Probability and Stochastic Processes
FINM 34000

1 Introduction

These notes are from FINM 34000, a half-quarter course in the Masters of Financial Mathematics program and are on stochastic processes. The prerequisite is an undergraduate course in probability and the natural follow-up to this is a course on stochastic calculus, also at the masters level.

A stochastic process is a collection of random variables X_t indexed by time. We view X_t as the value of some quantity at time t. There are various possibilities for the set of times. We will be looking at only three cases:

- Discrete with increments of 1: $t = 0, 1, 2, 3, \ldots$.
- Discrete with increments of Δt for some time increment Δt: $t = 0, \Delta t, 2\,\Delta t, 3\Delta t, \ldots$.
- Continuous: $t \in [0, \infty)$

In this course we will be considering several examples that are either discrete time or are continuous time but "purely jump". When Δt is very small, then the discrete process is much like a continuous time process and at the end we will discuss one important continuous time process, Brownian motion. Brownian motion and the corresponding stochastic calculus are the topic for FINM 34500 and will not be discussed much in this course.

2 Conditional expectation

If X is a random variable, then its expectation, $\mathbb{E}[X]$ can be thought of as the best guess for X given no information about the result of the trial. A conditional expectation can be considered as the best guess given some but not total information.

Let $X_1, X_2, \ldots$ be random variables which we think of as a time series with the data arriving one at a time. At time n we have viewed the values $X_1, \ldots, X_n$. If Y is another random variable, then $E(Y \mid X_1, \ldots, X_n)$ is the best guess for Y given $X_1, \ldots, X_n$. We will assume that Y is an *integrable* random variable which means $\mathbb{E}[|Y|] < \infty$. To save some space we will write $\mathcal{F}_n$ for "the information contained in $X_1, \ldots, X_n$" and $E[Y \mid \mathcal{F}_n]$ for $E[Y \mid X_1, \ldots, X_n]$. We view $\mathcal{F}_0$ as no information. The best guess should satisfy the following properties.

- If we have no information, then the best guess is the expected value. In other words, $E[Y \mid \mathcal{F}_0] = \mathbb{E}[Y]$.

- The conditional expectation $E[Y \mid \mathcal{F}_n]$ should only use the information available at time n. In other words, it should be a function of $X_1, \ldots, X_n$,

$$E[Y \mid \mathcal{F}_n] = \phi(X_1, \ldots, X_n).$$

  We say that $E[Y \mid \mathcal{F}_n]$ is $\mathcal{F}_n$*-measurable*.

The definitions in the last paragraph are certainly vague. We can use measure theory to be precise. We assume that the random variables $Y, X_1, X_2, \ldots$ are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Here $\mathcal{F}$ is a σ-*algebra* or σ-*field* of subsets of Ω, that is, a collection of subsets satisfying

- $\emptyset \in \mathcal{F}$;
- $A \in \mathcal{F}$ implies that $\Omega \setminus A \in \mathcal{F}$;
- $A_1, A_2, \ldots \in \mathcal{F}$ implies that $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

The information $\mathcal{F}_n$ is the smallest sub σ-algebra $\mathcal{G}$ of $\mathcal{F}$ such that $X_1, \ldots, X_n$ are $\mathcal{G}$-measurable. The latter statement means that for all $t \in \mathbb{R}$, the event $\{X_j \leq t\} \in \mathcal{F}_n$. The "no information" σ-algebra $\mathcal{F}_0$ is the trivial σ-algebra containing only $\emptyset$ and Ω.

The definition of conditional expectation is a little tricky, so let us try to motivate it by considering some examples.

2.1 Conditional expectation with respect to a partition

Suppose that $\mathcal{A} = \{A_1, A_2, \ldots\}$ is a partition of the sample space Ω, that is, exactly one of the events $A_1, A_2, \ldots$ occurs, or in the language of sets,

$$A_j \cap A_k = \emptyset \quad \text{if } j \neq k, \qquad \bigcup_{n=1}^{\infty} A_n = \Omega.$$

We can also view $\mathcal{A}$ as a source of partial information: we will be told which of the events $A_1, A_2, \ldots$ occur but not the exact outcome. If X is a random variable, we want to define $E(X \mid \mathcal{A})$. As a simple example, suppose we roll two fair dice, one red and one green, and we let $\mathcal{A}$ denote the information obtained by seeing the value on the green die, that is, the partition

$$A_j = \{\text{green die comes up } j\}.$$

We let X be the sum of the two rolls.

One of the properties is that $E(X \mid \mathcal{A})$ should be $\mathcal{A}$-measurable. This means that we only need to know which partition event has occurred in order determine $E(X \mid \mathcal{A})$. Another way of saying this is that $E(X \mid \mathcal{A})$ is constant on each A_j. A little thought shows that we should choose

$$E(X \mid \mathcal{A}) = \frac{\mathbb{E}[X\, 1_{A_j}]}{\mathbb{P}(A_j)} \quad \text{on} \quad A_j.$$

This way the conditional expectation satisfies the law of total expectation:

$$\mathbb{E}[X] = \sum_{j=1}^{\infty} \mathbb{E}[X\, 1_{A_j}] = \sum_{j=1}^{\infty} \mathbb{P}(A_j)\, E[X \mid A_j] = \mathbb{E}\left(E[X \mid \mathcal{A}]\right).$$

It also satisfies a stronger version. Let us call an event V $\mathcal{A}$-measurable if it can be written as the (finite or countable) union of sets in the partition. Then

$$\mathbb{E}[X\, 1_V] = \mathbb{E}\left[E(X \mid \mathcal{A})\, 1_V\right].$$

In our dice rolling example, we leave it to you to check that

$$E(X \mid \mathcal{A}) = j + \frac{7}{2} \quad \text{on} \;\; A_j.$$

2.2 Conditional expectation for continuous random variables

Suppose that (X, Y) have a joint density

$$f(x, y), \quad 0 < x, y < \infty,$$

with marginal densities

$$f(x) = \int_{-\infty}^{\infty} f(x, y)\, dy, \qquad g(y) = \int_{-\infty}^{\infty} f(x, y)\, dx.$$

The conditional density $f(y|x)$ is defined by

$$f(y|x) = \frac{f(x, y)}{f(x)}.$$

This is well defined provided that $f(x) > 0$, and if $f(x) = 0$, then x is an "impossible" value for X to take. We can write

$$\mathbb{E}[Y \mid X = x] = \int_{-\infty}^{\infty} y\, f(y \mid x)\, dy.$$

We can use this as the definition of conditional expectation in this case,

$$E[Y \mid X] = \int_{-\infty}^{\infty} y\, f(y \mid X)\, dy = \frac{\int_{-\infty}^{\infty} y\, f(X, y)\, dy}{f(X)}.$$

Note that $E[Y \mid X]$ is a random variable which is determined by the value of the random variable X. Since it is a random variable, we can take its expectation

$$\begin{aligned}
\mathbb{E}\left[E[Y \mid X]\right] &= \int_{-\infty}^{\infty} \mathbb{E}[Y \mid X = x]\, f(x)\, dx \\
&= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} y\, f(y \mid x)\, dy\right] f(x)\, dx \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y\, f(x, y)\, dy\, dx \\
&= \mathbb{E}[Y].
\end{aligned}$$

Similarly, suppose that A is an event that only depends on X, say $a \leq X \leq b$. We can show that

$$\mathbb{E}\left[E[Y \mid X]\, 1_A\right] = \mathbb{E}[Y\, 1_A].$$

2.3 General definition

These calculation illustrate basic properties of conditional expectation. Suppose we are interested in the value of a random variable Y and we are going to be given data $X_1, \ldots, X_n$. Once we observe the data, we make our best prediction $E[Y \mid X_1, \ldots, X_n]$. If we average our best prediction given $X_1, \ldots, X_n$ over all the possible values of $X_1, \ldots, X_n$, we get the best prediction of Y. In other words,

$$\mathbb{E}[Y] = \mathbb{E}\left[E[Y \mid \mathcal{F}_n]\right].$$

More generally, suppose that A is an $\mathcal{F}_n$-measurable event, that is to say, if we observe the data $X_1, \ldots, X_n$, then we know whether or not A has occurred. An example of an $\mathcal{F}_4$-measurable event would be

$$A = \{X_1 \geq X_2, X_4 < 4\}.$$

Using similar reasoning, we can see that if A is $\mathcal{F}_n$-measurable, then

$$\mathbb{E}\left[Y\, 1_A\right] = \mathbb{E}\left[E[Y \mid \mathcal{F}_n]\, 1_A\right].$$

At this point, we have not derived this relation mathematically; in fact, we have not even defined the conditional expectation. Instead, we will use this reasoning to motivate the following definition.

Definition The conditional expectation $E[Y \mid \mathcal{F}_n]$ is the unique random variable satisfying the following.

- $E[Y \mid \mathcal{F}_n]$ is $\mathcal{F}_n$-measurable.

- For every $\mathcal{F}_n$-measurable event A,

$$\mathbb{E}\left[E[Y \mid \mathcal{F}_n]\, 1_A\right] = \mathbb{E}\left[Y\, 1_A\right].$$

We have used different fonts for the E of conditional expectation and the $\mathbb{E}$ of usual expectation in order to emphasize that the conditional expectation is a random variable. However, most authors use the same font for both leaving it up to the reader to determine which is being referred to. Another notation that is sometimes used is

$$\mathbb{E}_n[X] = E(X \mid \mathcal{F}_n).$$

Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and Y is an integrable random variable. Suppose $\mathcal{G}$ is a sub σ-algebra of $\mathcal{F}$. Then $E[Y \mid \mathcal{G}]$ is defined to be the unique (up to an event of measure zero) $\mathcal{G}$-measurable random variable such that if $A \in \mathcal{G}$,

$$\mathbb{E}\left[Y\, 1_A\right] = \mathbb{E}\left[E[Y \mid \mathcal{G}]\, 1_A\right].$$

Uniqueness follows from the fact that if Z_1, Z_2 are $\mathcal{G}$-measurable random variables with

$$\mathbb{E}\left[Z_1\, 1_A\right] = \mathbb{E}\left[Z_2\, 1_A\right]$$

for all $A \in \mathcal{G}$, then $\mathbb{P}\{Z_1 \neq Z_2\} = 0$. Existence of the conditional expectation can be deduced from the Radon-Nikodym theorem. Let $\mu(A) = \mathbb{E}\left[Y 1_A\right]$. Then μ is a (signed) measure on $(\Omega, \mathcal{G}, \mathbb{P})$ with $\mu \ll \mathbb{P}$, and hence there exists an L^1 random variable Z with $\mu(A) = \mathbb{E}\left[Z 1_A\right]$ for all $A \in \mathcal{G}$.

Although the definition does not give an immediate way to calculate the conditional expectation, in many cases one can compute it. We will give a number of properties of the conditional expectation most of which follow quickly from the definition.

Proposition 2.1. *Suppose $X_1, X_2, \ldots$ is a sequence of random variables and $\mathcal{F}_n$ denotes the information at time n. The conditional expectation $E[Y \mid \mathcal{F}_n]$ satisfies the following properties.*

- *If Y is $\mathcal{F}_n$-measurable, then $E[Y \mid \mathcal{F}_n] = Y$.*
- *If A is an $\mathcal{F}_n$-measurable event, then $\mathbb{E}\left[E[Y \mid \mathcal{F}_n]\, 1_A\right] = \mathbb{E}\left[Y\, 1_A\right]$. In particular,*

$$\mathbb{E}\left[E[Y \mid \mathcal{F}_n]\right] = \mathbb{E}[Y].$$

- *Suppose $X_1, \ldots, X_n$ are independent of Y. Then $\mathcal{F}_n$ contains no useful information about Y and hence*

$$E[Y \mid \mathcal{F}_n] = \mathbb{E}[Y].$$

- *Linearity. If Y, Z are random variables and a, b are constants, then*

$$E[aY + bZ \mid \mathcal{F}_n] = a\, E[Y \mid \mathcal{F}_n] + b\, E[Z \mid \mathcal{F}_n]. \tag{1}$$

- *Projection or Tower Property. If $m < n$, then*

$$E\,[\,E[Y \mid \mathcal{F}_n] \mid \mathcal{F}_m\,] = E[Y \mid \mathcal{F}_m]. \tag{2}$$

- *If Z is an $\mathcal{F}_n$-measurable random variable, then when conditioning with respect to $\mathcal{F}_n$, Z acts like a constant,*

$$E[YZ \mid \mathcal{F}_n] = Z\, E[Y \mid \mathcal{F}_n]. \tag{3}$$

The proof of this proposition is not very difficult given our choice of definition for the conditional expectation. We will discuss only a couple of cases here, leaving the rest for the reader. To prove the linearity property, we know that $a\, E[Y \mid \mathcal{F}_n] + b\, E[Z \mid \mathcal{F}_n]$ is an $\mathcal{F}_n$-measurable random variable. Also if $A \in \mathcal{F}_n$, then linearity of expectation implies that

$$\begin{aligned}
&\mathbb{E}\left[1_A \left(a\, E[Y \mid \mathcal{F}_n] + b\, E[Z \mid \mathcal{F}_n]\right)\right] \\
&\quad = a\mathbb{E}\left[1_A\, E[Y \mid \mathcal{F}_n]\right] + b\, \mathbb{E}\left[1_A\, E[Z \mid \mathcal{F}_n]\right] \\
&\quad = a\mathbb{E}\left[1_A\, Y\right] + b\mathbb{E}\left[1_A\, Z\right] \\
&\quad = \mathbb{E}\left[1_A\, (aY + bZ)\right].
\end{aligned}$$

Uniqueness of the conditional expectation then implies (1).

We first show the "constants rule" (3) for $Z = 1_A, A \in \mathcal{F}_n$, as follows. If $B \in \mathcal{F}_n$, then $A \cap B \in \mathcal{F}_n$ and

$$\mathbb{E}\left[1_B\, E(YZ \mid \mathcal{F}_n)\right] = \mathbb{E}\left[1_B\, E(1_A\, Y \mid \mathcal{F}_n)\right]$$

$$= \mathbb{E}\left[1_B\, 1_A\, Y\right] = \mathbb{E}\left[1_{A\cap B} Y\right] = \mathbb{E}\left[1_{A\cap B} E(Y \mid \mathcal{F}_n)\right]$$

$$= \mathbb{E}\left[1_B\, 1_A\, E(Y \mid \mathcal{F}_n)\right] = \mathbb{E}\left[1_B\, Z\, E(Y \mid \mathcal{F}_n)\right].$$

Hence $E(YZ \mid \mathcal{F}_n) = Z\, E(Y \mid \mathcal{F}_n)$ by definition. Using linearity, the rule holds for simple random variables of the form

$$Z = \sum_{j=1}^{n} a_j\, 1_{A_j}, \qquad A_j \in \mathcal{F}_n.$$

We can then prove it for nonnegative Y by approximating from below by nonnegative simple random variables and using the monotone convergence theorem, and then for general Y by writing $Y = Y^+ - Y^-$. These are standard techniques in Lebesgue integration theory.

Example 2.1. Suppose that $X_1, X_2, \ldots$ are independent random variables with $\mathbb{E}[X_j] = \mu$ for each j. Let $S_n = X_1 + \cdots + X_n$, and let $\mathcal{F}_n$ denote the information contained in $X_1, \ldots, X_n$. Then if $m < n$,

$$\begin{aligned} E[S_n \mid \mathcal{F}_m] &= E[S_m \mid \mathcal{F}_m] + E[S_n - S_m \mid \mathcal{F}_m] \\ &= S_m + \mathbb{E}[S_n - S_m] \\ &= S_m + \mu\,(n-m). \end{aligned}$$

Example 2.2. In the same setup as Example 2.1 suppose that $\mu = 0$ and $\mathbb{E}[X_j^2] = \sigma^2$ for each j. Then if $m < n$,

$$\begin{aligned} E[S_n^2 \mid \mathcal{F}_m] &= E([S_m + (S_n - S_m)]^2 \mid \mathcal{F}_m) \\ &= E[S_m^2 \mid \mathcal{F}_m] + 2\,E[S_m(S_n - S_m) \mid \mathcal{F}_m] \\ &\qquad + E[(S_n - S_m)^2 \mid \mathcal{F}_m]. \end{aligned}$$

Since S_m is $\mathcal{F}_m$-measurable and $S_n - S_m$ is independent of $\mathcal{F}_m$,

$$E[S_m^2 \mid \mathcal{F}_m] = S_m^2,$$

$$E[S_m(S_n - S_m) \mid \mathcal{F}_m] = S_m\,E[S_n - S_m \mid \mathcal{F}_m] = S_m\,\mathbb{E}[S_n - S_m] = 0,$$

$$E[(S_n - S_m)^2 \mid \mathcal{F}_m] = \mathbb{E}[(S_n - S_m)^2] = \mathrm{Var}(S_n - S_m) = \sigma^2\,(n-m),$$

and hence,

$$E[S_n^2 \mid \mathcal{F}_m] = S_m^2 + \sigma^2\,(n-m).$$

Example 2.3. In the same setup as Example 2.1, let us also assume that $X_1, X_2, \ldots$ are identically distributed. We will compute $E[X_1 \mid S_n]$. Note that the information contained in the one data point S_n is less than the information contained in $X_1, \ldots, X_n$. However, since the random variables are identically distributed, it must be the case that

$$E[X_1 \mid S_n] = E[X_2 \mid S_n] = \cdots = E[X_n \mid S_n].$$

Linearity implies that

$$n\,E[X_1 \mid S_n] = \sum_{j=1}^{n} E[X_j \mid S_n] = E[X_1 + \cdots + X_n \mid S_n] = E[S_n \mid S_n] = S_n.$$

Therefore,

$$E[X_1 \mid S_n] = \frac{S_n}{n}.$$

It may be at first surprising that the answer does not depend on $\mathbb{E}[X_1]$.

Definition If $X_1, X_2, \ldots$ is a sequence of random variables, then the associated (discrete time) *filtration* is the collection $\{\mathcal{F}_n\}$ where $\mathcal{F}_n$ denotes the information in $X_1, \ldots, X_n$.

