process* X *is called* strongly mixing *if for any measurable set of sequences* A_1, A_2,

$$\lim_{n\to\infty} P\{X \in A_1, \theta^n X \in A_2\} = P\{X \in A_1\}P\{X \in A_2\}.$$

Remark 6.20. *By the Daniell-Kolmogorov extension theorem, it is sufficient to prove that for any* k *and any measurable subsets* A_1, A_2 *of* S^{k+1}

$$\lim_{n\to\infty} P\{(X_0, \ldots, X_k) \in A_1, (X_n, \ldots, X_{n+k}) \in A_2\} = P\{(X_0, \ldots, X_k) \in A_1\}P\{(X_0, \ldots, X_k) \in A_2\}.$$

Proposition 6.21. *Every strongly mixing stationary process is ergodic.*

Proof. Take $A = A_1 = A_2$ to be shift invariant. Thus,

$$\theta^n X \in A \quad \text{if and only if} \quad X \in A.$$

Apply the mixing property

$$P\{X \in A\} = \lim_{n\to\infty} P\{X \in A, \theta^n X \in A\} = P\{X \in A\}^2.$$

Thus, the only possibilities for $P\{X \in A\}$ are 0 and 1. □

Remark 6.22.

1. *We now see that independent and identically sequences are strongly mixing. Thus, the strong law is a consequence of the ergodic theorem.*

2. *For X an ergodic Markov chain with stationary distribution π, then for $n > k$*

$$P_\pi\{X_0 = x_0, \ldots, X_k = x_k, X_n = x_n, \ldots, X_{n+k} = x_{n+k}\}$$

$$= \pi\{x_0\}T(x_0, x_1)\cdots T(x_{k-1}, x_k)T^{n-k}(x_k, x_n)T(x_n, x_{n+1})\cdots T(x_{n+k-1}, x_{n+k}).$$

as $n \to \infty$, this expression converges to

$$\pi\{x_0\}T(x_0, x_1)\cdots T(x_{k-1}, x_k)\pi\{x_n\}T(x_n, x_{n+1})\cdots T(x_{n+k-1}, x_{n+k})$$

$$= P_\pi\{X_0 = x_0, \ldots, X_k = x_k\}P_\pi\{X_0 = x_n, \ldots, X_k = x_{n+k}\}.$$

Consequently, X is ergodic and the ergodic theorem holds for any ϕ satistying $E_\pi|\phi(X_0)| < \infty$. To obtain the limit for arbitrary initial distribution, choose a state x. Let X and $\tilde{X}$ be two independent Markov chains with common transition operator T. Let X have initial distribution π and $\tilde{X}$ have initial distribution α. Then,

$$Y_n = I_{\{x\}}(X_n) \quad \textit{and} \quad \tilde{Y}_n = I_{\{x\}}(\tilde{X}_n).$$

are both positive recurrent delayed renewal sequences. Set the coupling time

$$\tau = \min\{n; Y_n = \tilde{Y}_n = 1\}.$$

Then by the proof of the renewal theorem,

$$P\{\tau < \infty\} = 1.$$

Now define

$$Z_n = \begin{cases} \tilde{X}_n, & \textit{if } \tau > n \\ X_n, & \textit{if } \tau \le n \end{cases}$$

Then, by the strong Markov property, Z is a Markov process with transition operator T and initial distribution α. In other words, Z and $\tilde{X}$ have the same distribution.

We can obtain the result on the ergodic theorem for Markov chains by noting that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{n=0}^{n-1} f(Z_n) = \lim_{n\to\infty} \frac{1}{n}\sum_{n=0}^{n-1} f(X_n).$$

Exercise 6.23. *let X be a stationary Gaussian process with covariance function c. Show that*

$$\lim_{k\to\infty} c(k) = 0$$

implies that X is ergodic.

6.4 Entropy

For this section, we shall assume that $\{X_n; n \in \mathbb{Z}\}$ is a stationary and ergodic process with *finite* state space S

Definition 6.24. *Let Y be a random variable on a finite state space F.*

1. *Define the* mass function
$$p_Y(y) = P\{Y = y\}, \quad y \in F.$$

2. *Define the* entropy
$$H(Y) = -\sum_{y\in F} p_Y(y)\log p_Y(y).$$

Exercise 6.25. *Entropy has the following properties:*

1. $H(Y) \geq 0$.
2. *If* Y *is a constant random variable,* $H(Y) = 0$.
3. *If* F *has* n *outcomes, then the uniform distribution on* F *maximizes the entropy at*
$$-\sum_{y\in F}\frac{1}{n}\log\frac{1}{n} = \log n.$$

For a pair of random variables Y and $\tilde{Y}$ with respective finite state spaces F and $\tilde{F}$, define the *joint mass function*
$$p_{Y,\tilde{Y}}(y,\tilde{y}) = P\{Y = y, \tilde{Y} = \tilde{y}\}$$
and the *conditional mass function*
$$p_{Y|\tilde{Y}}(y|\tilde{y}) = P\{Y = y|\tilde{Y} = \tilde{y}\}.$$

