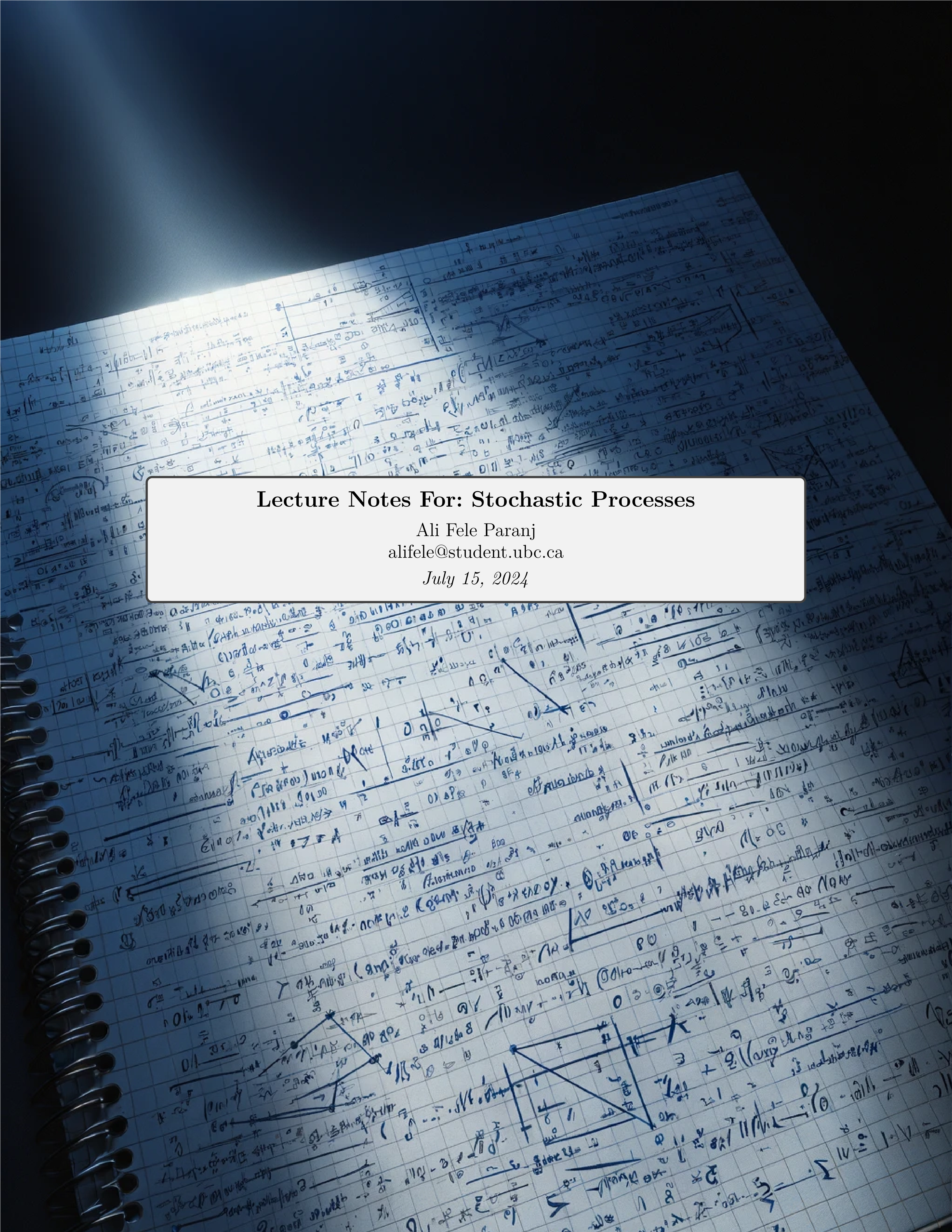


Lecture Notes For: Stochastic Processes

Ali Fele Paranj
alifele@student.ubc.ca

July 15, 2024

0. Contents

1. Probability Theory

1.1 Foundamentals

The main concept in the field of statistics and probability is the set theory. Basically all we deal with the sets. The whole theroy of statistics can be built on that. Let's discuss some foundamental concepts in statistics and then build the theory.

1.1.1 Random Experiment

To understand the meaning of random experiment, do not over think! The first thing that comes into our minds when we hear the word "random experiment" is its definition! In a nutshell, random experiment is an experiment that its outcome is unkown to us. Like:

- Tossing two coin
- Rolling a dice
- Measuring the number of possible ReadWrite operations on a piece of EEPROM chip

Do not overthink about that. Yes we can go further and discuss stuff like "we can compute the exact movement of dice or coin so it is not random but determenistic" and etc. Here I will not touch the philosophical topics that are very deep and do not necessarily converge the a unified point of view!