One assumption in the definition of a filtration, which may sometimes not reflect reality, is that information is never lost. If $m < n$, then everything known at time m is still known at time n. Sometimes a filtration is given starting at time $n = 1$ and sometimes starting at $n = 0$. If it starts at time $n = 1$, we define $\mathcal{F}_0$ to be "no information".

More generally, a *(discrete time) filtration* $\{\mathcal{F}_n\}$ is an increasing sequence of σ-algebras.

Exercises

Exercise 2.1. Suppose (X, Y) are discrete random variables with joint probabilities

$$\begin{array}{cc} X\backslash Y & \begin{array}{cccc} \;1\; & \;2\; & \;3\; & \;4\; \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{pmatrix} .01 & .1 & .2 & .09 \\ .1 & .05 & .05 & 0 \\ .1 & .1 & .1 & .1 \end{pmatrix} \end{array}$$

For example, $\mathbb{P}\{X = 2, Y = 3\} = .05$.

1. Find the marginal distributions for X and Y.
2. Find $\mathbb{E}[X], \mathbb{E}[Y], E(X \mid Y), E(Y \mid X)$ and use these to check directly that
$$\mathbb{E}[X] = \mathbb{E}[E(X \mid Y)], \quad \mathbb{E}[Y] = \mathbb{E}[E(Y \mid X)].$$
3. Let A be the event $A = \{Y \text{ is even}\}$. Which of the following facts hold?
$$\mathbb{E}[E(X|Y)\, 1_A] = \mathbb{E}[X\, 1_A], \qquad \mathbb{E}[E(Y|X)\, 1_A] = \mathbb{E}[Y\, 1_A].$$

Exercise 2.2. Suppose we roll two dice, a red and a green one, and let X be the value on the red die and Y the value on the green die. Let $Z = XY$.

1. Let $W = E(Z \mid X)$. What are the possible values for W? Give the distribution of W.
2. Do the same exercise for $U = E(X \mid Z)$.
3. Do the same exercise for $V = E(Y \mid X, Z)$

Exercise 2.3. Suppose we roll two dice, a red and a green one, and let X be the value on the red die and Y the value on the green die. Let $Z = X/Y$.

1. Find $E[(X+2Y)^2 \mid X]$.
2. Find $E[(X+2Y)^2 \mid X, Z]$.
3. Let $W = E[Z \mid X]$. What are the possible values for W? Give the distribution of W.

Exercise 2.4. Suppose $X_1, X_2, \ldots$ are independent random variables with

$$\mathbb{P}\{X_j = 2\} = 1 - \mathbb{P}\{X_j = -1\} = \frac{1}{3}.$$

Let $S_n = X_1 + \cdots + X_n$ and let $\mathcal{F}_n$ denote the information in $X_1, \ldots, X_n$.

1. Find $\mathbb{E}[X_1], \mathbb{E}[X_1^2], \mathbb{E}[X_1^3]$.
2. Find $\mathbb{E}[S_n], \mathbb{E}[S_n^2], \mathbb{E}[S_n^3]$.
3. If $m < n$, find
$$E[S_n \mid \mathcal{F}_m], \quad E[S_n^2 \mid \mathcal{F}_m], \quad E[S_n^3 \mid \mathcal{F}_m].$$
4. If $m < n$, find $E[X_m \mid S_n]$.

Exercise 2.5. Repeat Exercise 2.4 assuming that

$$\mathbb{P}\{X_j = 3\} = \mathbb{P}\{X_j = -1\} = \frac{1}{2}.$$

Exercise 2.6. Let $X_1, X_2, \ldots$ be independent, identically distributed random variables with mean μ. Let $S_n = X_1 + \cdots + X_n$. Find

$$\mathbb{E}[X_1 \mid S_n, X_n].$$

3 Simple (symmetric) random walk

Here we will study a simple but fundamental model for random movement. Let $X_1, X_2, \ldots$ be independent "coin-flipping" random variables

$$\mathbb{P}\{X_j = 1\} = \mathbb{P}\{X_j = -1\} = \frac{1}{2}.$$

We call the sum

$$S_n = X_1 + \cdots + X_n$$

(one-dimensional) simple (symmetric) random walk. Here we are assuming that $S_0 = 0$; we could also choose other initial conditions. It is closely related to the binomial model in financial mathematics. Note that $\mathbb{E}[X_j] = 0, \mathrm{Var}[X_j] = \mathbb{E}[X_j^2] = 1$ and hence

$$\mathbb{E}[S_n] = 0, \quad \mathrm{Var}[S_n] = \mathrm{Var}[X_1 + \cdots + X_n] = n.$$

Our goal in this section is to answer the following questions.

- How far does one go in n steps?
- More generally, what is the distribution of S_n?
- What is the probability of being at the origin after n steps?
- Does one return to the origin infinitely often? (The answer will be yes.)
- If a, b are integers, what is the probability that we reach b before we reach $-a$?
- If we allow the random walker to walk in more than one dimension, do we still return to the origin infinitely often?

We know that $\mathrm{Var}[S_n] = n$ and hence the typical distance is of order $\sqrt{n}$. More precisely, the central limit theorem tells us that the distribution of $S_n/\sqrt{n}$ approaches a standard normal,

$$\lim_{n\to\infty} \mathbb{P}\left\{a \le \frac{S_n}{\sqrt{n}} \le b\right\} = \int_a^b \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}\, dx.$$

Can we be more precise than this? We first note that if n is odd, S_n is odd while if n is even, S_n is even. Let us consider $\mathbb{P}\{S_{2n} = 2k\}$ where k is an integer. If life is good, we can use the normal approximation to say that

$$\begin{aligned}
\mathbb{P}\{S_{2n} = 0\} &\sim \mathbb{P}\{-1 \le S_{2n} \le 1\} \\
&= \mathbb{P}\left\{-\frac{1}{\sqrt{2n}} \le \frac{S_{2n}}{\sqrt{2n}} \le \frac{1}{\sqrt{2n}}\right\} \\
&\sim \int_{-1/\sqrt{2n}}^{1/\sqrt{2n}} \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}\, dx \\
&\sim \frac{2}{\sqrt{2n}} \quad \frac{1}{\sqrt{2\pi}} \\
&\sim \frac{1}{\sqrt{\pi n}}.
\end{aligned}$$

There is another (more rigorous) way to derive this using the binomial theorem. In order for $S_{2n} = 0$ we must have exactly n $+1$s and exactly n -1s in $2n$ trials. Using the binomial distribution we get an exact expression for the probability

$$\mathbb{P}\{S_{2n} = 0\} = \binom{2n}{n} \left(\frac{1}{2}\right)^n \left(1 - \frac{1}{2}\right)^n = \frac{(2n)!}{2^{2n}\, n!\, n!}.$$

The right-hand side is exact, but it is not very illuminating. The numerator and denominator are both very large and it is not easy so see what the ratio is as $n \to \infty$. However, using Stirling's formula which estimates the factorial one can show that the right-hand side is indeed asymptotic to $(\pi n)^{-1/2}$.

Stirling's formula states that as $n \to \infty$,

$$n! \sim \sqrt{2\pi}\, n^{n+\frac{1}{2}}\, e^{-n},$$

where the $\sim$ sign means asymptotic, that is, the ratio of the two sides goes to one. We will not prove this here but if we plug it in we see that

$$\frac{(2n)!}{2^{2n}\, n!\, n!} \sim \frac{\sqrt{2\pi}\,(2n)^{n+\frac{1}{2}}\, e^{-2n}}{2^{2n}\,(\sqrt{2\pi}\, n^{n+\frac{1}{2}}\, e^{-n})^2} = \frac{1}{\sqrt{\pi\, n}}.$$

More generally, if we define r by $2k = r\sqrt{2n}$, we might guess that

$$\begin{aligned}
\mathbb{P}\{S_{2n} = 2k\} &= \mathbb{P}\{2k-1 \le S_{2n} \le 2k+1\} \\
&= \mathbb{P}\left\{\frac{2k-1}{\sqrt{2n}} \le \frac{S_{2n}}{\sqrt{2n}} \le \frac{2k+1}{\sqrt{2n}}\right\} \\
&\sim \int_{(2k-1)/\sqrt{2n}}^{(2k+1)/\sqrt{2n}} \frac{1}{\sqrt{2\pi}} e^{-(2k/\sqrt{2n})^2/2} \\
&= \frac{2}{\sqrt{2n}}\,\frac{1}{\sqrt{2\pi}} e^{-r^2/2} = \frac{1}{\sqrt{\pi n}}\, e^{-r^2/2}.
\end{aligned}$$

We can derive this using the binomial distribution and Stirling's formula. Note that $S_{2n} = 2k$ means that there are $(n+k)$ $+1$'s and $(n-k)$ (-1)'s. Hence

$$\mathbb{P}\{S_{2n} = n+k\} = \binom{2n}{n+k} \left(\frac{1}{2}\right)^{n+k} \left(\frac{1}{2}\right)^{n-k} = \frac{(2n)!}{4^n\,(n+k)!\,(n-k)!}.$$

If we use Stirling's formula, we see that the right hand side is asymptotic to

$$\frac{\sqrt{2\pi}\, e^{-2n}\,(2n)^{2n+\frac{1}{2}}}{4^n\,(\sqrt{2\pi}\, e^{-(n+k)}\,(n+k)^{n+k+\frac{1}{2}})\,(\sqrt{2\pi}\, e^{-(n-k)}\,(n-k)^{n-k+\frac{1}{2}})},$$

which simplifies to

$$\frac{1}{\sqrt{\pi}} \sqrt{\frac{n}{(n+k)\,(n-k)}} \left(1+\frac{k}{n}\right)^{-k} \left(1-\frac{k}{n}\right)^{k} \left(1-\frac{k^2}{n}\right)^{-n}.$$

If we now let $2k = r\sqrt{2n}$, we get

$$\sqrt{\frac{n}{(n+k)\,(n-k)}} \sim \frac{1}{\sqrt{n}},$$

$$\left(1+\frac{k}{n}\right)^{-k} = \left(1+\frac{r}{\sqrt{2n}}\right)^{-r\sqrt{2n}/2} = \left[\left(1+\frac{r}{\sqrt{2n}}\right)^{\sqrt{2n}}\right]^{-r/2} \sim e^{-r^2/2}$$

$$\left(1-\frac{k}{n}\right)^{k} = \left(1-\frac{r}{\sqrt{2n}}\right)^{r\sqrt{2n}/2} = \left[\left(1-\frac{r}{\sqrt{2n}}\right)^{\sqrt{2n}}\right]^{r/2} \sim e^{-r^2/2}$$

$$\left(1-\frac{k^2}{n}\right)^{-n}=\left(1-\frac{r^2}{n}\right)^{n}\sim e^{r^2/2}.$$

This gives

$$\mathbb{P}\{S_{2n}=2k\}\sim\frac{1}{\sqrt{\pi n}}\,e^{-r^2/2},\qquad \text{where}\quad r=\frac{2k}{\sqrt{2n}}.$$

Let us consider another problem. Suppose we start our random walker at point one and let it walk until it reaches 0 or N? What is the probability that it reaches N first? To answer this question it is easier to answer a larger question: let $q(k)$ be the probability that the walker reaches N before 0 assuming it starts at point k. Note that if $1\leq k\leq N-1$, then by considering the first step of the walker we get

$$q(k)=\frac{1}{2}\,q(k-1)+\frac{1}{2}\,q(k+1),\qquad k=1,2,\ldots,N-1. \tag{4}$$

We also have $q(0)=0, q(N)=1$. Note that (4) gives $N-1$ linear equations in $N-1$ unknowns $(q(1),q(2),\ldots,q(N-1))$. This suggests that there is a unique solution, and indeed we will show this now. Let $q=q(1)$. Then since

$$q(1)=\frac{1}{2}\,q(0)+\frac{1}{2}\,q(2)=\frac{1}{2}\,q(2)$$

we get $q(2)=2q$. Similarly, by continuing this we get that for all k, $q(k)=kq$ and since $q(N)=1$ we must have $q=1/N$. We have established an estimate for random walk that goes under a fun name.

Fact Gambler's Ruin Estimate If a,b are positive integers, then the probability that a simple random walk starting at the origin reaches b before reaching $-a$ is

$$\frac{a}{a+b},$$

and hence the probability to reach $-a$ first is

$$1-\frac{a}{a+b}=\frac{b}{a+b}.$$

There is another way of thinking about this which will give us a first look at "martingale" reasoning. Flipping coins should give a fair game. If I flip coins until I get either b heads or $-a$ tails, and I receive \$$b$ in the first case and lose \$$a$ in the second case, then this should be a fair game. The expected winning in this game is

$$b\cdot\frac{a}{a+b}-a\cdot\frac{b}{a+b}=0.$$

We can also ask: what is the expected number of steps the walker starting at the origin takes until getting to $-N$ or N? Let $e(k)$ be the expected number of steps assuming the walker starts at k. Then $e(-N) = 0, e(N) = 0$ and

$$e(k) = 1 + \frac{1}{2}\, e(k-1) + \frac{1}{2}\, e(k+1), \quad k = -(N-1), -(N-2), , \ldots, N-2, N-1.$$

The 1 corresponds to the first step and the expected number of steps after that corresponds to the expected number of steps starting at the place one moves on the first step. Again there is a unique solution to this but we will just give it

$$e(k) = (N+k)\,(N-k).$$

It is straightforward to show that this works.

We will now show that the random walker returns infinitely often to the origin. Let V be the total number of visits to the origin,

$$V = \sum_{j=0}^{\infty} I_j$$

where I_j is the indicator function of the event that the walker is at the origin at time j. Recall that $\mathbb{E}[I_j] = \mathbb{P}\{S_j = 0\}$ and since $\mathbb{P}\{S_j = 0\} = 0$ if j is odd, we can write

$$\mathbb{E}[V] = \sum_{j=0}^{\infty} \mathbb{P}\{S_{2j} = 0\}.$$

We have seen that

$$\mathbb{P}\{S_{2j} = 0\} \sim \frac{c}{n^{1/2}},$$

where $c = 1/\sqrt{\pi}$. This tells us that

$$\mathbb{E}[V] = \infty.$$

Recall from calculus that

$$\sum_{n=1}^{\infty} n^{-p} < \infty$$

if and only if $p > 1$.

However, we cannot say that $V = \infty$ just because $\mathbb{E}[V] = \infty$. In order to finish this we will compute $\mathbb{E}[V]$ in another way. Let q denote the probability that the random walk starting at the origin ever returns to the origin. We are interested whether $q = 1$ or $q < 1$. Let us suppose that $q < 1$. Then the probability of exactly k visits is given by

$$\mathbb{P}\{V = k\} = (1-q)\, q^{k-1}, \quad k = 1, 2, 3, \ldots.$$

To see this, we note that in order to have exactly k visits, we need to return exactly $k-1$ times. The probability of returning is q and the probability of not returning (sometimes called the escape probability) is $1-q$. This is a geometric distribution giving the number of trials until a success where a success here is "never returning". Using this we get

$$\mathbb{E}[V] = \sum_{k=1}^{\infty} k\,(1-q)\,q^{k-1} = \frac{1}{1-q}.$$

For us the key thing is that if $q < 1$, then $\mathbb{E}[V] < \infty$. Therefore, since we have $\mathbb{E}[V] = \infty$, we must have $q = 1$.

What happens if we allow the walker to walk in more than one dimension? Let $\mathbb{Z}^d$ denote the d-dimensional integer grid

$$\mathbb{Z}^d = \{(x_1, x_2, \ldots, x_d) : x_j \text{ integer}\}.$$

Each point in $\mathbb{Z}^d$ has $2d$ "nearest neighbors", that is, points distance one away. Simple random walk in $\mathbb{Z}^d$ is the process where a walker at each time moves to a nearest neighbor with probability $1/2d$ for each of the possible choices. As in the one-dimensional case, the lattice splits into "even" and "odd" points where in this case even means that $x_1 + \cdots + x_d$ is even. It can be shown that the probability that the walker starting at the origin is at the origin at time $2n$ is asymptotic to

$$\frac{c_d}{n^{d/2}}, \qquad c_d = \left(\frac{d}{\pi}\right)^{d/2} 2^{1-d}.$$

Let us not worry about the constant c_d but just focus on the exponent $d/2$. Let us given a simple heuristic why this should be correct. In $2n$ steps, a random walker is typically at a distance of order $\sqrt{n}$ from the origin. The number of points in the d-dimensional lattice that are distance about $\sqrt{n}$ from the origin is on the order of $(\sqrt{n})^d$. Hence the probability of being at a particular one should be like $1/(\sqrt{n})^d = n^{-d/2}$.

Using this we see that the expected number of visits to the origin

$$\mathbb{E}[V] = \sum_{j=0}^{\infty} \mathbb{P}\{S_{2j} = 0\}$$

is infinite for $d = 1, 2$ but is finite for larger d.

Theorem 1. *(Pólya) Simple random walk in one and two dimensions is recurrent, that is, returns to the origin infinitely often. Simple random walk in three or more dimensions is transient, that is, only returns to the origin finitely often and eventually goes to infinity.*

Exercise

Exercise 3.1. Suppose we change the probabilities in simple random walk so that

$$\mathbb{P}\{X_j = 1\} = p, \qquad \mathbb{P}\{X_j = -1\} = 1 - p,$$

where $1/2 < p < 1$. Let

$$q_n = \mathbb{P}\{S_{2n} = 0\}$$

where we start at the origin.

- Give an exact expression for q_n.
- Show that

$$\sum_{n=1}^{\infty} q_n < \infty$$

  and conclude that the random walk does not return to the origin infinitely often.

Exercise 3.2. Use the central limit theorem to find

$$\lim_{n\to\infty} \mathbb{P}\{S_n > 2\sqrt{n}\}.$$

Do this for both the symmetric simple random walk and the asymmetric random walk in Exercise 3.1.