With this we define the entropy
$$H(Y,\tilde{Y}) = -E[\log p_{Y,\tilde{Y}}(Y,\tilde{Y})] = -\sum_{(y,\tilde{y})\in F\times\tilde{F}} p_{Y,\tilde{Y}}(y,\tilde{y})\log p_{Y,\tilde{Y}}(y,\tilde{y})$$

and the *conditional entropy*,
$$\begin{aligned}
H(Y|\tilde{Y}) &= -E[\log p_{Y|\tilde{Y}}(Y|\tilde{Y})] \\
&= -\sum_{(y,\tilde{y})\in F\times\tilde{F}} p_{Y,\tilde{Y}}(y,\tilde{y})\log p_{Y|\tilde{Y}}(y|\tilde{y}) = -\sum_{(y,\tilde{y})\in F\times\tilde{F}} p_{Y,\tilde{Y}}(y,\tilde{y})\log\frac{p_{Y,\tilde{Y}}(y,\tilde{y})}{p_{\tilde{Y}}(\tilde{y})} \\
&= -\sum_{(y,\tilde{y})\in F\times\tilde{F}} p_{Y,\tilde{Y}}(y,\tilde{y})\log p_{Y,\tilde{Y}}(y,\tilde{y}) + \sum_{(y,\tilde{y})\in F\times\tilde{F}} p_{Y,\tilde{Y}}(y,\tilde{y})\log p_{\tilde{Y}}(\tilde{y}) \\
&= H(Y,\tilde{Y}) + \sum_{\tilde{y}\in\tilde{F}} p_{\tilde{Y}}(\tilde{y})\log p_{\tilde{Y}}(\tilde{y}) = H(Y,\tilde{Y}) - H(\tilde{Y})
\end{aligned}$$

or
$$H(Y,\tilde{Y}) = H(Y|\tilde{Y}) + H(\tilde{Y}).$$

We can also write
$$H(Y|\tilde{Y}) = -\sum_{(y,\tilde{y})\in F\times\tilde{F}} p_{\tilde{Y}}(\tilde{y})p_{Y|\tilde{Y}}(y|\tilde{y})\log p_{Y|\tilde{Y}}(y|\tilde{y}) = -\sum_{\tilde{y}\in\tilde{F}} p_{\tilde{Y}}(\tilde{y})\sum_{y\in F} p_{Y|\tilde{Y}}(y|\tilde{y})\log p_{Y|\tilde{Y}}(y|\tilde{y}).$$

Because the function $\phi(t) = t\log t$ is convex, we can use Jensen's inequality to obtain

$$\begin{aligned} H(Y|\tilde{Y}) &= -\sum_{y\in F}\sum_{\tilde{y}\in\tilde{F}} \phi(p_{Y|\tilde{Y}}(y|\tilde{y}))p_{\tilde{Y}}(\tilde{y}) \le -\sum_{y\in F}\phi\left(\sum_{\tilde{y}\in\tilde{F}} p_{\tilde{Y}}(\tilde{y})p_{Y|\tilde{Y}}(y|\tilde{y})\right) \\ &= -\sum_{\tilde{y}\in\tilde{F}}\phi(p_Y(y)) = -\sum_{\tilde{y}\in\tilde{F}} p_Y(y)\log p_Y(y) = H(Y) \end{aligned}$$

Therefore,

$$H(Y,\tilde{Y}) \le H(Y) + H(\tilde{Y})$$

with equality if and only if Y and $\tilde{Y}$ are independent.

Let's continue this line of analysis with $m+1$ random variables $X_0,\ldots,X_m$

$$\begin{aligned} H(X_0,\ldots,X_m) &= H(X_m|X_{m-1},\ldots,X_0) + H(X_0,\ldots,X_{m-1}) \\ &= H(X_m|X_{m-1},\ldots,X_0) + H(X_{m-1}|X_{m-2},\ldots,X_0) + H(X_0,\ldots,X_{m-2}) \\ &= H(X_m|X_{m-1},\ldots,X_0) + H(X_{m-1}|X_{m-2},\ldots,X_0) + \\ &\quad \cdots + H(X_2|X_0,X_1) + H(X_1|X_0) + H(X_0). \end{aligned}$$

This computation above shows that $H(X_m,\ldots,X_0)/m$ is the average of $\{H(X_k|X_{k-1},\ldots,X_0); 0\le k\le m\}$.