The random experiments can be modeled and despite the fact that a random experiment is random, we can deduce many useful information from modeling that. To model a random experiment, we use three important concepts: sample space, events, probability. In the following section, we will discuss each of them in detail.

1.1.2 Sample Space

Definition 1.1 — Sample Space. Sample space Ω is simply a set that contains *all possible outcomes* of a random experiment/

For each of random experiments described above, we can define a sample space. For example:

- Ω of Tossing Tow Coins:
$$\Omega = \{HH, HT, TH, TT\}$$
- Ω of Rolling a Dice:
$$\Omega = \{1, 2, 3, 4, 5, 6\}$$
- Ω of Rolling Two Dices:
$$\Omega = \{(1,1), (1,2), \ldots, (1,6), \ldots, (6,6)\}$$
- Ω of Number of possible ReadWrite operations on a EEPROM chip:
$$\Omega = \mathbb{N}$$

1.1.3 Events

Definition 1.2 — Events. Event E is a set of outcomes of a random experiment and is the subset of sample space Ω.
$$E \in \Omega$$

For example for any of the sample spaces specified above, we can define so many possible events. In fact any set that is a subset of the sample space is a valid event of that sample space. For example:

- Tossing Three Coins
  - There are at least on Heads:
$$E = \{HHH, HHT, HTH, THH, HTT, THT, TTH\}$$
  - There are only two Tails:
$$E = \{TTH, THT, HTT\}$$
- Rolling Two Dices
  - The sum of two dices is 4:
$$E = \{(1,3), (2,2), (3,1)\}$$
  - there are at least one prime number in the outcome:
$$E = \{(1,2), (1,3), (1,5), (2,1), (3,1), (5,1), (2,2), (2,3), (2,5), \ldots, (5,5)\}$$

Since we have define everything on the basics of set theory, then now we can correspond the everyday concepts to specific operations in the set theory.

■ **Example 1.1** The Mapping Between Everyday Language and Sets in the Theory of Probability

- At least one of two events $A, B \in \Omega$ happens: $E = A \cup B$.
- Tow events $A, B \in \Omega$ occures at the same time: $E = A \cap B$.
- Event $A \in \Omega$ does not happen: $E = \overline{A} = \Omega - A$.
- The event A happens but B does not happen: $E = A - B$.

■

In probability and statistics, we are dealing with three important concepts: sample space Ω, event E, and probability P.

Definition 1.3 — Disjoint events. If two events has no common elements (i.e. $A \cap B = \varnothing$) then we say that two events are *disjoint*. Basically, if two sets in the venn diagram has nothing is common they are considerent to be disjoint sets.

For example for the random experiment of tossing two coins, the events 1) both coins are heads: $A = \{HH\}$ and 2) both coins are tails: $B = \{TT\}$. Two events A, B are two disjoint events. **Two events being disjoing is NOT the same as being independent**. We will talk about independet events in future.

Note that since the events are basically sets, we can use theorems of set theory to solve the problems.

Theorem 1.1 — De Morgan's Laws. If A, B are two sets then:

$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$

Proof. the proof is left as an excerise! □

1.1.4 Probability

The last foundamental ingeridient in modeling a random experiment, is to define a probability for each event. The probability should intuitively reflect how likely an event is probable to happen. This probability should satisfy some foundamental properties which are explained as follows.

Definition 1.4 — Axioms of probability (Kolmogorov axioms). Suppose that $A, B \in \Omega$ is an event and $\mathbb{P}$ is a probability function. Then $\mathbb{P}$ should satisfy the following properites:

1. $0 \leq \mathbb{P}(A) \leq 1$
2. $\mathbb{P}(\Omega) = 1$
3. For the events $E_1, E_2, ..., E_n \in \Omega$ that are mutually exclusive (i.e. disjoint events):

$$\mathbb{P}(\bigcup_i E_i) = \sum_i \mathbb{P}(E_i)$$

These axioms are called the foundamental axioms of probability and also the Kolmogorov axioms. We are free to define any kind of probabilty function that we want but it is important that 1) It should align with our common sense, 2) It should satisfy the Kolmogorov axioms.

Using the axioms above, we can observe and prove several interesting properties of the probability function. In the following box we have expressed some of them.