Exercise 3.3. For the symmetric simple random walk, find the following for $m < n < u$.

$$\mathbb{E}[S_m\, S_n], \quad \rho(S_m, S_n), \quad \mathbb{E}[S_m\, S_n\, S_u].$$

Here ρ denotes the correlation coefficient.

Exercise 3.4. Let us call m an upswing time for (symmetric) simple random walk if $S_m = S_{m-4} + 4$, that is, if we have had four consecutive $+1$ values. Find the expected number of steps until we have an upswing time. (Hint: a very similar problem was discussed in the August review and you should feel free to consult those notes.)

4 (Time homogeneous) Markov chains

A discrete time stochastic process

$$X_0, X_1, X_2, X_3, \ldots$$

is called Markov if the only information about the past that is relevant for predicting the future is the current value. In other words, if $\mathcal{F}_n$ denotes the information available in $X_0, X_1, \ldots, X_n$, then the conditional distribution of the future

$$X_{n+1}, X_{n+2}, \ldots$$

given $\mathcal{F}_n$ is the same as the conditional distribution given X_n.

We will study (time homogeneous) Markov chains, which are such processes whose state space S (set of possible values for the chain) is a finite or countably infinite set. If S is finite we will call it a finite Markov chain. Markov chains are specified by their transition probabilities

$$p(x,y) = \mathbb{P}\{X_1 = y \mid X_0 = x\} = \mathbb{P}\{X_{n+1} = y \mid X_n = x\}.$$

The second equality is the time homogeneity which we will assume throughout. For a finite Markov chain whose state space has N elements, the transition probabilities can be written as an $N \times N$ transition matrix

$$\mathbf{P} = [p(x,y)]_{x,y \in S}.$$

We can also view the probabilities as a matrix with an infinite state space if we are happy talking about infinite matrices. The transition probabilities satisfy two conditions:

$$p(x,y) \geq 0, \quad x,y \in S,$$

and for each $x \in S$,

$$\sum_{y \in S} p(x,y) = 1.$$

This just means that starting in state x, we will definitely go to some state in the next step and hence the sum of the probabilities equals 1. For a finite Markov chain, this means that the transition matrix has all nonnegative entries and the sum of the elements of each row equals one. Such matrices are called stochastic matrices.

Example 4.1. The simple random walk in $\mathbb{Z}^d$ is a Markov chain whose state space is $\mathbb{Z}^d$ and has transition probabilities

$$p(x,y) = \frac{1}{2d}, \qquad \text{if } |x-y| = 1.$$

Example 4.2. If the state space has two elements, say $S = \{0,1\}$ then the transition matrix is determined by two numbers $p(0,1) = p$ and $p(1,0) = q$ and the transition matrix is

$$\mathbf{P} = \begin{array}{c c} & \begin{array}{cc} 0 & \quad 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} \end{array}.$$

The n-step transition probabilities for a Markov chain are defined by

$$p_n(x,y) = \mathbb{P}\{X_n = y \mid X_0 = x\} = \mathbb{P}\{X_{n+m} = y \mid X_m = x\}.$$

Using the law of total probability, we get a relation between the transition probabilities that goes under the fancy name, Chapman-Kolmogorov equations:

$$p_{m+n}(x,y) = \sum_{z \in S} p_m(x,z)\, p_n(z,y).$$

This basically just says that if you are going to be in state y after $m+n$ steps, you must be in some state z after m steps and we can add up all the possibilities. To be a little more formal we assume $X_0 = x$ and write

$$\begin{aligned}\mathbb{P}\{X_{m+n} = y\} &= \sum_{z \in S} \mathbb{P}\{X_m = z\}\, \mathbb{P}\{X_{m+n} = y \mid X_m = z\} \\ &= \sum_{z \in S} p_m(x,z)\, p_n(z,y)\end{aligned}$$

In the case of finite chains, the Chapman-Kolmogorov equations are the same as matrix multiplication.

Fact For a finite Markov chain, the n-step transition probabilities are the entries of the transition matrix raised to the n power, $\mathbf{P}^n$.

This is also true for $n = 0$ with $\mathbf{P}^0 = \mathbf{I}$ where $\mathbf{I}$ denotes the identity matrix. (This is immediate from the definition since $\mathbb{P}\{X_0 = y \mid X_0 = x\}$ is equal to 0 or 1 depending on whether $y \neq x$ or $y = x$.) This is also true for infinite state spaces if we understand the nth power of an infinite matrix to be defined by the Chapman-Kolmogorov equations. Taking powers of matrices is easy to do computationally for finite chains but it is not so obvious how to do it for infinite chains.

To complete the description of a Markov chain, we must give the initial probability distribution which we write as

$$\pi_0(x) = \mathbb{P}\{X_0 = x\}.$$

Then we can write

$$\pi_n(y) := \mathbb{P}\{X_n = y\} = \sum_{x \in S} \mathbb{P}\{X_0 = x\}\, \mathbb{P}\{X_n = y \mid X_0 = x\} = \sum_{x \in S} \pi_0(x)\, p_n(x,y).$$

For a finite Markov chain, if we view $\boldsymbol{\pi}_0, \boldsymbol{\pi}_n$ as vectors, then this becomes the matrix equation

$$\boldsymbol{\pi}_n = \boldsymbol{\pi}_0\, \mathbf{P}^n.$$

The order of multiplication is important; the vector $\boldsymbol{\pi}_0$ is on the left.

Example 4.3. We will consider three variants of simple random walk, each restricted to the set $\{0,1,2,3,4\}$ and then consider $\mathbf{P}^n$ for large n. When the walker is inside, it moves like simple random walk,

$$p(j, j+1) = p(j, j-1) = \frac{1}{2} \quad \text{if } j = 1,2,3.$$

The three cases will differ on how the walker moves when it reaches a boundary point.

- Absorbing boundary. When the walker reaches the boundary, it stops moving. This gives $p(0,0) = p(4,4) = 1$ and

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array}\!\!\begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}, \qquad \mathbf{P}^n \approx \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array}\!\!\begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ .75 & 0 & 0 & 0 & .25 \\ .50 & 0 & 0 & 0 & .50 \\ .25 & 0 & 0 & 0 & .75 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}.$$

- Reflecting boundary. When the walker reaches the boundary it immediately goes back to the interior. This gives $p(0,1) = p(4,3) = 1$,

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array}\!\!\begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \end{array},$$

$$\mathbf{P}^{2n} \approx \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array}\!\!\begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} .25 & 0 & .50 & 0 & .25 \\ 0 & .50 & 0 & .50 & 0 \\ .25 & 0 & .50 & 0 & .25 \\ 0 & .50 & 0 & .50 & 0 \\ .25 & 0 & .50 & 0 & .25 \end{pmatrix} \end{array}, \qquad \mathbf{P}^{2n+1} \approx \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 3 \\ 4 \end{array}\!\!\begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 0 & .50 & 0 & .50 & 0 \\ .25 & 0 & .50 & 0 & .25 \\ 0 & .50 & 0 & .50 & 0 \\ .25 & 0 & .50 & 0 & .25 \\ 0 & .50 & 0 & .50 & 0 \\ .25 & 0 & .50 & 0 & .25 \end{pmatrix} \end{array}.$$

  In this case, the large powers look different depending on whether it is an even or an odd power.

- Partially reflecting boundary. When the walker reaches the boundary it tries to take a usual random walk step but if that would take the walker outside, the step is rejected and the walker stays in place. This gives $p(0,0) = p(0,1) = p(4,4) = (4,3) = \frac{1}{2}$,

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array}\!\!\begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \end{array}, \qquad \mathbf{P}^n = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array}\!\!\begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} .2 & .2 & .2 & .2 & .2 \\ .2 & .2 & .2 & .2 & .2 \\ .2 & .2 & .2 & .2 & .2 \\ .2 & .2 & .2 & .2 & .2 \\ .2 & .2 & .2 & .2 & .2 \end{pmatrix} \end{array}.$$

An invariant probability for a Markov chain is a probability distribution $\boldsymbol{\pi}$ such that if we start in this distribution we stay in this distribution forever. This can be written as

$$\boldsymbol{\pi}\,\mathbf{P} = \boldsymbol{\pi}, \qquad \pi(y) = \sum_x \pi(x)\,p(x,y).$$

In other words, it is a left eigenvector for the matrix $\mathbf{P}$ that is also a probability distribution (all nonnegative entries that add to one).

A Markov chain is called irreducible if one can get from any site to any other site with positive probability in some number of steps. More precisely, it is irreducible if for every $x, y \in S$, there exist nonnegative integers m, n with $p_m(x,y) > 0, p_n(y,x) > 0$. An irreducible chain is called aperiodic ("not periodic") if for every $x \in S$, there exists an integer N_x such that $p_n(x,x) > 0$ for all $n \geq N_x$. For the random walk on $\{0, 1, \ldots, 4\}$ with different boundary conditions we have: the absorbing walk is not irredicible; the reflecting walk is irreducible but not aperiodic; and finally the partially reflecting walk is irreducible and aperiodic.

Theorem 2. *If $\mathbf{P}$ is the transition matrix for a finite irreducible Markov chain, then there exists a unique invariant probability $\boldsymbol{\pi}$,*

$$\boldsymbol{\pi}\,\mathbf{P} = \boldsymbol{\pi}.$$

- *If $\mathbf{P}$ is aperiodic, then for every $x, y \in S$,*

$$\lim_{n\to\infty} p_n(x,y) = \boldsymbol{\pi}(y).$$

- *If $\mathbf{P}$ is not aperiodic, then the last equation does not hold but we still have*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} p_j(x,y) = \boldsymbol{\pi}(y).$$

In other words, the invariant probability $\pi(y)$ for an irreducible Markov chain gives the long range fraction of time spent in the state y. There is another way of viewing the invariant probability $\boldsymbol{\pi}$. Suppose we start the Markov chain at x and consider the first return time for the chain

$$T = \min\{j \geq 1 : X_j = x\}.$$

More generally we can let T_k be the time between the $(k-1)$st and kth return,

$$T_k = \min\{j \geq 1 : X_{T_{k-1}+j} = x\}.$$

Then $T_1, T_2, \ldots$ are independent, identically distributed random variables and by the law of large numbers we know that

$$\frac{T_1 + \cdots + T_k}{k} \sim \mathbb{E}[T].$$

Note that the number of times in $\{1, 2, \ldots, T_k\}$ that the chain has spent in state x has been k. But we know that the fraction of time that the chain spends in x is $\pi(x)$. This gives the rule

$$\pi(x) = \frac{1}{\mathbb{E}[T]}.$$

Example 4.4. Let us consider the two state Markov chain with transition matrix

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} 0 & \quad 1 \end{array} \\ \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} \end{array}.$$

where $0 < p, q < 1$. This is irreducible and aperiodic. The invariant probability satisfies

$$\boldsymbol{\pi}\,\mathbf{P} = \boldsymbol{\pi}.$$

If we write $\boldsymbol{\pi} = (\pi(0)\ \pi(1))$, then the equations are

$$(1-p)\,\pi(0) + q\,\pi(1) = \pi(0), \qquad p\,\pi(0) + (1-q)\,\pi(1) = \pi(1).$$

These two equations are redundant and only give us the relation $p\,\pi(0) = q\,\pi(1)$. The extra condition $\pi(0) + \pi(1) = 1$ gives the unique solution

$$\boldsymbol{\pi} = \left(\frac{q}{p+q}, \frac{p}{p+q}\right).$$

Starting at state 0, the expected time to return to 0 is $1/\pi(0) = (p+q)/q$. Similarly, the expected time starting at 1 to return to 1 is $1/\pi(1) = (p+q)/p$.

Let us consider the case $p = q$ so that $\pi(0) = \pi(1) = 1/2$ and the expected return time is 2. While the expected return time does not depend on the value of p, the distribution certainly does. If p is very close to 1, then most of the time the chain is switching states and most of the time it takes exactly two steps to return. However, if p is close to 0, then the chain usually stays in the same state and the return time is most often equal to 1. However, in the rare case that the chain moves on the first step it will take a long time for it to change again and return to the initial state.

Suppose A is a subset of the Markov chain and we are interested in the amount of time it takes in order for the chain to leave A,

$$T = \min\{j : X_j \notin A\}.$$

This is sometimes called the passage time and the expectation $\mathbb{E}[T]$ is the mean passage time. Expectations are easier to compute than the entire distribution of the random variable. Suppose $x \in A$ and $X_0 = x$. Using indicator functions we can write

$$T = \sum_{n=0}^{\infty} 1\{X_n \in A, n < T\} = \sum_{n=0}^{\infty} \sum_{y \in A} 1\{X_n = y, n < T\} = \sum_{y \in A} \sum_{n=0}^{\infty} 1\{X_n = y, n < T\},$$

and using linearity of expectation we have

$$\mathbb{E}[T] = \sum_{y \in A} G_A(x, y)$$

where $G_A(x, y)$ is the Green's function (or Green's matrix)

$$G_A(x, y) = \sum_{n=0}^{\infty} \mathbb{P}\{X_n = y, n < T\}.$$

This represents the expected number of visits to y before leaving A assuming we start at x. The matrix

$$\mathbf{M} = [G_A(x, y)]_{x,y \in A}$$

is called the Green's matrix.

Let $\mathbf{Q}$ be the matrix $\mathbf{P}$ restricted to the rows and columns corresponding to sites in A. Then

- $\mathbb{P}\{X_n = y, n < T \mid X_0 = x\}$ is the (x, y) entry of the matrix $\mathbf{Q}^n$. In particular,

$$\mathbf{M} = \mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \cdots = [\mathbf{I} - \mathbf{Q}]^{-1}.$$

The second equality is similar to the identity for the geometric sum

$$1 + t + t^2 + \cdots = \frac{1}{1-t}, \quad |t| < 1.$$

It can be justified in the same way by multiplying to see that

$$[\mathbf{I} - \mathbf{Q}] \left[\mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \cdots\right] = \mathbf{I}.$$

We summarize.

Fact The (x, y) entry of the Green's matrix $\mathbf{M}$ gives the expected number of visits to y before leaving A assuming $X_0 = x$. The mean passage time starting at x to leave A is given by the sum of the entries in the x-row of $\mathbf{M}$

4.1 An example

Consider a Markov chain with four states $\{1, 2, 3, 4\}$ and transition probabilities

$$p(1, 1) = p(1, 2) = \frac{1}{2},$$

$$p(2, 1) = p(2, 3) = \frac{1}{2},$$

$$p(3, 2) = p(3, 4) = \frac{1}{2},$$

$$p(4, 1) = p(4, 2) = p(4, 3) = \frac{1}{3}.$$

The transition matrix $\mathbf{P}$ is given by

$$\mathbf{P} = \begin{array}{c c} & \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \end{pmatrix} \end{array}.$$

Note that this is irreducible.

To find the invariant probability we solve the equation

$$\boldsymbol{\pi}\,\mathbf{P} = \boldsymbol{\pi}.$$

We could use a computer but we will do this by hand. This gives us four equations in four unknowns:

$$\frac{1}{2}\,\pi(1) + \frac{1}{2}\,\pi(2) + \frac{1}{3}\,\pi(4) = \pi(1),$$
$$\frac{1}{2}\,\pi(1) + \frac{1}{2}\,\pi(3) + \frac{1}{3}\,\pi(4) = \pi(2),$$
$$\frac{1}{2}\,\pi(2) + \frac{1}{3}\,\pi(4) = \pi(3),$$
$$\frac{1}{2}\,\pi(3) = \pi(4).$$

Plugging the last equation into the third gives

$$\pi(3) = \frac{3}{5}\,\pi(2), \quad \pi(4) = \frac{1}{2}\,\pi(3) = \frac{3}{10}\,\pi(2)$$

Plugging these into the first two equation gives

$$\pi(2) = \frac{5}{6}\,\pi(1).$$

which gives the invariant probability

$$\begin{bmatrix} \pi(1) & \frac{5}{6}\,\pi(1) & \frac{1}{2}\,\pi(1) & \frac{1}{4}\,\pi(1) \end{bmatrix}.$$

We now impose the condition that the sum of the components equals one. Since

$$1 + \frac{5}{6} + \frac{1}{2} + \frac{1}{4} = \frac{31}{12}$$

we get

$$\boldsymbol{\pi} = \begin{bmatrix} \frac{12}{31} & \frac{10}{31} & \frac{6}{31} & \frac{3}{31} \end{bmatrix}.$$

On the average, the expected fraction of time spent in state 4 is $\pi(4) = 3/31$. The expected time starting at state 4 to return to state 4 is

$$\frac{1}{\pi(4)} = \frac{31}{3}.$$

We will now consider the mean passage time until reaching the state 2. In this case $A = \{1, 3, 4\}$ and

$$\mathbf{Q} = \begin{array}{c} \\ 1 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 3 & 4 \end{array} \\ \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{3} & 0 \end{pmatrix} \end{array}, \qquad \mathbf{I} - \mathbf{Q} = \begin{array}{c} \\ 1 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 3 & 4 \end{array} \\ \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 1 & -\frac{1}{2} \\ -\frac{1}{3} & -\frac{1}{3} & 1 \end{pmatrix} \end{array},$$

and the Green's matrix is

$$\mathbf{M} = [\mathbf{I} - \mathbf{Q}]^{-1} = \begin{array}{c} \\ 1 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 3 & 4 \end{array} \\ \begin{pmatrix} 2 & 0 & 0 \\ \frac{2}{5} & \frac{6}{5} & \frac{3}{5} \\ \frac{4}{5} & \frac{2}{5} & \frac{6}{5} \end{pmatrix} \end{array}$$

You may note that in this chain, if one starts in state 1, one cannot get to states 3 or 4 without going through state 2. This explains the two zeroes in the first row. The mean passage time to 2 starting in state 4 is the sum of the row associated to state 4, that is,

$$\frac{4}{5} + \frac{2}{5} + \frac{6}{5} = \frac{12}{5}.$$

4.2 Infinite state spaces

We have already seen an interesting Markov chain with an infinite state space — simple random walk in $\mathbb{Z}^d$. There we saw different behavior depending on the dimension d.