Also, by adapting the computation above, we can use Jensen's inequality to show that

$$H(X_m|X_{m-1},\ldots,X_0) \le H(X_m|X_{m-1},\ldots,X_1).$$

Now assume that $\{X_n; n\ge 1\}$ is stationary, then

$$H(X_{m-1}|X_{m-2},\ldots,X_0) = H(X_m|X_{m-1},\ldots,X_1) \ge H(X_m|X_{m-1},\ldots,X_0).$$

Thus, the sequence $\{H(X_m|X_{m-1},\ldots,X_0); m\ge 1\}$ is monotonically decreasing and nonnegative. We can use this to define the *entropy of the process.*

$$H(X) = \lim_{m\to\infty} H(X_m|X_{m-1},\ldots,X_0) = \lim_{m\to\infty}\frac{1}{m}H(X_m,\ldots,X_0).$$

The last equality expresses the fact that the Cesaro limit of a converging sequence is the same as the limit of the sequence.

This sets us up for the following theorem.

Theorem 6.26 (Shannon-McMillan-Breiman)**.** *Let X be a stationary and ergodic process having a finite state space, then*

$$\lim_{n\to\infty} -\frac{1}{n}\log p_{X_0,\ldots,X_n}(X_0,\ldots,X_{n-1}) = H(X) \quad a.s$$

Remark 6.27. *In words, this say that for a stationary process, the likelihood function*

$$p_{X_0,\ldots,X_n}(X_0,\ldots,X_n)$$

is for large n near to

$$\exp -nH(X).$$

Proof. By the basic properties of conditional probability,

$$p_{X_0,\ldots,X_{n-1}}(x_0,\ldots,x_{n-1}) = p_{X_0}(x_0)\prod_{k=1}^{n-1} p_{X_k|X_{k-1},\ldots,X_0}(x_k|x_{k-1},\ldots,x_0).$$

Consequently,

$$\frac{1}{n}\log p_{X_0,\ldots,X_{n-1}}(x_0,\ldots,x_{n-1}) = \frac{1}{n}\left(\log p_{X_0}(x_0) + \sum_{k=1}^{n}\log p_{X_k|X_{k-1},\ldots,X_0}(x_k|x_{k-1},\ldots x_0)\right).$$

Recall here that the stationary process is defined for all $n \in \mathbb{Z}$. Using the filtration $\mathcal{F}_n^k = \sigma\{X_{k-j}; 1 \leq j \leq n\}$, we have, by the martingale convergence theorem, that the uniformly integrable martingale

$$\lim_{n\to\infty} p_{X_k|X_{k-1},\ldots,X_{k-n}}(X_k|X_{k-1},\ldots X_{k-n}) = p_{X_k|X_{k-1},\ldots}(X_k|X_{k-1},\ldots)$$

almost surely and in L^1.

By the ergodic theorem,

$$\lim_{n\to\infty} -\frac{1}{n}\sum_{k=1}^{n-1}\log p_{X_k|X_{k-1},\ldots}(X_k|X_{k-1},\ldots) = -E[\log p_{X_0|X_{-1},\ldots}(X_0|X_{-1},\ldots)] \text{ a.s. and in } L^1.$$

Now,

$$\begin{aligned} H(X) &= \lim_{m\to\infty} H(X_m|X_{m-1},\ldots,X_0) = \lim_{m\to\infty} -E[\log p_{X_m|X_{m-1},\ldots,X_0}(X_m|X_{m-1},\ldots,X_0)] \\ &= \lim_{m\to\infty} -E[\log p_{X_0|X_{-1},\ldots,X_{-m}}(X_0|X_{-1},\ldots,X_{-m})] = -E[\log p_{X_0|X_{-1},\ldots}(X_0|X_{-1},\ldots)]. \end{aligned}$$

The theorem now follows from the exercise following the proof of the ergodic theorem. Take

$$g_k(X) = -\log p_{X_0|X_{-1},\ldots,X_{-k}}(X_0|X_{-1}\ldots,X_{-k}).$$

and

$$g(X) = -\log p_{X_0|X_{-1},\ldots}(X_0|X_{-1}\ldots).$$

Then,

$$g_k(\theta^k X) = -\log p_{X_k|X_{k-1},\ldots,X_0}(X_k|X_{k-1}\ldots,X_0),$$

and

$$\begin{aligned} \lim_{n\to\infty} -\frac{1}{n}\log p_{X_0,\ldots,X_{n-1}}(X_0,\ldots,X_{n-1}) &= \lim_{n\to\infty} -\frac{1}{n}\left(\log p_{X_0}(X_0) + \sum_{k=1}^{n}\log p_{X_k|X_{k-1},\ldots,X_0}(X_k|X_{k-1},\ldots X_0)\right). \\ &= H(X) \end{aligned}$$

almost surely. □

Example 6.28. *For an ergodic Markov chain having transition operator T in its stationary distribution π,*

$$\begin{aligned} H(X_m|X_{m-1},\ldots,X_0) &= H(X_m|X_{m-1}) \\ &= -\sum_{x,\tilde{x}\in S} P\{X_m=\tilde{x}, X_{m-1}=x\}\log P\{X_m=\tilde{x}|X_{m-1}=x\} \\ &= -\sum_{x,\tilde{x}\in S}\pi\{x\}T(x,\tilde{x})\log T(x,\tilde{x}). \end{aligned}$$