Theorem 1.2 — Basic Properties of the Probability Function. Suppose that $\mathbb{P}$ is a probability function and $A, B \in \Omega$ are events of the sample space Ω. We can show that the probability function has the following properties:

1. $\mathbb{P}(\varnothing) = 0$
2. If $A \subset B$ then $\mathbb{P}(A) \leq \mathbb{P}(B)$.
3. $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$.

4. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

Proof. The properties can be proved using the basic set theory theorems.

1. Since $\varnothing$ is the complement of Ω, so these two sets are disjoint (i.e. $\varnothing \cap \Omega = \varnothing$). On the other hand from the set theory we know that $\varnothing \cup \Omega = \Omega$. So $\mathbb{P}(\varnothing \cup \Omega) = \mathbb{P}(\Omega)$. On the other hand, using the third axiom we can write: $\mathbb{P}(\varnothing \cup \Omega) = \mathbb{P}(\varnothing) + \mathbb{P}(\Omega)$. Comparing the two recent equations we can conclude that $\mathbb{P}(\varnothing) = 0$.

The proofs for 2,3,4 are left as a exercise. However, the solutions can be found in the book "Statistical Modeling and Computation by Kroese" chapter 1.

□

■ **Example 1.2 — Defining a simple probability function.** Let's define a probability function for the rolling n dice experiment that is both aligned with our common sense and also satisfy the Kolmogorov equations. Suppose that the Ω is the sample space and $E \in \Omega$ is an event. Then let's define:

$$\mathbb{P}(E) = \frac{|E|}{|\Omega|}$$

in which the $|E|$ means the cardinality (number of elements) of the set E.

■

Utilizing the properties of the probability function, we can derive some very important notions, one of which is reflected in the following proposition.

Proposition 1.1 — Conditional expansion - Law of total probabilities. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\mathfrak{F}$ be a finite collection of events $\mathfrak{F} = \{F_1, F_2, \cdots, F_n\}$ that partitions Ω. I.e.

(i) $F_i \cap F_j = \varnothing \qquad i \neq j$,

(ii) $\bigcap_i F_i = \Omega$.

Let $E \in \mathcal{F}$ be any nonempty event. Then we can write

$$\mathbb{P}(E) = \sum_i \mathbb{P}(E|F_i)\mathbb{P}(F_i).$$

Proof. Since $\mathfrak{F}$ partitions Ω and $E \neq \varnothing$, then $\{E \cap F_i\}_i$ is a partition of E. Thus

$$\mathbb{P}(E) = \mathbb{P}(\bigcup_i (E \cap F_i)) = \sum_i \mathbb{P}(E \cap F_i) = \sum_i \mathbb{P}(E|F_i)\mathbb{P}(F_i).$$

This completes the proof. □

In dealing with random variables, either continuous or discrete, using the notion of the law of total probabilities helps us to simplify some of the calculations significantly. The following examples are some places that we use this idea to simplify calculations by a lot.

■ **Example 1.3** Let $X_1, X_2, X_3, \cdots$ be i.i.d. real-valued random variable. Let T be a positive integer valued random variable. Define the the real-valued random variable N as

$$N = \sum_{i=1}^{T} X_i.$$

What is the probability generating function for N.

Solution For the generating probability function we have

$$G_N(s) = \mathbb{E}\Big[s^N\Big] = \mathbb{E}\Big[s^{X_1+X_2+\cdots+X_T}\Big].$$

The problem in evaluating the expression above is that the number of random variables X_i to be summed up is also a random variable. So the first step is to make this a non-random variable buy conditional expansion.

$$G_N(s) = \mathbb{E}\Big[s^N\Big] = \sum_{i\in\mathbb{N}} \mathbb{E}\Big[s^{X_1+\cdots+X_T}\Big|T=i\Big]\mathbb{P}(T=i) = \sum_{i\in\mathbb{N}} \mathbb{E}\Big[s^{X_1+\cdots+X_i}\Big]\mathbb{P}(T=i)$$

Since X_i are all i.i.d., then we can write

$$G_N(s) = \sum_{i\in\mathbb{N}} (\mathbb{E}\Big[s^{X_1}\Big])^n \mathbb{P}(T=i) = G_T(G_{X_1}(s)).$$

■

■ **Example 1.4** Let X, Y be two independent random variables. Define $Z = X+Y$. Find the PDF of Z.