- If $d = 1, 2$, the random walk is recurrent, that is, returns to all points infinitely often.
- If $d \geq 3$, the random walk is transient, that is, visits points only finitely often and wanders off to infinity.

We can use the same term for irreducible Markov chains — they are recurrent if they visit points infinitely often and they are transient if they visit points only finitely often.

We can divide recurrent Markov chains into two cases. Let us consider simple random walk in one dimension. We have seen that

$$\mathbb{P}\{S_{2n} = 0\} \sim \frac{1}{\sqrt{\pi n}}.$$

In particular,

$$\lim_{n \to \infty} \mathbb{P}\{S_n = 0\} = 0,$$

and in the long run, the fraction of time spent at the origin goes to zero. We call recurrent chains with this property null recurrent.

There are also irreducible chains on infinite state spaces that spend positive fractions of times at sites. An irreducible Markov chain is called positive recurrent if there exists an invariant probability. For an infinite state space, an invariant probability is a probability function on S with the property that $\pi\,\mathbf{P} = \boldsymbol{\pi}$, that is, for every $x \in S$,

$$\pi(x) = \sum_{y \in S} \pi(y)\, p(y,x). \tag{5}$$

Let us give an example: random walk with partially reflecting boundary at the origin with probability q of moving to the right where $0 < q < 1/2$. To be precise,

$$p(n, n-1) = 1-q, \qquad p(n, n+1) = q, \qquad n = 1, 2, \ldots,$$
$$p(0,0) = 1-q, \quad p(0,1) = q.$$

Suppose we set

$$\pi(y) = \left(\frac{q}{1-q}\right)^y, \quad y = 0, 1, 2, \ldots$$

Note that

$$q\left(\frac{q}{1-q}\right)^{y-1} + (1-q)\left(\frac{q}{1-q}\right)^{y+1} = \left(\frac{q}{1-q}\right)^y$$

From this we see that this satisfies (5). We want to make it into a probability distribution, so we choose

$$\pi(y) = c\left(\frac{q}{1-q}\right)^y, \quad y = 0, 1, 2, \ldots$$

where

$$c = \left[\sum_{y=0}^{\infty}\left(\frac{q}{1-q}\right)^y\right]^{-1} = \left[\frac{1}{1-\frac{q}{1-q}}\right]^{-1} = \frac{1-2q}{1-q}.$$

Here we have used $0 < q < 1/2$ to guarantee that the infinite sum converges. For this chain, the long-range fraction of time that the chain is in state y is

$$\pi(y) = \frac{1-2q}{1-q}\left(\frac{q}{1-q}\right)^y.$$

There is a way to determine whether or not an irreducible Markov chain is positive recurrent. Basically, one tries to find an invariant probability, that is, a probability function that satisfies (5). If one finds one, then it is positive recurrent and that is the invariant probability. If one shows that there is no such probability function, then it is not positive recurrent.

We can also relate the invariant probability to the return time for the chain. Suppose that $X_0 = x$ and let T be the return time

$$T = \min\{j \geq 1 : X_j = x\}$$

where we set $T = \infty$ if no such time exists. Then there is a trichotomy:

- If the chain is positive recurrent, then

$$\mathbb{E}[T] = \frac{1}{\pi(x)}.$$

- If the chain is null recurrent, then with probability one, $T < \infty$ but

$$\mathbb{E}[T] = \infty.$$

- If the chain is transient, then $\mathbb{P}\{T = \infty\} > 0$.

In particular, in the case of simple random walk in one or two dimensions, the chain is null recurrent. Therefore, even though we know that the walker will return to the origin, the expected amount of time until return is infinite!

An overused joke among mathematicians, especially probabilists, who are trying to find a particular restaurant is "don't worry, random walk in two dimensions is recurrent so we will eventually get there". After this someone adds "but the expected time is infinite!"

Exercises

Exercise 4.1. Suppose X_n is a Markov chain with state space $S = \{A, B, C\}$ and transition matrix

$$\mathbf{P} = \begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{pmatrix} 1/4 & 1/2 & 1/4 \\ 1/3 & 1/3 & 1/3 \\ 1/2 & 1/4 & 1/4 \end{pmatrix} \end{array}.$$

1. Find the long range fraction of time that the chain spends in the three states A, B, C.

2. Suppose an investor gains \$1 when the chain is in state B, \$1.50 when the chain is in state C, and loses \$.50 when the chain is in state A. Then the long range earnings after n steps is proportional to $c\,n$ for some constant c. What is c?

3. Suppose that $X_0 = A$. Find

$$\mathbb{P}\{X_4 = B \mid X_3 = C\}, \qquad \mathbb{P}\{X_3 = C \mid X_4 = B\}.$$

Exercise 4.2. Consider the following Markov chain with state space $\{0, 1, 2, 3\}$.

- When the chain is in state j, then $3 - j$ fair coins are flipped. If there are exactly k heads on these flips, we move to state k. In particular, if we are in state 3 we always move to state 0.

1. Write down the transition matrix for this chain.
2. Give the invariant probability for the chain.
3. Suppose we start at state 0. What is the probability that after three steps we are in state 3?
4. Suppose we start at state 0. What is the expected number of steps until we return to state 0?

Exercise 4.3. Suppose X_n is a finite irreducible Markov chain whose transition matrix is doubly stochastic, that is, both the rows and the columns add up to one. Show that the uniform distribution is the invariant probability distribution for the chain.

Exercise 4.4. Consider a finite Markov chain on $\{0, 1, 2, 3, 4\}$ with the following rule. If we are at $j \geq 1$ we always move one step down. But if are at the origin we choose any of the other four states at random. In other words,

$$p(j, j-1) = 1, \quad j = 1, 2, 3, 4,$$

$$p(0,1) = p(0,2) = p(0,3) = p(0,4) = \frac{1}{4}.$$

1. Explain why this an irreducible chain.
2. Give the invariant probability.
3. Suppose the chain starts in state 4. What is the expected amount of time before it returns to state 4?
4. Suppose the chain starts at 0. What is the probability that it will reach state 4 before it reaches state 3?
5. Suppose the chain starts at 0. What is the probability that it will reach state 4 before it reaches state 2?

Exercise 4.5. Take a standard 52 deck of cards. We will do the following simple "shuffle" of the cards. Choose one of the 51 cards that are not the top card of the deck (uniformly) and move that card to the top of the deck leaving all the other cards in the same order. This is a Markov chain whose state space is the set of 52! orderings (shuffles, permutations) of the deck.

1. Is this an irreducible Markov chain?

2. Is the transition matrix doubly stochastic? (See Exercise 4.3.)

3. Suppose we start with a particular ordering of the cards. Assume I do one "shuffle" every second. What is the expected amount of time until the deck is back to the original order?

Exercise 4.6. Suppose we have a Markov chain with state space $\{0,1,2,3\}$ and transition probabilities

$$\begin{array}{c c} & \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} & \begin{pmatrix} 1/3 & 2/3 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 3/5 & 2/5 \end{pmatrix} \end{array}.$$

1. Suppose that $X_0 = 2$. Find the probability that $X_4 = 3$.

2. Suppose that $X_0 = 2$. Find the probability that the following all happen: $X_4 = 3, X_5 = 2, X_6 = 1, X_7 = 1$.

3. Find the invariant probability.

4. Suppose that $X_0 = 2$. What is the expected amount of time until the chain returns to state 2?

5. Suppose that $X_0 = 2$. What is the expected amount of time until the chain reaches state 3?

Exercise 4.7. Consider the infinite Markov chain with state space $\{0,1,2,\ldots\}$ that moves according to the following rules.

- If the state is currently in state $j > 0$ it moves to state $j-1$.
- If the state is current in state 0, then a Poisson random variable with mean 2 is sampled and one moves to that state. In other words,

$$p(0,k) = e^{-2}\,\frac{2^k}{k!}.$$

Suppose that $X_0 = 0$.

1. What is the expected number of steps until we return to 0 for the first time?

2. Show that this is a positive recurrent chain and give the invariant probability π.

5 Martingales

5.1 Introduction

A martingale is a model of a fair game. Suppose $X_1, X_2, \ldots$ is a sequence of random variables to which we associate the filtration $\{\mathcal{F}_n\}$ where $\mathcal{F}_n$ is the information contained in $X_1, \ldots, X_n$.

Definition A sequence of random variables $M_0, M_1, \ldots$ is called a martingale with respect to the filtration $\{\mathcal{F}_n\}$ if:

- For each n, M_n is an $\mathcal{F}_n$-measurable random variable with $\mathbb{E}[|M_n|] < \infty$.
- If $m < n$, then
$$E[M_n \mid \mathcal{F}_m] = M_m. \tag{6}$$

  We can also write (6) as
$$E[M_n - M_m \mid \mathcal{F}_m] = 0.$$

If we think of M_n as the winnings of a game, then this implies that no matter what has happened up to time m, the expected winnings in the next $n-m$ games is 0. Sometimes one just says "$M_0, M_1, \ldots$ is a martingale" without reference to the filtration. In this case, the assumed filtration is $\mathcal{F}_n$, the information in $M_0, \ldots, M_n$. In order to establish (6) it suffices to show for all n,
$$E[M_{n+1} \mid \mathcal{F}_n] = M_n. \tag{7}$$
In order to see that this suffices, we can use the tower property (2) for conditional expectation to see that

$$E[M_{n+2} \mid \mathcal{F}_n] = E\left[\, E[M_{n+2} \mid \mathcal{F}_{n+1}] \;\mid \mathcal{F}_n\right] \;= E[M_{n+1} \mid \mathcal{F}_n] = M_n,$$

and so forth. Also note that if M_n is a martingale, then

$$\mathbb{E}[M_n] = \mathbb{E}\left[E[M_n \mid \mathcal{F}_0]\right] = \mathbb{E}[M_0].$$

Example 5.1. Suppose $X_1, X_2, \ldots$ are independent random variables with $\mathbb{E}[X_j] = 0$ for each j. Let $S_0 = 0$ and $S_n = X_1 + \cdots + X_n$. In Example 2.2, we showed that if $m < n$, then $E[S_n \mid \mathcal{F}_m] = S_m$. Hence, S_n is a martingale with respect to $\mathcal{F}_n$, the information in $X_1, \ldots, X_n$.

Example 5.2. Suppose $X_n, S_n, \mathcal{F}_n$ are as in Example 5.1 and also assume $\mathrm{Var}[X_j] = \mathbb{E}[X_j^2] = \sigma_j^2 < \infty$. Let
$$A_n = \sigma_1^2 + \cdots + \sigma_n^2,$$
$$M_n = S_n^2 - A_n,$$

where $M_0 = 0$. Then M_n is a martingale with respect to $\mathcal{F}_n$. To see this, we compute as in Example 2.2,

$$\begin{aligned}
E[S_{n+1}^2 \mid \mathcal{F}_n] &= E[(S_n + X_{n+1})^2 \mid \mathcal{F}_n] \\
&= E[S_n^2 \mid \mathcal{F}_n] + 2E[S_n X_{n+1} \mid \mathcal{F}_n] + E[X_{n+1}^2 \mid \mathcal{F}_n] \\
&= S_n^2 + 2S_n\, E[X_{n+1} \mid \mathcal{F}_n] + \mathbb{E}[X_{n+1}^2] \\
&= S_n^2 + 2S_n\, \mathbb{E}[X_{n+1}] + \mathbb{E}[X_{n+1}^2] \\
&= S_n^2 + \sigma_{n+1}^2.
\end{aligned}$$

Therefore,

$$\begin{aligned}
E[M_{n+1} \mid \mathcal{F}_n] &= E[S_{n+1}^2 - A_{n+1} \mid \mathcal{F}_n] \\
&= S_n^2 + \sigma_{n+1}^2 - (A_n + \sigma_{n+1}^2) = M_n.
\end{aligned}$$

There are various ways to view a martingale. One can consider M_n as the price of an asset (although we allow negative values of M_n) or as the winnings in a game. We can also consider

$$\Delta M_n = M_n - M_{n-1}$$

as either the change in the asset price or as the amount won in the game at time n. Negative values indicate drops in price or money lost in the game. The basic idea of stochastic integration is to allow one to change one's portfolio (in the asset viewpoint) or change one's bet (in the game viewpoint). However, we are not allowed to see the outcome before betting. We make this precise in the next example.

Example 5.3. Discrete stochastic integral. Suppose that $M_0, M_1, \ldots$ is a martingale with respect to the filtration $\mathcal{F}_n$. For $n \geq 1$, let $\Delta M_n = M_n - M_{n-1}$. Let B_j denote the "bet" on the jth game. We allow negative values of B_j which indicate betting that the price will go down or the game will be lost. Let W_n denote the winnings in this strategy: $W_0 = 0$ and for $n \geq 1$,

$$W_n = \sum_{j=1}^{n} B_j\, [M_j - M_{j-1}] = \sum_{j=1}^{n} B_j\, \Delta M_j.$$

Let us assume that for each n there is a number $K_n < \infty$ such that $|B_n| \leq K_n$. We also assume that we cannot see the result of nth game before betting. This last assumption can be expressed mathematically by saying that B_n is $\mathcal{F}_{n-1}$-measurable. In other words, we can adjust our bet based on how well we have been doing. We claim that under these assumptions, W_n is a martingale with respect to $\mathcal{F}_n$. It is clear that W_n is measurable with respect to $\mathcal{F}_n$, and integrability follows from the estimate

$$\begin{aligned}
\mathbb{E}[|W_n|] &\leq \sum_{j=1}^{n} \mathbb{E}[|B_j||M_j - M_{j-1}|] \\
&\leq \sum_{j=1}^{n} K_j\, (\mathbb{E}[|M_j|] + \mathbb{E}[|M_{j-1}|]) < \infty.
\end{aligned}$$

Also,

$$\begin{aligned} E[W_{n+1} \mid \mathcal{F}_n] &= E[W_n + B_{n+1}(M_{n+1} - M_n) \mid \mathcal{F}_n] \\ &= E[W_n \mid \mathcal{F}_n] + E[B_{n+1}(M_{n+1} - M_n) \mid \mathcal{F}_n]. \end{aligned}$$

Since W_n is $\mathcal{F}_n$-measurable, $E[W_n \mid \mathcal{F}_n] = W_n$. Also, since B_{n+1} is $\mathcal{F}_n$-measurable and M is a martingale,

$$E[B_{n+1}(M_{n+1} - M_n) \mid \mathcal{F}_n] = B_{n+1}\, E[M_{n+1} - M_n \mid \mathcal{F}_n] = 0.$$

Therefore,

$$E[W_{n+1} \mid \mathcal{F}_n] = W_n.$$

Example 5.3 demonstrates an important aspect of martingales. One cannot change a discrete-time martingale to a game in one's favor with a betting strategy in a finite amount of time. However, the next example shows that if we are allowed an infinite amount of time we can beat a fair game.

Example 5.4. Martingale betting strategy. Let $X_1, X_2, \ldots$ be independent random variables with

$$\mathbb{P}\{X_j = 1\} = \mathbb{P}\{X_j = -1\} = \frac{1}{2}. \tag{8}$$

We will refer to such random variables as "coin-tossing" random variables where 1 corresponds to heads and -1 corresponds to tails. Let $M_0 = 0, M_n = X_1 + \cdots + X_n$. We have seen that M_n is a martingale. We will consider the following betting strategy. We start by betting \$1. If we win, we quit; otherwise, we bet \$2 on the next game. If we win the second game, we quit; otherwise we double our bet to \$4 and play. Each time we lose, we double our bet. At the time that we win, we will be ahead \$1. With probability one, we will eventually win the game, so this strategy is a way to beat a fair game. The winnings in this game can be written as

$$W_n = \sum_{j=1}^{n} B_j\, \Delta M_j = \sum_{j=1}^{n} B_j\, X_j,$$

where the bet $B_1 = 1$ and for $j > 1$,

$$B_j = 2^{j-1} \text{ if } X_1 = X_2 = \cdots = X_{j-1} = -1,$$

and otherwise $B_j = 0$. This is an example of a discrete stochastic integral as in the previous example, and hence, we know that W_n must be a martingale. In particular, for each n, $\mathbb{E}[W_n] = 0$. We can check this directly by noting that $W_n = 1$ unless $X_1 = X_2 = \cdots = X_n = -1$ in which case

$$W_n = -1 - 2^1 - 2^2 - \cdots - 2^{n-1} = -[2^n - 1].$$

This last event happens with probability $(1/2)^n$, and hence

$$\mathbb{E}[W_n] = 1 \cdot [1 - 2^{-n}] - [2^n - 1] \cdot 2^{-n} = 0.$$

However, we will eventually win which means that with probability one

$$W_\infty = \lim_{n\to\infty} W_n = 1,$$

and

$$1 = \mathbb{E}[W_\infty] > \mathbb{E}[W_0] = 0.$$

We have beaten the game (but it takes an infinite amount of time to guarantee it).

If the condition (6) is replaced with

$$E[M_n \mid \mathcal{F}_m] \geq M_m,$$

then the process is called a submartingale. If it is replaced with

$$E[M_n \mid \mathcal{F}_m] \leq M_m,$$

then it is called a supermartingale. In other words, games that are always in one's favor are submartingales and games that are always against one are supermartingales. (In most games in Las Vegas, one's winnings give a supermartingale.) Under this definition, a martingale is both a submartingale and a supermartingale. The terminology may seem backwards at first: submartingales get bigger and supermartingales get smaller. The terminology was set to be consistent with the related notion of subharmonic and superharmonic functions. Martingales are related to harmonic functions.

5.2 Optional sampling theorem

Suppose $M_0, M_1, M_2, \ldots$ is a martingale with respect to the filtration $\{\mathcal{F}_n\}$. In the last section we discussed the discrete stochastic integral. Here we will consider a particular case of a betting strategy where one bets 1 up to some time and then one bets 0 afterwards. Let T be the "stopping time" for the strategy. Then the winnings at time n is

$$M_0 + \sum_{j=1}^{n} B_j\,[M_j - M_{j-1}],$$

where $B_j = 1$ if $j \leq T$ and $B_j = 0$ if $j > T$. We can write this as

$$M_{n\wedge T},$$

where $n \wedge T$ is shorthand for $\min\{n, T\}$. The time T is random, but it must satisfy the following condition to be allowable.

Definition A nonnegative integer-valued random variable T is a stopping time with respect to the filtration $\{\mathcal{F}_n\}$ if for each n the event $\{T = n\}$ is $\mathcal{F}_n$-measurable.

The following theorem is a special case of the discrete stochastic integral. It restates the fact that one cannot beat a martingale in finite time. It is one version of the optional sampling theorem; it is also called the optional stopping theorem.

Theorem 3 (Optional Sampling Theorem I)**.** *Suppose T is a stopping time and M_n is a martingale with respect to $\{\mathcal{F}_n\}$. If $Y_n = M_{n\wedge T}$, then Y_n is a martingale. In particular, for each n,*

$$\mathbb{E}\left[M_{n\wedge T}\right] = \mathbb{E}\left[M_0\right].$$

If T is bounded, that is, if there exists $k < \infty$ such that $\mathbb{P}\{T \leq k\} = 1$, then

$$\mathbb{E}\left[M_T\right] = \mathbb{E}\left[M_0\right]. \tag{9}$$

The final conclusion (9) of the theorem holds since $\mathbb{E}[M_{n\wedge T}] = \mathbb{E}[M_T]$ for $n \geq k$. What if the stopping time T is not bounded but $\mathbb{P}\{T < \infty\} = 1$? Then, we cannot conclude (9) without further assumptions. To see this we need only consider the martingale betting strategy of the previous section. If we define

$$T = \min\{n : X_n = 1\} = \min\{n : W_n = 1\},$$

then with probability one $T < \infty$ and $W_T = 1$. Hence,

$$1 = \mathbb{E}\left[W_T\right] > \mathbb{E}\left[W_0\right] = 0.$$

Often one does want to conclude (9) for unbounded stopping times, so it is useful to give conditions under which it holds. Let us try to derive the equality and see what conditions we need to impose. First, we will assume that we stop, $\mathbb{P}\{T < \infty\} = 1$, so that M_T makes sense. For every $n < \infty$, we know that

$$\mathbb{E}[M_0] = \mathbb{E}[M_{n\wedge T}] = \mathbb{E}[M_T] + \mathbb{E}[M_{n\wedge T} - M_T].$$

If we can show that

$$\lim_{n\to\infty} \mathbb{E}\left[|M_{n\wedge T} - M_T|\right] = 0,$$

then we have (9). The random variable $M_{n\wedge T} - M_T$ is zero if $n \wedge T = T$, and

$$M_{n\wedge T} - M_T = 1\{T > n\}\left[M_n - M_T\right].$$

If $\mathbb{E}[|M_T|] < \infty$, then one can show that

$$\lim_{n\to\infty} \mathbb{E}\left[|M_T|\, 1\{T > n\}\right] = 0.$$

In the martingale betting strategy example, this term did not cause a problem since $W_T = 1$ and hence $\mathbb{E}[|W_T|] < \infty$.

If $\mathbb{P}\{T < \infty\} = 1$, then the random variables $X_n = |M_T| \, 1\{T > n\}$ converge to zero with probability one. If $\mathbb{E}[|M_T|] < \infty$, then we can use the dominated convergence theorem to conclude that

$$\lim_{n\to\infty} \mathbb{E}[X_n] = 0.$$

Finally, in order to conclude (9) we will make the hypothesis that the other term acts nicely.

Theorem 4 (Optional Sampling Theorem II). *Suppose T is a stopping time and M_n is a martingale with respect to $\{\mathcal{F}_n\}$. Suppose that $\mathbb{P}\{T < \infty\} = 1, \mathbb{E}[|M_T|] < \infty$, and for each n,*

$$\lim_{n\to\infty} \mathbb{E}[|M_n| \, 1\{T > n\}] = 0. \tag{10}$$

Then,

$$\mathbb{E}[M_T] = \mathbb{E}[M_0].$$

Let us check that the martingale betting strategy does *not* satisfy the conditions of the theorem (it better not since it does not satisfy the conclusion!) In fact, it does not satisfy (10). For this strategy, if $T > n$, then we have lost n times and $W_n = 1 - 2^n$. Also, $\mathbb{P}\{T > n\} = 2^{-n}$. Therefore,

$$\lim_{n\to\infty} \mathbb{E}[|W_n| \, 1\{T > n\}] = \lim_{n\to\infty} (2^n - 1) \, 2^{-n} = 1 \neq 0.$$

Checking condition (10) can be difficult in general. We will give one criterion which is useful.

Theorem 5 (Optional Sampling Theorem III). *Suppose T is a stopping time and M_n is a martingale with respect to $\{\mathcal{F}_n\}$. Suppose that $\mathbb{P}\{T < \infty\} = 1, \mathbb{E}[|M_T|] < \infty$, and that there exists $C < \infty$ such that for each n,*

$$\mathbb{E}\left[|M_{n\wedge T}|^2\right] \leq C. \tag{11}$$

Then,

$$\mathbb{E}[M_T] = \mathbb{E}[M_0].$$

To prove this theorem, first note that with probability one,

$$|M_T|^2 = \lim_{n\to\infty} |M_{T\wedge n}|^2 \, 1\{T \leq n\},$$

and hence by the Hölder inequality and the monotone convergence theorem,

$$\mathbb{E}[|M_T|]^2 \leq \mathbb{E}\left[|M_T|^2\right] = \lim_{n\to\infty} \mathbb{E}\left[|M_{T\wedge n}|^2 | \, 1\{T \leq n\}\right] \leq C.$$

We need to show that (11) implies (10). If $b > 0$, then for every n,

$$\mathbb{E}\left[|M_n| \, 1\{|M_n| \geq b, T > n\}\right] \leq \frac{\mathbb{E}[|M_{n\wedge T}|^2]}{b} \leq \frac{C}{b}.$$

Therefore,

$$\begin{aligned}\mathbb{E}\left[|M_n| \, 1\{T > n\}\right] &= \mathbb{E}\left[|M_n| \, 1\{T > n, |M_n| \geq b\}\right] \\ &\qquad + \mathbb{E}\left[|M_n| \, 1\{T > n, |M_n| < b\}\right] \\ &\leq \frac{C}{b} + b \, \mathbb{P}\{T > n\}.\end{aligned}$$

Hence,

$$\limsup_{n\to\infty} \mathbb{E}\left[|M_n| \, 1\{T > n\}\right] \leq \frac{C}{b} + b \lim_{n\to\infty} \mathbb{P}\{T > n\} = \frac{C}{b}.$$

Since this holds for every $b > 0$ we get (10).

Example 5.5. Gambler's ruin for random walk. Here we give another derivation of a result from Section 3. Let $X_1, X_2, \ldots$ be independent, coin-tosses as in (8) and let

$$S_n = 1 + X_1 + \cdots + X_n.$$

S_n is called *simple (symmetric) random walk* starting at 1. We have shown that S_n is a martingale. Let $K > 1$ be a positive integer and let T denote the first time n such that $S_n = 0$ or $S_n = K$. Then $M_n := S_{n\wedge T}$ is a martingale. Also $0 \leq M_n \leq K$ for all n, so (11) is satisfied. We can apply the optional sampling theorem to deduce that

$$1 = M_0 = \mathbb{E}[M_T] = 0 \cdot \mathbb{P}\{M_T = 0\} + K \cdot \mathbb{P}\{M_T = K\}.$$

By solving, we get

$$\mathbb{P}\{M_T = K\} = \frac{1}{K}.$$

This relation is sometimes called the *gambler's ruin estimate* for the random walk. Note that

$$\lim_{K\to\infty} \mathbb{P}\{M_T = K\} = 0.$$

If we consider 1 to be the starting stake of a gambler and K to be the amount held by a casino, this shows that with a fair game, the gambler will almost surely lose. If $\tau = \min\{n : S_n = 0\}$, then the last equality implies that $\mathbb{P}\{\tau < \infty\} = 1$. The property that the walk always returns to the origin is called *recurrence*.

Example 5.6. Let $S_n = X_1 + \cdots + X_n$ be simple random walk starting at 0. We have seen that

$$M_n = S_n^2 - n$$

is a martingale. Let J, K be positive integers and let

$$T = \min\{n : S_n = -J \text{ or } S_n = K\}.$$

As in Example 5.5, we have

$$0 = \mathbb{E}[S_0] = \mathbb{E}[S_T] = [1 - \mathbb{P}\{S_T = K\}] \cdot (-J) + \mathbb{P}\{S_T = K\} \cdot K,$$

and solving gives

$$\mathbb{P}\{S_T = K\} = \frac{J}{J+K}.$$

In Exercise 5.9 it is shown that there exists $C < \infty$ such that for all n, $\mathbb{E}[M_{n\wedge T}^2] \leq C$. Hence we can use Theorem 5 to conclude that

$$0 = \mathbb{E}[M_0] = \mathbb{E}\left[M_T\right] = \mathbb{E}\left[S_T^2\right] - \mathbb{E}\left[T\right].$$

Moreover,

$$\begin{aligned} \mathbb{E}\left[S_T^2\right] &= J^2\, \mathbb{P}\{S_T = -J\} + K^2\, \mathbb{P}\{S_T = K\} \\ &= J^2\, \frac{K}{J+K} + K^2\, \frac{J}{J+K} = JK. \end{aligned}$$

Therefore,

$$\mathbb{E}[T] = \mathbb{E}\left[S_T^2\right] = JK.$$

In particular, the expected amount of time for the random walker starting at the origin to get distance K from the origin is K^2.

Example 5.7. As in Example 5.6, let $S_n = X_1 + \cdots + X_n$ be simple random walk starting at 0. Let

$$T = \min\{n : S_n = 1\}, \quad T_J = \min\{n : S_n = 1 \text{ or } S_n = -J\}.$$

Note that $T = \lim_{J\to\infty} T_J$ and

$$\mathbb{P}\{T = \infty\} = \lim_{J\to\infty} \mathbb{P}\left\{S_{T_J} = -J\right\} = \lim_{J\to\infty} \frac{1}{J+1} = 0.$$

Therefore, $\mathbb{P}\{T < \infty\} = 1$, although Example 5.6 shows that for every J,

$$\mathbb{E}[T] \geq \mathbb{E}\left[T_J\right] = J,$$

and hence $\mathbb{E}[T] = \infty$. Also, $S_T = 1$, so we do not have $\mathbb{E}[S_0] = \mathbb{E}[S_T]$. From this we can see that (10) and (11) are not satisfied by this example.

5.3 Martingale convergence theorem

The martingale convergence theorem describes the behavior of a martingale M_n as $n \to \infty$.

Theorem 6 (Martingale Convergence Theorem)**.** *Suppose M_n is a martingale with respect to $\{\mathcal{F}_n\}$ and there exists $C < \infty$ such that $\mathbb{E}[|M_n|] \leq C$ for all n. Then there exists a random variable M_∞ such that with probability one*

$$\lim_{n\to\infty} M_n = M_\infty.$$

It does *not* follow from the theorem that $\mathbb{E}[M_\infty] = \mathbb{E}[M_0]$. For example, the martingale betting strategy satisfies the conditions of the theorem since

$$\mathbb{E}[|W_n|] = (1 - 2^{-n}) \cdot 1 + (2^n - 1) \cdot 2^{-n} \leq 2.$$

However, $W_\infty = 1$ and $W_0 = 0$.

We will prove the martingale convergence theorem. The proof uses a well-known financial strategy — buy low, sell high. Suppose $M_0, M_1, \ldots$ is a martingale such that

$$\mathbb{E}[|M_n|] \leq C < \infty,$$

for all n. Suppose $a < b$ are real numbers. We will show that it is impossible for the martingale to fluctuate infinitely often below a and above b. Define a sequence of stopping times by

$$S_1 = \min\{n : M_n \leq a\}, \quad T_1 = \min\{n > S_1 : M_n \geq b\},$$

and for $j > 1$,

$$S_j = \min\{n > T_{j-1} : M_n \leq a\},$$
$$T_j = \min\{n > S_j : M_n \geq b\}.$$

We set up the discrete stochastic integral

$$W_n = \sum_{k=0}^{n} B_k \, [M_k - M_{k-1}],$$

with $B_n = 0$ if $n - 1 < S_1$ and

$$B_n = 1 \quad \text{if } S_j \leq n - 1 < T_j,$$

$$B_n = 0 \quad \text{if } T_j \leq n - 1 < S_{j+1}.$$

In other words, every time the "price" drops below a we buy a unit of the asset and hold onto it until the price goes above b at which time we sell. Let U_n denote the number of times by time n that we have seen a fluctuation; that is,

$$U_n = j \quad \text{if} \quad T_j < n \leq T_{j+1}.$$

We call U_n the number of *upcrossings* by time n. Every upcrossing results in a profit of at least $b-a$. From this we see that

$$W_n \geq U_n\,(b-a) - (a-M_n)_+ \geq U_n\,(b-a) - |a-M_n|.$$

The term $(a-M_n)_+$ represents a possible loss caused by holding a share of the asset at the current time (if we are holding a share and M_n is less than the price we paid for it). Since W_n is a martingale, we know that $\mathbb{E}[W_n] = \mathbb{E}[W_0] = 0$, and hence

$$\mathbb{E}[U_n] \leq \frac{\mathbb{E}[|a-M_n|]}{b-a} \leq \frac{|a| + \mathbb{E}[|M_n|]}{b-a} \leq \frac{|a|+C}{b-a}.$$

This holds for every n, and hence

$$\mathbb{E}[U_\infty] \leq \frac{|a|+C}{b-a} < \infty.$$

In particular with probability one, $U_\infty < \infty$, and hence there are only a finite number of fluctuations. We now allow a, b to run over all rational numbers to see that with probability one,

$$\liminf_{n\to\infty} M_n = \limsup_{n\to\infty} M_n.$$

Therefore, the limit

$$M_\infty = \lim_{n\to\infty} M_n$$

exists. We have not yet ruled out the possibility that M_∞ is $\pm\infty$, but it is not difficult to see that if this occurred with positive probability, then $\mathbb{E}[|M_n|]$ would not be uniformly bounded.

To illustrate the martingale convergence theorem, we will consider another example of a martingale called Pólya's urn. Suppose we have an urn with red and green balls. At time $n=0$, we start with one red ball and one green ball. At each positive integer time we choose a ball at random from the urn (with each ball equally likely to be chosen), look at the color of the ball, and then put the ball back in with another ball of the same color. Let R_n, G_n denote the number of red and green balls in the urn after the draw at time n so that

$$R_0 = G_0 = 1, \qquad R_n + G_n = n+2,$$

and let

$$M_n = \frac{R_n}{R_n + G_n} = \frac{R_n}{n+2}$$

be the fraction of red balls at this time. Let $\mathcal{F}_n$ denote the information in the data $M_1, \ldots, M_n$, which one can check is the same as the information in $R_1, R_2, \ldots, R_n$. Note that the probability that a red ball is chosen at time n depends only on the number (or fraction) of red balls in the urn before choosing. It does not depend on what order the red and green balls were put in. This is an example of the Markov property that we discussed in Section 4.

We can describe the rule of Pólya's urn by

$$\mathbb{P}\{R_{n+1} = R_n + 1 \mid \mathcal{F}_n\} = 1 - \mathbb{P}\{R_{n+1} = R_n \mid \mathcal{F}_n\} =$$

$$\mathbb{P}\{R_{n+1} = R_n + 1 \mid M_n\} = \frac{R_n}{n+2} = M_n.$$

We claim that M_n is a martingale with respect to $\mathcal{F}_n$. To check this,

$$\begin{aligned} E\left[M_{n+1} \mid \mathcal{F}_n\right] &= E\left[M_{n+1} \mid M_n\right] \\ &= M_n \frac{R_n + 1}{n+3} + [1 - M_n] \frac{R_n}{n+3} \\ &= \frac{R_n(R_n+1)}{(n+2)(n+3)} + \frac{(n+2-R_n)R_n}{(n+2)(n+3)} \\ &= \frac{R_n(n+3)}{(n+2)(n+3)} = M_n. \end{aligned}$$

Since $\mathbb{E}[|M_n|] = \mathbb{E}[M_n] = \mathbb{E}[M_0] = 1/2$, this martingale satisfies the conditions of the martingale convergence theorem. (In fact, the same argument shows that every martingale that stays nonnegative satisfies the conditions.) Hence, there exists a random variable M_∞ such that with probability one,

$$\lim_{n\to\infty} M_n = M_\infty.$$

It turns out that the random variable M_n is really random in the sense that it has a nontrivial distribution. In Exercise 5.4 you will show that for each n, the distribution of M_n is uniform on

$$\left\{\frac{1}{n+2}, \frac{2}{n+2}, \ldots, \frac{n+1}{n+2}\right\},$$

and from this it is not hard to see that M_∞ has a uniform distribution on $[0,1]$. You will also be asked to simulate this process to see what happens. There is a lot of randomness in the first few draws to see what fraction of red balls the urn will settle down to. However, for large n this ratio changes very little; for example, the ratio after 2000 draws is very close to the ratio after 4000 draws.

While Pólya's urn seems like a toy model, it arises in a number of places. We will give an example from Bayesian statistics. Suppose that we perform independent trials of an experiment where the probability of success for each experiment is θ (such trials are called *Bernoulli trials*). Suppose that we do not know the value of θ, but want to try to deduce it by observing trials. Let $X_1, X_2, \ldots$ be independent random variables with

$$\mathbb{P}\{X_j = 1\} = 1 - \mathbb{P}\{X_j = 0\} = \theta.$$

The (strong) law of large numbers implies that with probability one,

$$\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = \theta. \tag{12}$$

Hence, if were able to observe infinitely many trials, we could deduce θ exactly.