Solution First, We need to find $F_Z(z) = \mathbb{P}(Z<z)$. For this we can write

$$F_Z(z) = \mathbb{P}(X+Y<z)$$

Again, we can use conditional expansion to write

$$F_Z(z) = \int_{\mathbb{R}} \mathbb{P}(X+Y<z\Big|Y=y)f_Y(y)\ dy = \int_{\mathbb{R}} \mathbb{P}(X<z-y)f_Y(y)\ dy = \int_{\mathbb{R}} F_X(z-y)f_Y(y)\ dy.$$

Then differentiating F_z with respect to z we will get the PDF

$$f_Z(z) = \frac{d}{dz}F_Z(z) = \int_{\mathbb{R}} f_X(z-y)f_Y(y)\ dy = (f_X * f_Y)(z).$$

■

■ **Example 1.5** Let X, Y be two real valued random variable, not necessarily independent. Calculate $\mathbb{P}(X<Y)$.

Solution To calculate this we can again use the law of total probabilities. In particular

$$\mathbb{P}(X<Y) = \int_{\mathbb{R}} \mathbb{P}(X<y)f_Y(y)\ dy = \int_{\mathbb{R}} F_{X|Y}(y)f_Y(y)\ dy.$$

■

1.1.5 Isomorphism between random experiments

Often, there is this intuition that certain random experiments are really the same, although they might look very different from each other. For instance, consider two random experiments. In one, we are playing a dice successively and asking what is the probability that after 5 plays, 1 is not appeared. The second experiment is that we have 6 Urns and we place balls in them successively, i.e. at each step one ball is placed in one of the urns and the chance of a ball to end up in any of the urns in equal. These two experiment, although very different, but looks very similar. There is one way that we can formalize this wage intuition, and that is the notion of isomorphism between sets. We say two sets are isomorphic if there is a bijection between them. And the reason that the previously mentioned experiments feel the same is that the sample space Ω of these two experiments are in fact isomorphic.

1.2 Random Variables

Often, we are interested in the some measurements of the outcome of a random experiment rather than knowing the outcome it self. For instance, if the experiment of tossing two dice, we might be interested in asking the question if the sum of two dice is 6, and not concerned over whether the actual outcome was (3,3) or (2,4), etc. These quantities of interest are called random variables. The following definition put this into a more formal definition.

Definition 1.5 Let $(\Omega, E, \mathbb{P})$ be a probability space. Then a random variable X is a function $X : \Omega \to S$, where S called the state space.

■ **Remark** The state space S must have some properties, i.e. being measurable, etc. You can read more about this on the Wikipedia of random variables. Also, the state space S if often $\mathbb{R}$, or in the case of a discrete time Markov chain, S is a finite set (that can be the edge set of a graph).

Since the value of a random variable is determined by the outcomes of the random experiment, we can assign probabilities to the possible values of the random variable. We use the following notation for this purpose.

Definition 1.6 — Notation for probability of random variables. Let X be a random variable. Then we define event

$$E = \{X = a\} = \{\omega \in \Omega : X(\omega) = a\}.$$

Then the following notations are usually used interchangeably:

$$\mathbb{P}(X = a) = \mathbb{P}(\{X = a\})$$

both of which is simply $\mathbb{P}(E)$.

■ **Example 1.6** Let X be a random variable defined to be the sum of two fair dice. Then

$$\begin{aligned}
\mathbb{P}(\{X = 2\}) &= \mathbb{P}(\{(1, 1)\}) = \frac{1}{36}, \\
\mathbb{P}(\{X = 3\}) &= \mathbb{P}(\{(1, 2), (2, 1)\}) = \frac{2}{36}, \\
\mathbb{P}(\{X = 13\}) &= \mathbb{P}(\varnothing) = 0.
\end{aligned}$$

■

■ **Example 1.7** Suppose that we toss a coin having probability p of coming up heads. We continue tossing the coin until we see a heads. Let the random variable N be the number of times we toss the coin. Describe this random variable.