Clearly, we cannot deduce θ with 100% assurance if we see only a finite number of trials. Indeed, if $0 < \theta < 1$, there is always a chance that the first n trials will all be failures and there is a chance they will all be successes. The Bayesian approach to statistics is to assume that θ is a random variable with a certain *prior distribution*. As we observe the data we update to a *posterior distribution*. We will assume we know nothing initially about the value and choose the prior distribution to the uniform distribution on $[0, 1]$ with density

$$f_0(\theta) = 1, \quad 0 < \theta < 1.$$

Suppose that after observing n trials, we have had $S_n = X_1 + \cdots + X_n$ successes. If we know θ, then the distribution of S_n is binomial,

$$\mathbb{P}\{S_n = k \mid \theta\} = \binom{n}{k} \theta^k (1-\theta)^{n-k}.$$

We use a form of the Bayes rule to update the density

$$f_{n,k}(\theta) := f_n(\theta \mid S_n = k) = \frac{\mathbb{P}\{S_n = k \mid \theta\}}{\int_0^1 \mathbb{P}\{S_n = k \mid x\}\, dx} = C_{n,k}\, \theta^k (1-\theta)^{n-k},$$

where $C_{n,k}$ is the appropriate constant so that $f_{n,k}$ is a probability density. This is the *beta density* with parameters $k+1$ and $n-k+1$. The probability of a success on the $(n+1)$st trial given that $S_n = k$ is the conditional expectation of θ given $S_n = k$. A little computation which we omit shows that

$$\mathbb{E}[\theta \mid S_n = k] = \int_0^1 \theta\, f_{n,k}(\theta)\, d\theta = \frac{k+1}{n+2} = \frac{S_n+1}{n+2}.$$

These are exactly the transition probabilities for Pólya's urn if we view $S_n + 1$ as the number of red balls in the urn (S_n is the number of red balls added to the urn). The martingale convergence theorem can now be viewed as the law of large numbers (12) for θ. Even though we do not initially know the value of θ (and hence treat it as a random variable) we know that the conditional value of θ given $\mathcal{F}_n$ approaches θ.

Example 5.8. We end with a simple example where the conditions of the martingale convergence theorem do not apply. Let $S_n = X_1 + \cdots + X_n$ be simple symmetric random walk starting at the origin as in the previous section. Then one can easily see that $\mathbb{E}[|S_n|] \to \infty$. For this example, with probability one

$$\limsup_{n\to\infty} S_n = \infty,$$

$$\liminf_{n\to\infty} S_n = -\infty.$$

5.4 Square integrable martingales

Definition A martingale M_n is called square integrable if for each n, $\mathbb{E}\left[M_n^2\right] < \infty$.

Note that this condition is not as strong as (11). We do not require that there exists a $C < \infty$ such that $\mathbb{E}\left[M_n^2\right] \le C$ for each n. Mean zero random variables X, Y are orthogonal if $\mathbb{E}[XY] = 0$. Independent neab-zero random variables are orthogonal, but orthogonal random variables need not be independent. If $X_1, \ldots, X_n$ are pairwise orthogonal random variables with mean zero, then $\mathbb{E}[X_j X_k] = 0$ for $j \neq k$ and by expanding the square we can see that

$$\mathbb{E}\left[(X_1 + \cdots + X_n)^2\right] = \sum_{j=1}^{n} \mathbb{E}[X_j^2].$$

This can be thought of as a generalization of the Pythagorean theorem $a^2 + b^2 = c^2$ for right triangles. The increments of a martingale are not necessarily independent, but for square integrable martingales they are orthogonal as we now show.

Proposition 5.1. *Suppose that M_n is a square integrable martingale with respect to $\{\mathcal{F}_n\}$. Then if $m < n$,*

$$\mathbb{E}\left[(M_{n+1} - M_n)(M_{m+1} - M_m)\right] = 0.$$

Moreover, for all n,

$$\mathbb{E}[M_n^2] = \mathbb{E}[M_0^2] + \sum_{j=1}^{n} \mathbb{E}\left[(M_j - M_{j-1})^2\right].$$

Proof. If $m < n$, then $M_{m+1} - M_m$ is $\mathcal{F}_n$-measurable, and hence

$$\begin{aligned} &E\left[(M_{n+1} - M_n)(M_{m+1} - M_m) \mid \mathcal{F}_n\right] \\ &\qquad = (M_{m+1} - M_m)\, E[M_{n+1} - M_n \mid \mathcal{F}_n] = 0. \end{aligned}$$

Hence

$$\begin{aligned} &\mathbb{E}\left[(M_{n+1} - M_n)(M_{m+1} - M_m)\right] \\ &\qquad = \mathbb{E}\left[E\left[(M_{n+1} - M_n)(M_{m+1} - M_m) \mid \mathcal{F}_n\right]\right] = 0. \end{aligned}$$

Also, if we set $M_{-1} = 0$,

$$\begin{aligned} M_n^2 &= \left[M_0 + \sum_{j=1}^{n}(M_j - M_{j-1})\right]^2 \\ &= M_0^2 + \sum_{j=1}^{n}(M_j - M_{j-1})^2 + \sum_{j\neq k}(M_j - M_{j-1})(M_k - M_{k-1}). \end{aligned}$$

Taking expectations of both sides gives the second conclusion. $\square$

The natural place to discuss the role of orthogonality in the study of square integrable martingales is $L^2 = L^2(\Omega, \mathcal{F}, \mathbb{P})$, the space of square integrable random variables. This is a (real) Hilbert space under the inner product

$$(X, Y) = \mathbb{E}[XY].$$

Two mean zero random variables are orthogonal if and only if $(X, Y) = 0$. The conditional expectation has a nice interpretation in L^2. Suppose Y is a square integrable random variable and $\mathcal{G}$ is a sub- σ-algebra. Then $L^2(\Omega, \mathcal{G}, \mathbb{P})$ is a closed subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$ and the conditional expectation $\mathbb{E}[Y \mid \mathcal{G}]$ is the same as the Hilbert space projection onto the subspace. It can also be characterized as the $\mathcal{G}$-measurable random variable Z that minimizes the mean-squared error

$$\mathbb{E}[(Y - Z)^2].$$

The reason L^2 rather than L^p for other values of p is so useful is because of the inner product which gives the idea of orthogonality.

5.5 Integrals with respect to random walk

Suppose that $X_1, X_2, \ldots$ are independent, identically distributed random variables with mean zero and variance σ^2. The two main examples we will use are:

- Binomial or coin-tossing random variables,

$$\mathbb{P}\{X_j = 1\} = \mathbb{P}\{X_j = -1\} = \frac{1}{2},$$

  in which case $\sigma^2 = 1$.

- Normal increments where $X_j \sim N(0, \sigma^2)$. We write $Z \sim N(\mu, \sigma^2)$ if Z has a normal distribution with mean μ and variance σ^2.

Let $S_n = X_1 + \cdots + X_n$ and let $\{\mathcal{F}_n\}$ denote the filtration generated by $X_1, \ldots, X_n$. A sequence of random variables $J_1, J_2, \ldots$ is called predictable (with respect to $\{\mathcal{F}_n\}$) if for each n, J_n is $\mathcal{F}_{n-1}$-measurable. Recall that this is the condition that makes J_n allowable "bets" on the martingale in the sense of the discrete stochastic integral.

Suppose $J_1, J_2, \ldots$ is a predictable sequence with $\mathbb{E}[J_n^2] < \infty$ for each n. The integral of J_n with respect to S_n is defined by

$$Z_n = \sum_{j=1}^{n} J_j \, X_j = \sum_{j=1}^{n} J_j \, \Delta S_j.$$

There are three important properties that the integral satisfies.

- Martingale property. The integral Z_n is a martingale with respect to $\{\mathcal{F}_n\}$. We showed this in Section 5.

- Linearity. If J_n, K_n are predictable sequences and a, b constants, then $aJ_n + bK_n$ is a predictable sequence and

$$\sum_{j=1}^{n}(aJ_j + bK_j)\,\Delta S_j = a\sum_{j=1}^{n} J_j\,\Delta S_j + b\sum_{j=1}^{n} K_j\,\Delta S_j.$$

  This is immediate.

- Variance rule

$$\mathrm{Var}\left[\sum_{j=1}^{n} J_j\,\Delta S_j\right] = \mathbb{E}\left[(\sum_{j=1}^{n} J_j\,\Delta S_j)^2\right] = \sigma^2\sum_{j=1}^{n}\mathbb{E}\left[J_j^2\right].$$

  To see this we first use the orthogonality of martingale increments to write

$$\mathbb{E}\left[(\sum_{j=1}^{n} J_j\,\Delta S_j)^2\right] = \sum_{j=1}^{n}\mathbb{E}\left[J_j^2 X_j^2\right].$$

  Since J_j is $\mathcal{F}_{j-1}$-measurable and X_j is independent of $\mathcal{F}_{j-1}$, we can see that

$$\begin{aligned}
\mathbb{E}\left[J_j^2 X_j^2\right] &= \mathbb{E}\left[E[J_j^2 X_j^2 \mid \mathcal{F}_{j-1}]\right] \\
&= \mathbb{E}\left[J_j^2\, E[X_j^2 \mid \mathcal{F}_{j-1}]\right] \\
&= \mathbb{E}\left[J_j^2\,\mathbb{E}[X_j^2]\right] = \sigma^2\,\mathbb{E}[J_j^2].
\end{aligned}$$

Exercises

Exercise 5.1. Suppose $S_n = X_1 + \cdots + X_n$ is simple symmetric random walk in one dimension. Let $\mathcal{F}_n$ denote the information in $X_1, X_2, \ldots, X_n$. For each of the following say if the process is a martingale, submartingale, or supermartingale (it can be more than one and it might be none of these) with respect to $\mathcal{F}_n$. Give reasons (citing a fact from lecture or notes is a sufficient reason).

1. $M_n = S_n$
2. $M_n = S_n^2$
3. $M_n = S_n^3$
4. $M_n = e^{S_n}$.
5. $M_n = S_{n+1} - S_n$.

6.

$$M_n = S_0\, X_1 + S_1\, X_2 + \cdots + S_{n-1}\, X_n = \sum_{j=1}^{n} S_{j-1}\, X_j .$$

Exercise 5.2. In the same set-up as the previous exercise, state which of the following are stopping times for the random walk. Give reasons.

1. T is the first time n such that $S_n < 0$.
2. T is the first time n such that $S_n > S_{2n}$.
3. T is the first time that
$$\frac{S_n}{n} > S_1.$$
4. Let τ be the first time m that $S_m \geq 4$ and let T be the first time n after τ that $S_n \leq -5$.

Exercise 5.3. Here are some statements about martingales. Say whether they are always true. If always true give reason (citing a fact from the lecture or notes is fine). If it is not always true give an example to show this. Let $M_n, n = 0, 1, 2, \ldots$ be a martingale with respect to $\{\mathcal{F}_n\}$ with $M_0 = 1$.

1. For all positive integers n, $\mathbb{E}[M_n] = 1$.
2. With probability one, the limit
$$M_\infty := \lim_{n\to\infty} M_n \tag{13}$$
exists and is finite.
3. Suppose the limit M_∞ exists as in (13) and is finite. Then $\mathbb{E}[M_\infty] = 1$.
4. Suppose we assume know that with probability one $M_n \geq 0$ for all n? Does this imply that the limit in (13) exists with probability one?
5. If we assume that $M_n \geq 0$ for all n does the answer to part 3 change?

Exercise 5.4. This exercise concerns Pólya's urn and has a computing/simulation component. Let us start with one red and one green ball as in the lecture and let M_n be the fraction of red balls at the nth stage.

1. Show that the distribution of M_n is uniform on the set
$$\left\{\frac{1}{n+2}, \frac{2}{n+2}, \ldots, \frac{n+1}{n+2}\right\}.$$
(Use mathematical induction, that is, note that it is obviously true for $n = 0$ and show that if it is true for n then it is true for $n + 1$.)

2. Write a short program that will simulate this urn. Each time you run the program note the fraction of red balls after 1000 draws and after 2000 draws. Compare the two fractions. Then, repeat this twenty times.

Exercise 5.5. In this exercise, we consider simple, asymmetric independent random variables with

$$\mathbb{P}\{X_j = 1\} = 1 - \mathbb{P}\{X_j = -1\} = q, \qquad \frac{1}{2} < q < 1.$$

Let $S_0 = 0$ and $S_n = X_1 + \cdots + X_n$. Let $\mathcal{F}_n$ denote the information contained in $X_1, \ldots, X_n$.

1. Which of these is S_n: martingale, submartingale, supermartingale (more than one answer is possible)?

2. For which values of r is $M_n = S_n - rn$ a martingale?

3. Let $\theta = (1-q)/q$ and let
$$M_n = \theta^{S_n}.$$
Show that M_n is a martingale.

4. Let a, b be positive integers, and
$$T_{a,b} = \min\{j : S_j = b \text{ or } S_j = -a\}.$$
Use the optional sampling theorem to determine
$$\mathbb{P}\left\{S_{T_{a,b}} = b\right\}.$$

5. Let $T_a = T_{a,\infty}$. Find
$$\mathbb{P}\{T_a < \infty\}.$$

Exercise 5.6. Let $X_1, X_2, \ldots$ be independent, identically distributed random variables with

$$\mathbb{P}\{X_j = 2\} = \frac{1}{3}, \quad \mathbb{P}\{X_j = \frac{1}{2}\} = \frac{2}{3}.$$

Let $M_0 = 1$ and for $n \geq 1$, $M_n = X_1 X_2 \cdots X_n$.

1. Show that M_n is a martingale.

2. Explain why M_n satisfies the conditions of the martingale convergence theorem.

3. Let $M_\infty = \lim_{n\to\infty} M_n$. Explain why $M_\infty = 0$. (Hint: there are at least two ways to show this. One is to consider $\log M_n$ and use the law of large numbers. Another is to note that with probability one M_{n+1}/M_n does not converge.)

4. Use the optional sampling theorem to determine the probability that M_n ever attains a value as large as 64.

5. Does there exist a $C < \infty$ such that $\mathbb{E}[M_n^2] \leq C$ for all n?

Exercise 5.7. Let $X_1, X_2, \ldots$ be independent, identically distributed random variables with

$$\mathbb{P}\{X_j = 1\} = q, \quad \mathbb{P}\{X_j = -1\} = 1 - q.$$

Let $S_0 = 0$ and for $n \geq 1$, $S_n = X_1 + X_2 + \cdots + X_n$. Let $Y_n = e^{S_n}$.

1. For which value of q is Y_n a martingale?
2. For the remaining parts of this exercise assume q takes the value from part 1. Explain why Y_n satisfies the conditions of the martingale convergence theorem.
3. Let $Y_\infty = \lim_n Y_n$. Explain why $Y_\infty = 0$. (Hint: there are at least two ways to show this. One is to consider $\log Y_n$ and use the law of large numbers. Another is to note that with probability one Y_{n+1}/Y_n does not converge.)
4. Use the optional sampling theorem to determine the probability that Y_n ever attains a value greater than 100.
5. Does there exist a $C < \infty$ such that $\mathbb{E}[Y_n^2] \leq C$ for all n?

Exercise 5.8. Consider the martingale betting strategy as discussed in Section 5. Let W_n be the "winnings" at time n, which for positive n equals either 1 or $1 - 2^n$.

1. Is W_n a square integrable martingale?
2. If $\Delta_n = W_n - W_{n-1}$ what is $\mathbb{E}[\Delta_n^2]$?
3. What is $\mathbb{E}[W_n^2]$?
4. What is $E(\Delta_n^2 \mid \mathcal{F}_{n-1})$?

Exercise 5.9. Suppose $S_n = X_1 + \cdots + X_n$ is simple random walk starting at 0. For any K, let

$$T = \min\{n : |S_n| = K\}.$$

1. Explain why for every j,

$$\mathbb{P}\{T \leq j + K \mid T > j\} \geq 2^{-K}.$$

2. Show that there exists $c < \infty, \alpha > 0$ such that for all j,

$$\mathbb{P}\{T > j\} \leq c\, e^{-\alpha j}.$$

Conclude that $\mathbb{E}[T^r] < \infty$ for every $r > 0$.

3. Let $M_n = S_n^2 - n$. Show there exists $C < \infty$ such that for all n,

$$\mathbb{E}\left[M_{n \wedge T}^2\right] \leq C.$$

Exercise 5.10. Suppose that $X_1, X_2, \ldots$ are independent random variables with $\mathbb{E}[X_j] = 0$, $\mathrm{Var}[X_j] = \sigma_j^2$, and suppose that

$$\sum_{n=1}^{\infty} \sigma_n^2 < \infty.$$

Let $S_0 = 0$ and $S_n = X_1 + \cdots + X_n$ for $n > 0$. Let $\mathcal{F}_n$ denote the information contained in $X_1, \ldots, X_n$.

1. Show that S_n is a martingale with respect to $\{\mathcal{F}_n\}$.

2. Show that there exists $C < \infty$ such that for all n, $\mathbb{E}[S_n^2] \leq C$.

3. Show that with probability one the limit

$$S_\infty = \lim_{n \to \infty} S_n,$$

exists.

4. Show that

$$\mathbb{E}[S_\infty] = 0, \quad \mathrm{Var}[S_\infty] = \sum_{n=1}^{\infty} \sigma_n^2.$$

6 (Markovian) Jump Processes

We will consider continuous time processes X_t, that is, for each $t \geq 0$, X_t is a random variable. We make two assumptions.

- Simple jump process: by this we mean that the process starts at a point, and then stays there for a (random) positive amount of time until it decides to move to a new state. After moving to the new state, it will stay there for a while and then chooses another state, etc.

- (Homogeneous) Markov process. Given that the process is in state x at time t ($X_t = x$), the distribution of the future depends only on x and not on how one got to site x or how long it took to get there. More precisely, the distribution of $\{X_s : s \geq t\}$ given $\{X_r : r \leq t\}$ is the same as the distribution given just X_t. Moreover, the distribution of $X_{t+s}, s \geq 0$ given $X_t = x$ is the same as the distribution of $X_s, s \geq 0$ given $X_0 = x$.

Sometimes it is useful to consider these processes as a two-stage process: first the process decides when it wants to make a move and then chooses which site to jump to. There are other times when it is easier to frame it as a single process.

6.1 Waiting times: exponential random variables

The (time homogeneous) Markov assumption implies that the waiting times have the memoryless property (this is sometimes also called a Markov property). Suppose we are at state x and we let T be the time until we move. The memoryless property is

$$\mathbb{P}\{T > t + s \mid T > t\} = \mathbb{P}\{T > s\}.$$

This assumption implies (see August review notes) that the waiting time T must have an exponential distribution with $\lambda = \lambda_x$, that is, it has density

$$f(t) = \lambda\, e^{-\lambda t}, \quad t > 0,$$

and expectation $\mathbb{E}[T] = 1/\lambda$. Another way of saying this is

$$\mathbb{P}\{T > t\} = e^{-\lambda t}.$$

There is an important property of exponential random variables. Suppose $T_1, T_2, \ldots, T_n$ are independent exponential random variables with rates $\lambda_1, \lambda_2, \ldots, \lambda_n$. Let

$$T = \min\{T_1, T_2, \ldots, T_n\}.$$

We often view the exponential random variables as the times at which "alarm clocks" go off, and then T is the time until the first alarm goes off.

Fact The distribution of T is exponential with rate $\lambda := \lambda_1 + \cdots + \lambda_n$. Moreover, the probability that the kth alarm is the first one to off is λ_k/λ, that is,

$$\mathbb{P}\{T = T_k\} = \frac{\lambda_k}{\lambda}. \tag{14}$$

This is not hard to show if we realize that the event $\{T > t\}$ is the same as the event $\{T_1 > t, T_2 > t, \ldots, T_n > t\}$.

$$\begin{aligned}
\mathbb{P}\{T > t\} &= \mathbb{P}\{T_1 > t, T_2 > t, \ldots, T_n > t\} \\
&= \mathbb{P}\{T_1 > t\}\, \mathbb{P}\{T_2 > t\} \cdots \mathbb{P}\{T_n > t\} \\
&= e^{-\lambda_1 t}\, e^{-\lambda_2 t} \cdots e^{-\lambda_n t} \\
&= e^{-(\lambda_1 + \cdots + \lambda_n)t} = e^{-\lambda t}.
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}\{T = T_1\} &= \mathbb{P}\{T_2 > T_1, T_3 > T_1, \ldots, T_n > T_1\} \\
&= \int_0^\infty \mathbb{P}\{T_2 > t, T_3 > t, \ldots, T_n > t\}\, d\mathbb{P}\{T_1 = t\} \\
&= \int_0^\infty [e^{-\lambda_2 t}\, e^{-\lambda_3 t} \cdots e^{-\lambda_n t}]\, \lambda_1\, e^{-\lambda_1 t}\, dt \\
&= \frac{\lambda_1}{\lambda_1 + \cdots + \lambda_n}.
\end{aligned}$$

6.2 Poisson process

The Poisson process is one of the basic jump process. There are many ways of viewing the Poisson process. We will start by going fishing. Suppose we are out on the lake fishing on a Saturday morning. We make some assumptions. We will first write them down informally, and then be more mathematical.

- We start with 0 fish.
- The "rate" at which we catch stays constant. Let us say that on the average we catch λ fish per hour.
- If $s < t$, the number of fish that we catch between time s and time t does not depend on how our day has gone up to time s.
- We catch at most one fish at a time.

When we make this formal there are two equivalent ways to think about this. One is to let X_t denote the number of fish that we have caught by time t (measured in hours). Then the assumptions become

1. $X_0 = 0$.
2. For all t, the random variables $X_{t+s} - X_t, s \geq 0$ has the same distribution as $X_s, s \geq 0$.
3. For all t the random variables $X_{t+s} - X_t, s \geq 0$ are independent of the information up to time t, that is, independent of $\{X_r : r \leq t\}$.
4. If we catch a fish at time t then for all s just a bit smaller than t, $X_t = X_s + 1$. More precisely, if we consider X_t as a function of t, then it always takes integer values, is nondecreasing, and the size of the jumps is always equal to 1.

Conditions (1) – (3) are the hypotheses of a wider class of processes called Lévy processes. The last condition is particular to the Poisson process and in fact characterizes it among all Lévy processes.

The second way to view the Poisson process is in terms of the waiting times between fish. Let T_1 be the amount of time until the first fish is caught and for $j > 1$, let T_j be the amount of time between the $(j-1)$st fish and the jth fish. Then the assumptions are equivalent to saying that these times are independent, identically distributed, with rate λ and with the memoryless property. In terms of $\{X_t\}$ we let T_1 be the first t such that $X_t = 1$ and, more generally, T_k is the first t such that $X_{T_{k-1}+t} = k$.

- The waiting times $T_1, T_2, \ldots$ for the Poisson process with rate λ are independent each exponential with rate λ.

We now ask the question: for fixed t, what is the distribution of the random variable X_t? Since X_t takes values in the nonnegative integers, it suffices to give $p_t(k) = \mathbb{P}\{X_t = k\}$ for all such k. We will do this two different ways, and (thank goodness!) will get the same answer both ways.

For a large integer n we write

$$X_t = I_{1,n} + I_{2,n} + \cdots + I_{n,n}$$

where $I_{k,n} = X_{k/n} - X_{(k-1)/n}$, that is, the number of fish caught in the small time interval $[\frac{(k-1)t}{n}, \frac{kt}{n}]$. Since we only catch one fish at a time, for large n we expect that $I_{k.n}$ equals zero or one. The expected number of fish caught in one of the small intervals is $\lambda t/n$ and hence X_t is well approximated by a binomial random variable with parameters n and $p = \lambda t/n$. Using the binomial distribution we see that for large n,

$$\mathbb{P}\{X_t = k\} \sim \binom{n}{k} \left(\frac{\lambda t}{n}\right)^k \left(1 - \frac{\lambda t}{n}\right)^{n-k}.$$

This approximation gets better and better as $n \to \infty$ and so we get

$$\mathbb{P}\{X_t = k\} = \lim_{n\to\infty} \binom{n}{k} \left(\frac{\lambda t}{n}\right)^k \left(1 - \frac{\lambda t}{n}\right)^{n-k} = e^{-\lambda t} \frac{(\lambda t)^k}{k!}.$$

The limit is not difficult if we remember some standard limits from calculus. Note that we can write

$$\binom{n}{k} \left(\frac{\lambda t}{n}\right)^k \left(1 - \frac{\lambda t}{n}\right)^{n-k} = \frac{(\lambda t)^k}{k!} \left[\frac{n \cdot (n-1) \cdots (n-k+1)}{\underbrace{n \cdot n \cdots n}_{k}} \left(1 - \frac{\lambda t}{n}\right)^{-k}\right] \left(1 - \frac{\lambda t}{n}\right)^n.$$

For fixed k and t, as $n \to \infty$ the term in the square brackets goes to 1 and the final term goes to $e^{-\lambda t}$.

More generally,

- The distribution of $X_{s+t} - X_s$ is Poisson with mean λt.

This is why the process is called the Poisson process.

Poisson is also the French word for fish, but this is just a coincidence!

Let us summarize by restating the definition using the facts we have derived.

Definition A Poisson process with rate λ is a collection of random variables $\{X_t, t \geq 0\}$ satisfying the following.

- $X_0 = 0$.
- For each t, the random variables $X_{t+s} - X_t, s \geq 0$ are independent of $\{X_r : r \leq t\}$.
- For each s, t, $X_{s+t} - X_t$ has a Poisson distribution with mean λs.

Another way to derive the Poisson distribution is to use differential equations. Let $p_t(k) = \mathbb{P}\{X_t = k\}$ and view $p_t(k)$ as an infinite collection of functions of t indexed by k. We have the initial condition

$$p_0(0) = 1, \qquad p_0(k) = 0 \text{ if } k > 0. \tag{15}$$

Let us compute

$$\frac{d}{dt}\, p_t(k) = \lim_{\Delta t \to 0} \frac{p_{t+\Delta t}(k) - p_t(k)}{\Delta t}.$$

Let us consider Δt positive but very small. Since the time increment is small we expect at most one jump in the time interval. One way to be at k at time $t + \Delta t$ is to be at $k-1$ at time t and have a jump: the probability of this is about $\lambda \Delta t$. The other possibility is to be at k at time t and stay there. The chance of leaving k would be about $\lambda\, \Delta t$ and hence we get the expression

$$p_{t+\Delta t}(k) = p_t(k-1)\, \lambda\, \Delta t + p_t(k)\, [1 - \lambda\, \Delta t] + (\text{small error})$$

where the small error if of smaller order than Δt. Plugging into the difference equation for $p_t(k)$ we get

$$\frac{d}{dt}\, p_t(k) = \lambda\, p_t(k-1) - \lambda\, p_t(k). \tag{16}$$

We need to solve this with the initial conditions (15). For $k = 0$, the equation is just

$$\frac{d}{dt}\, p_t(k) = -\lambda\, p_t(0), \qquad p_0(0) = 1,$$

($p_t(-1) = 0$ for all t by definition). This is the exponential differential equation and has the solution $p_t(0) = e^{-\lambda t}$. We can then solve the other equations one by one. While we could do this, we will take advantage of the fact that we already know what the answer should be

$$p_t(k) = e^{-\lambda t}\, \frac{(\lambda t)^k}{k!}.$$

By differentiating, one can check that these satisfy (16) with initial value (15).

The Poisson process is the prototypical process modeling “rare events”. We think of this in terms of the approximation by the binomial. If n is large and k is an integer, then the event “jump at a time between $\frac{k}{n}$ and $\frac{k+1}{n}$” is rare. In time 1, say, one gets n chances ("trials") but the probability of success for each one is about λ/n. Another thing modeled by Poisson processes is customers arriving at a store, assuming they arrive one at a time. We will use this to go through two examples. We will state the problems and then go through the solutions.

Example 6.1. Suppose that the number of customers arriving at an automobile dealership in Chicago follows a Poisson process with $\lambda = 2.5$ (time is measured in hours).

a. What is the probability that at most two customers arrive in the first hour?

   X_1 **has a Poisson distribution with mean** $2.5 \cdot 1 = 2.5$**.**

$$\mathbb{P}\{X_1 \leq 2\} = e^{-2.5} + \frac{2.5}{1!}e^{-2.5} + \frac{2.5^2}{2!}\,e^{-2.5} = .5438$$

b. Suppose that exactly two customers arrive in the first hour. What is the probability that there will be exactly three customers in the second hour?

   The question is asking

$$\mathbb{P}\{X_2 - X_1 = 3 \mid X_1 = 2\}.$$

   The number arriving in different intervals are independent so this is the same as the unconditioned probability $\mathbb{P}\{X_2 - X_1 = 3\}$**. Note that** $X_2 - X_1$ **has a Poisson distribution with mean** 2.5**.**

$$\mathbb{P}\{X_2 - X_1 = 3\} = e^{-2.5}\,\frac{(2.5)^3}{3!} = .2138$$

c. Suppose that exactly four customers arrived in the first two hours. What is the probability that exactly two customers arrived in the first hour?

   The question is asking about $\mathbb{P}\{X_1 = 2 \mid X_2 = 4\}$**.** X_1 **and** X_2 **are not independent so we go back to the definition of conditional probability.**

$$\begin{aligned}
\mathbb{P}\{X_1 = 2 \mid X_2 = 4\} &= \frac{\mathbb{P}\{X_1 = 2, X_2 = 4\}}{\mathbb{P}\{X_2 = 4\}} \\
&= \frac{\mathbb{P}\{X_1 = 2\}\,\mathbb{P}\{X_2 - X_1 = 2\}}{\mathbb{P}\{X_2 = 4\}} \\
&= \frac{e^{-2.5}\,\frac{(2.5)^2}{2!}\cdot e^{-2.5}\,\frac{(2.5)^2}{2!}}{e^{-5}\,\frac{(5)^4}{4!}} \\
&= \frac{4!}{2!\,2!}(1/2)^4 = \frac{3}{8}
\end{aligned}$$

d. An enthusiastic salesperson decides to wait until 10 customers have arrived before going to lunch. What is the expected amount of time she will have to wait?

The time until 10 **customers have arrived is** $T_1+T_2+\cdots+T_{10}$ **where** $T_1, \cdots, T_{10}$ **are independent random variables each exponential with rate** $\lambda = 2.5$**. Note that** $\mathbb{E}[T_j] = 1/\lambda = 2/5$ **and hence**

$$\mathbb{E}[T_1 + \cdots + T_{10}] = \mathbb{E}[T_1] + \cdots + \mathbb{E}[T_{10}] = 10 \cdot \frac{2}{5} = 4.$$

e. Let N denote the number of customers that arrive in the first two hours. Find $\mathbb{E}[N^2]$.

Note that N **has Poisson distribution with mean** $2{\cdot}2.5 = 5$**. You can calculate** $\mathbb{E}[N^2]$ **directly or look up that if** N **is Poisson with rate** 5 **then** $\mathrm{Var}[N] = 5$**. Then** $\mathbb{E}[N^2] = \mathrm{Var}[N] + (\mathbb{E}[N])^2 = 30$.

Before doing the second example, we will give a property about Poisson processes. One may recall (or look up now) the following fact about Poisson random variables: if X, Y are independent Poisson random variables with means λ_X, λ_Y, respectively, then $X + Y$ has a Poisson distribution with mean $\lambda_X + \lambda_Y$. This idea extends to Poisson processes.

Fact If X_t, Y_t are independent Poisson processes with parameters λ_X and λ_Y respectively, and $Z_t = X_t + Y_t$, then Z_t is a Poisson process with parameter $\lambda_Z = \lambda_X + \lambda_Y$.

To prove this one just shows that Z_t satisfies the conditions for a Poisson process.

Example 6.2. This dealership is running a competition with a dealership in Milwaukee. The number of customers at the Milwaukee dealership is a Poisson process with parameter $\lambda = 4$. Since they are in different cities we will assume that the number of customers going to one dealership is independent of the number going to the other dealership. Both stores open the same time on a Saturday morning (Chicago and Milwaukee are both in the central time zone).

a. What is the probability that a customer arrives in the Chicago dealership before any customer arrives in the Milwaukee dealership?

Let $J_1, J_2, \ldots$ **be the waiting times at the Milwaukee dealership. These are independent exponentials with rate** $\lambda = 4$ **and we are assuming that these are independent of** $T_1, T_2, \ldots$ **the waiting times in Chicago. Using a property of exponential distributions (see** (14)**) we see that**

$$\mathbb{P}\{T_1 < J_1\} = \frac{2.5}{2.5 + 4} = \frac{5}{13}$$

b. What is the probability that after the first hour, there have been a total of exactly six customers in the two dealerships?

Note that customers arrive at either dealership at rate $2.5+4=6.5$. Therefore the total number of customers Z_t is a Poisson process with rate 6.5.

$$\mathbb{P}\{Z_1 = 6\} = e^{-6.5}\,\frac{(6.5)^6}{6!} = .1575.$$

c. Given that there were exactly six customers in that first hour, what is the probability that exactly two of them went to the Chicago dealership?

Each time a customer comes the probability that they are from the Chicago dealership is $5/13$. Using the binomial distribution

$$\binom{6}{2}\,(5/13)^2\,(8/13)^4 = .3182$$

d. The Milwaukee dealership agrees to call Chicago as soon as the first customer arrives in Milwaukee. Let X denote the number of customers that have arrived in Chicago by that time. Find the distribution of X, that is, find the numbers $\mathbb{P}\{X=k\}$ for each k.

This is the number of successes before a failure when a success is "Chicago gets a customer before Milwaukee". This is an example of a geometric distribution. The probability of success is $5/13$.

$$\mathbb{P}\{X=k\} = (5/13)^k\,(8/13).$$

6.3 Continuous time finite Markov chains

When modeling (time-homogeneous) Markov chains that visit a finite set of states, one can use either discrete times or continuous times. We discussed the discrete time case in Section 4, and we will discuss the continuous case here. In the discrete case, one describes the transitions with the transition matrix which gives the probabilities of moving to other points. In the continuous time case, we will give the rates at which we move.

Let S be the finite state space. For each $x, y \in S$ with $x \neq y$ we specify the rate $\alpha(x,y)$ of making a transition from x to y. We think of this as saying that the probability in a small amount of time Δt of jumping to y is about $\alpha(x,y)\,\Delta t$. Let

$$\alpha_x = \sum_{y \neq x} \alpha(x,y)$$

be the sum of all the rates going out of x. When the process is at x we can view the move to the next step in two equivalent ways.

- The process waits at x up to time T where T is exponential with rate α_x. When it decides to move, it choose state y with probability
$$\frac{\alpha(x,y)}{\alpha_x}.$$
- There are independent alarm clocks at each $y \neq x$. The time T_y until the clock associated to y goes off is exponential with rate $\alpha(x,y)$. The process waits until the first clock goes off which is at time
$$T = \min_y T_y,$$
and proceeds to the point associated to the clock that rings.

The analogue of the transition matrix is the (infinitesimal) generator matrix $\mathbf{A} = [a(x,y)]$ where
$$a(x,y) = \begin{cases} \alpha(x,y) & \text{if } x \neq y \\ -\alpha_x & \text{if } x = y \end{cases}.$$
Note that the rows of $\mathbf{A}$ add up to 0. The transition probabilities are defined by
$$p_t(x,y) = \mathbb{P}\{X_t = y \mid X_0 = x\} = \mathbb{P}\{X_{t+s} = y \mid X_s = x\}.$$
which can be written in matrix form
$$\mathbf{P}_t = [p_t(x,y)].$$
For a fixed t, $\mathbf{P}_t$ is a stochastic matrix. By definition we have $\mathbf{P}_0 = \mathbf{I}$, the identity matrix. The transition matrices satisfy the Chapman-Kolmogorov equations
$$p_{t+s}(x,y) = \sum_z p_t(x,z)\, p_s(z,y).$$
The reasoning is the same as the discrete case.

For the discrete time chain, the n-step transition probabilities were obtained by taking the transition matrix to the nth power. The analogous fact is a little trickier in the continuous time case. Here we write a differential equation in t, similarly to the way we did for the Poisson process. If we assume that $X_0 = x$ and we want to know if $X_{t+\Delta t} = y$ we consider the various possibilities for where we are at time t. We get
$$p_{t+\Delta t}(x,y) - p_t(x,y) \approx -\alpha_y\, p_t(x,y)\, \Delta t + \sum_{z \neq y} p_t(x,z)\, \alpha(z,y)\, \Delta t.$$
The first term on the right represents the probability of being at y at time t but changing sites before time $t + \Delta t$ and the second term represents the probability of not being at y at time t but being there at time $t + \Delta t$. This gives the differential equations
$$\frac{d}{dt}\, p_t(x,y) = -\alpha_y\, p_t(x,y) + \sum_{z \neq y} p_t(x,z)\, \alpha(z,y). \tag{17}$$

In matrix form this becomes the differential equation

$$\frac{d}{dt}\mathbf{P}_t = \mathbf{P}_t\,\mathbf{A}. \tag{18}$$

We could do a similar expression where we consider our position at Δt (rather than at t) and get a similar equation

$$\frac{d}{dt}\mathbf{P}_t = \mathbf{A}\mathbf{P}_t.$$

If

$$\mathbf{M_t} = [b_t(j,k)]$$

is a matrix depending on t, the the time derivative is taken entry-wise,

$$\frac{d}{dt}\,\mathbf{M_t} = \left[\frac{d}{dt}\,b_t(j,k)\right].$$

The equations (17) and (18) with initial condition $\mathbf{P}_0 = \mathbf{I}$ have the same solution which we will just give here. Recall that the solution to the simple equation $f'(t) = \lambda\, f(t)$ with initial condition $f(0) = 1$ is $f(t) = e^{\lambda t}$. The matrix solution is the same although we must use power series expression for the exponential to define the matrix

$$\mathbf{P}_t = e^{t\mathbf{A}} := \sum_{n=0}^{\infty} \frac{(t\mathbf{A})^n}{n!} = \sum_{n=0}^{\infty} \frac{t^n\,\mathbf{A}^n}{n!}.$$

Once one thinks of this, it is not hard to show that the right-hand side satisfies (17) and (18).

Given $\mathbf{A}$ there are ways to write e^{tA} explicitly, but we will not describe it here. Instead we will discuss analogues of some of the other results about discrete time chains. The continuous time chain is called irreducible if one can get from any state to any other state. This is the same as saying for every x, z we can find a sequence of points

$$x = y_0, y_1, y_2, \ldots, y_k = z$$

such that $\alpha(y_{j-1}, y_j) > 0$ for each j. It turns out that for an irreducible continuous time chain, $\mathbf{P}_t$ has all strictly positive entries — in other words, if one can get from x to z and $t > 0$, by choosing the waiting times appropriately (remembering that waiting times have a continuous distribution) we can arrange to be there at exactly time t.

For an irreducible chain there is a unique invariant probability vector $\boldsymbol{\pi}$ that represents the long-range average amount of time spent in each state. The "equilibrium" equations that the component satisfy is

$$\pi(x) = -\alpha_x\,\pi(x) + \sum_{y \neq x} \pi(y)\,\alpha(y,x).$$

The first term on the left represents the rate at which mass is leaving x and the second term gives the rate at which mass is coming in. Equilibrium means that these two rates are the same. In matrix form the equation is

$$\boldsymbol{\pi}\,\mathbf{A} = 0.$$

Here it is important that we are multiplying with $\boldsymbol{\pi}$ on the left.

We will now consider the mean passage times. Suppose $x \neq y$, $X_0 = x$ and let

$$T = T_y = \min\{t : X_t = y\}$$

be the first time that the process reaches the state y. Let $e(z)$ denote the expected time assuming $X_0 = z$ where $e(y) = 0$. Then we claim that the following holds for $x \neq y$,

$$e(x) = \frac{1}{\alpha_x} + \sum_{z \neq x} \frac{\alpha(x,z)}{\alpha x}\, e(z).$$

The first term on the right is the expected time until the process decides to change state for the first time. The second term gives the expected amount of time after the first jump until y is reached. Since $e(y) = 0$, the sum is really over all z not equal to x or y. Dividing by α_x we get

$$\alpha_x\, e(x) = 1 + \sum_{z \neq x,y} \alpha(x,z)\, e(z).$$

We can write this in matrix form

$$\tilde{\mathbf{A}}\,\mathbf{e} = -\mathbf{1}.$$

Here $\tilde{\mathbf{A}}$ is the matrix $\mathbf{A}$ with the row and column associated to y removed, and $\mathbf{1}$ denotes the vector all of whose components equals 1. We get

$$\mathbf{e} = [-\tilde{\mathbf{A}}]^{-1}\,\mathbf{1}.$$

If we consider the time starting at x to return to x, we must be careful. If we start at x, then we will stay there for a while before jumping. We can define T to be the time for the process starting at x to leave x for the first time and then return to x for the first time. We can work out $\mathbb{E}[T]$ in terms of the invariant probability with a little thought. Let us write

$$T = Y + Z$$

where Y is the time it takes for the process to leave x and Z is the remaining amount of time it takes to return. We know that

$$\mathbb{E}[Y] = \frac{1}{\alpha_x}.$$

We also know that the long-range amount of time spent in x is $\pi(x)$ which implies that

$$\frac{\mathbb{E}(Y)}{\mathbb{E}(Y+Z)} = \pi(x).$$

Solving gives

$$\mathbb{E}[T] = \frac{1}{\alpha_x \, \pi(x)}.$$

We summarize the last two paragraphs.

Fact Suppose X_t is an irreducible continuous time finite Markov chain with generator $\mathbf{A}$ and invariant probability $\boldsymbol{\pi}$. Suppose that $X_0 = x$.

- The expected amount of time for the process to leave x and then return for the first time is
$$\frac{1}{\alpha_x \, \pi(x)}$$
- If $y \neq x$, the expected amount of time for the process to reach y is the x-component of
$$[-\tilde{\mathbf{A}}]^{-1} \, \mathbf{1},$$
where $\tilde{\mathbf{A}}$ is the matrix $\mathbf{A}$ with the row and column associated to state y removed.

Let us do a particular example. Suppose we have five states $\{A, B, C, D, E\}$ and rates

$$\alpha(A,B) = 1, \quad \alpha(B,C) = 4, \quad \alpha(C,D) = 1, \quad \alpha(C,E) = 2,$$

$$\alpha(D,A) = 2, \quad \alpha(D,C) = 3, \quad \alpha(D,E) = 1, \quad \alpha(E,A) = 1, \quad \alpha(E,B) = 1$$

with all other rates equal to zero. This matrix is

$$\mathbf{A} = \begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array} \begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{pmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -4 & 4 & 0 & 0 \\ 0 & 0 & -3 & 1 & 2 \\ 2 & 0 & 3 & -6 & 1 \\ 1 & 1 & 0 & 0 & -2 \end{pmatrix} \end{array}.$$

Note that the diagonal entries have been chosen so that the rows add up to zero. If we want to find the invariant probability, $\pi \mathbf{A} = 0$, we get the equations

$$-\pi(A) + 2\,\pi(D) + \pi(E) = 0$$

$$\pi(A) - 4\,\pi(B) + \pi(E) = 0$$

$$4\,\pi(B) - 3\,\pi(C) + 3\,\pi(D) = 0$$

$$\pi(C) - 6\,\pi(D) = 0$$

$$2\,\pi(C) + \pi(D) - 2\,\pi(E) = 0.$$

This is a redundant set of equations. As in the case of the discrete chain, we solve these equations up to a multiplicative constant and then choose the constant so that the sum of the components equals one. In this case we get

$$\boldsymbol{\pi} = \begin{bmatrix} \frac{34}{103} & \frac{15}{103} & \frac{24}{103} & \frac{4}{103} & \frac{26}{103} \end{bmatrix}.$$

The long-range fraction of time spent in state A is $34/103$. If we start in state C, then the expected time until the first return to C is

$$\frac{1}{\alpha_C\,\pi(C)} = \frac{1}{3\,(14/103)} = \frac{103}{42}.$$

Let us consider the mean passage time to the state A. If we remove the row and column associated to state A we get matrix

$$\tilde{\mathbf{A}} = \begin{array}{c} \\ B \\ C \\ D \\ E \end{array} \begin{array}{c} \begin{array}{cccc} B & C & D & E \end{array} \\ \begin{pmatrix} -4 & 4 & 0 & 0 \\ 0 & -3 & 1 & 2 \\ 0 & 3 & -6 & 1 \\ 1 & 0 & 0 & -2 \end{pmatrix} \end{array}.$$

Either by hand or by computer we can compute

$$-\tilde{\mathbf{A}}^{-1} = \begin{array}{c} \\ B \\ C \\ D \\ E \end{array} \begin{array}{c} \begin{array}{cccc} B & C & D & E \end{array} \\ \begin{pmatrix} 15/34 & 12/17 & 2/17 & 13/17 \\ 13/68 & 12/17 & 2/17 & 13/17 \\ 9/68 & 7/17 & 4/17 & 9/17 \\ 15/68 & 6/17 & 1/17 & 15/17 \end{pmatrix} \end{array},$$

$$-\tilde{\mathbf{A}}^{-1}\,\mathbf{1} = \begin{array}{c} B \\ C \\ D \\ E \end{array} \begin{pmatrix} 69/34 \\ 121/68 \\ 89/68 \\ 108/68 \end{pmatrix}.$$

The expected amount of time starting at C to reach A is $121/68$.

6.4 Compound Poisson process

Here we discuss a generalization of the Poisson process. As before we let λ be the rate at which we decide to jump. However, when we jump rather than taking a step of $+1$ we jump a random amount taken from some probability distribution. In order to keep this a Lv́ey process, we assume that these random jumps are independent and from the same probability distribution.

To be more formal suppose N_t is a Poisson process with rate λ and

$$Y_1,\ Y_2,\ Y_3\ \cdots$$

are independent, identically distributed random variables. Then we define the process

$$X_t = Y_1 + Y_2 + \cdots + Y_{N_t},$$

where by definition $X_t = 0$ if $N_t = 0$. We call such a process a compound Poisson process. We can choose any distribution for the random jumps Y_j, discrete or continuous. To be specific, we will discuss the case where Y_j has a continuous distribution with density $f^\#$,

$$\mu^\#[a,b] := \mathbb{P}\{a \leq Y_j \leq b\} = \int_a^b f^\#(x)\,dx.$$

Another way of saying this is that the probability that Y_j lies between x and $x + \Delta x$ is exactly $\mu^\#[x, x+\Delta x]$ and this is approximately $f^\#(x)\,\Delta(x)$. If we combine this with the jumping rule, we can say that the probability that there is a jump between time t and time Δt with the value of the jump being between x and $x + \Delta x$ is about

$$\lambda\,\Delta t\, m^\#[x, x+\Delta x]\,\Delta x = \mu[x, x+\Delta x]\,\Delta t\,\Delta x \tag{19}$$

where $\mu[a,b] := \lambda\,\mu^\#[a,b]$ with density $f(x) = \lambda\, f^\#(x)$. The measure μ is called the Lévy measure and f the Lévy density.

What we have done here is similar to what we did for the continuous time Markov chain which we can view either as a two-step process

- The rate at which we take a jump combined with the probability distribution for the jump,

or as a one-step process

- The rate at which we are jumping to a site.

Here the "rate" at which we are making a jump of size x is $\lambda f(x)$ where, as in the case of the density for continuous random variables, one must understand this in the sense of (19). Compound Poisson processes are generally described in terms of the measure μ but from these we can get

$$\lambda = \int_{-\infty}^{\infty} f(x)\,dx, \qquad \mu^\# = \frac{\mu}{\lambda}.$$

Proposition 6.1. *If X_t is a compound Poisson process with Lévy density f and*

$$m = \int_{-\infty}^{\infty} x\, f(x)\,dx, \qquad \sigma^2 = \int_{-\infty}^{\infty} x^2\, f(x)\,dx < \infty,$$

then

$$\mathbb{E}[X_t] = mt, \qquad \mathrm{Var}[X_t] = \sigma^2\, t.$$

Proof. To show this, we will use the characteristic function. Let ϕ denote the characteristic function of the Y_j,

$$\phi(s) = \mathbb{E}[e^{isY_j}] = \int_{-\infty}^{\infty} e^{isx} f^{\#}(x)\, dx.$$

Recall that the characteristic function for $Y_1 + \cdots + Y_n$ is ϕ^n. Also, $\phi'(0) = i\,\mathbb{E}[Y_j], \phi''(0) = -\mathbb{E}[Y_j^2]$. Let ψ be the characteristic function for X_t. Then

$$\begin{aligned}
\psi(s) &= \mathbb{E}[e^{isX_t}] \\
&= \sum_{n=0}^{\infty} \mathbb{P}\{N_t = n\}\, \mathbb{E}[e^{isX_t} \mid N_t(n)] \\
&= \sum_{n=0}^{\infty} e^{-\lambda t}\, \frac{(\lambda t)^n}{n!}\, \phi(s)^n \\
&= \sum_{n=0}^{\infty} e^{-\lambda t}\, \frac{(\lambda t \phi(s))^n}{n!} \\
&= \exp\left\{\lambda t \phi(s) - \lambda t\right\} \\
&= \exp\left\{t\,\lambda\,[\phi(s) - 1]\right\}.
\end{aligned}$$

The result can now be computed directly using

$$\mathbb{E}[X_t] = i\,\psi'(0). \quad \mathrm{Var}[X_t] = \mathbb{E}[X_t^2] - (\mathbb{E}[X_t])^2 = -\,\left[\psi''(0) - \psi'(0)^2\right].$$

□

Exercises

Exercise 6.1. Suppose X_t is a Poisson process with parameter $\lambda = 2$. Find the following.

1. $\mathbb{P}\{X_3 = 7\}$
2. $\mathbb{P}\{X_2\,(X_4 - X_2) = 0\}$
3. The expected amount of time until $X_t = 6$
4. $\mathbb{P}\{X_1 = 3 \mid X_3 = 7\}$.

Exercise 6.2. Suppose that the number of customers arriving at an automobile dealership in Chicago follows a Poisson process with $\lambda = 3$ (time is measured in hours).

a. What is the probability that at most two customers arrive in the first hour?

b. Suppose that exactly two customers arrive in the first hour. What is the probability that there will be exactly three customers in the second hour?

c. Suppose that exactly four customers arrived in the first two hours. What is the probability that exactly two customers arrived in the first hour?

d. An enthusiastic salesperson decides to wait until 10 customers have arrived before going to lunch. What is the expected amount of time she will have to wait?

e. Let N denote the number of customers that arrive in the first two hours. Find $\mathbb{E}[N^2]$.

Exercise 6.3. This dealership is running a competition with a dealership in Houston. The number of customers at the Houston dealership is a Poisson process with parameter $\lambda = 2$. Since they are in different cities we will assume that the number of customers going to one dealership is independent of the number going to the other dealership. Both stores open the same time on a Saturday morning.

a. What is the probability that a customer arrives in the Chicago dealership before any customer arrives in the Houston dealership?

b. What is the probability that after the first hour, there have been a total of exactly five customers in the two dealerships?

c. Given that there were exactly five customers in that first hour, what is the probability that exactly two of them went to the Chicago dealership?

d. The Houston dealership agrees to call Chicago as soon as the first customer arrives in Houston. Let X denote the number of customers that have arrived in Chicago by that time. Find the distribution of X, that is, find the numbers $\mathbb{P}\{X = k\}$ for each k.

Note: both Chicago and Houston are in the central time zone.

Exercise 6.4. Suppose that X_n is an irreducible discrete time finite Markov chain on a state space S with transition matrix $\mathbf{P}$ and let Y_t be a continuous time Markov chain with generator

$$\mathbf{A} = \mathbf{P} - \mathbf{I}.$$

Explain why the following facts are true.

1. $\mathbf{A}$ is a generator of a continuous time Markov chain and hence this problem makes sense.

2. The invariant probability for $\mathbf{P}$ is the same as the invariant probability for $\mathbf{A}$.

3. If x, y are different states, then the expected amount of time to get from x to y is the same for both the discrete and the continuous chain.

4. If $p(x, x) = 0$ for all x, then the expected return time to x starting at x for the discrete chain is the same as the expected time for the first return to x for the continuous-time chain.

5. If $p(x,x) > 0$, then the last statement does not hold.

Exercise 6.5. Suppose X_t is a continuous time Markov chain with state space $\{0,1,2,3\}$ with rates

$$\alpha(0,1) = 2, \;\; \alpha(1,2) = 3, \;\; \alpha(2,0) = 1, \;\; \alpha(2,3) = 1, \;\; \alpha(3,1) = 4,$$

with all other rates equal to zero.

1. Write down the generator $\mathbf{A}$.
2. Is this chain irreducible?
3. What is the invariant probability?
4. Suppose we start with $X_0 = 0$. What is the expected amount of time until reaching state 3?
5. Suppose we start with $X_0 = 0$. What is the expected amount of time until the chain leaves state 0 for the first time?
6. Suppose we start with $X_0 = 0$. What is the expected amount of time for the chain to leave 0 and then return to 0 for the first time?

Exercise 6.6. Answer the same questions as in Exercise 6.5 with state space $\{0,1,2,3,4\}$ and rates

$$\alpha(0,1) = \alpha(1,0) = \alpha(1,2) = \alpha(2,1) = \alpha(2,3) = \alpha(3,2) = \alpha(3,4) = \alpha(4,3) = 1.$$

This is a continuous time version of simple symmetric random walk with reflecting boundary.