Solution Although, we can always solve this kind of questions in an ad hoc way by just simply following our intuition, but it is always a best practice to try to fine tune our abstract thinking with our intuitive understandings in these kind of example. Then we can use of abstract thinking capability to solve problems that are almost impossible to address by solely depending on the intuition. So, it is a good idea to try to see how does the set Ω look like. The set Ω will be the set of all finite string of all T letters terminated with H. In other words

$$\Omega = \{H, TH, TTH, TTTH, TTTTH, \cdots\}.$$

Then the random variable $N : \Omega \to \mathbb{Z}$ is basically the length of the string. For instance, if $\omega = TTH \in \Omega$, then $N(\omega) = 3$. Let's calculate

$$\mathbb{P}(N = 3) = \mathbb{P}(\{\omega \in \Omega : N(\omega) = 3\}).$$

To solve this, we need to define appropriate events and then condition our probability on those events. Define F_n be the event where the n first outcomes are tails. For instance

$$F_1 = \{TH, TTH, TTTH, \cdots\},\ F_2 = \{TTH, TTTH, TTTTH, \cdots\},\ \cdots .$$

And let $E = \{N = 3\} = \{TTH\}$. Then we can condition $\mathbb{P}(E)$ on F_2

$$\mathbb{P}(E) = \mathbb{P}(E|F_2)\mathbb{P}(F_2) + \mathbb{P}(E|F_2^c)\mathbb{P}(F_2^c).$$

Note that $F_2^c = \{H, TH\}$, this $\mathbb{P}(E|F_2^c) = \mathbb{P}(E \cap F_2^c)/\mathbb{P}(F_2^c) = 0$. Now we need to determine $\mathbb{P}(F_2)$. Again, we can condition this event on F_1. Then we can write

$$\mathbb{P}(F_2) = \mathbb{P}(F_2|F_1)\mathbb{P}(F_1) + \mathbb{P}(F_2|F_1^c)\mathbb{P}(F_1^c).$$

with the same argument as above $\mathbb{P}(F_2|F_1^c) = 0$. Combining these equations we will get

$$\mathbb{P}(E) = \mathbb{P}(E|F_2)\mathbb{P}(F_2|F_1)\mathbb{P}(F_1).$$

Now these probabilities are easy to calculate which leads to the final answer

$$\mathbb{P}(E) = (1 - p)(1 - p)p.$$

And by induction we can conclude

$$\mathbb{P}(\{N = n\}) = (1 - p)^n p.$$

■

■ **Example 1.8** Suppose that independent trials, each of which results in m possible outcomes with respective probabilities $p_1, p_2, \ldots, p_m$ such that $\sum_{i=1}^{m} p_i = 1$. Are continually performed. Let X be the number of trials needed until each outcome has occurred at least once. Describe the properties of this random variable.

Solution It is sometime a good idea to try to imagine what does the sample space look like. Let $\Sigma = \{s_1, s_2, s_3, \ldots, s_m\}$ be a set of m distinct symbols. Then each time we are continually performing the experiment, we are getting each of these symbols with corresponding probability p_m. Thus the sample space will be the set of all infinite sequences of these symbols. In other words

$$\Omega = \{\text{all infinite sequence of symbols from } \Sigma\}.$$

Then the random number $X(\omega)$ for $\omega \in \Omega$ is basically the length of the prefix string of ω in which any of the symbols in Σ has been occurred at least once.

■

1.2.1 Cumulative Distribution of Random Variable

The notion of the cumulative distribution of a random variable comes handy in most of the future calculations. Also, this distribution can be used to derive other notions of distributions what are extremely important in applications.

Definition 1.7 — Cumulative distribution. Let X be a random variable $X : \Omega \to \mathbb{R}$. Then the cumulative distrubition $F : \mathbb{R} \to \mathbb{R}$ is defined as

$$F(x) = \mathbb{P}(\{X \leq x\}).$$

Proposition 1.2 The cumulative distribution of a random variable has the following properties.

(i) $\mathbb{P}(a < X \leq b) = F(b) - F(a)$.

(ii) $F(x)$ is a non-decreasing function of x.

Proof. (i)

$$\mathbb{P}(\{a < X \leq b\}) = \mathbb{P}(\{X \leq b\} \cap \{X \leq a\}^c) = -\mathbb{P}(\Omega) + \mathbb{P}(\{X \leq b\}) + \underbrace{\mathbb{P}(\{X \leq a\}^c)}_{1-\mathbb{P}(\{X \leq a\})} = F(b) - F(a).$$

(ii) Let $b_1, b_2 \in \mathbb{R}$ and $b_1 \leq b_2$. Then $\{X \leq b_1\} \subseteq \{X \leq b_2\}$. This implies

$$\mathbb{P}(\{X \leq b_1\}) \leq \mathbb{P}(\{X \leq b_2\}) \implies F(b_1) \leq F(b_2).$$

This implies that $F(x)$ is a non-decreasing function.

□

1.3 Probability Generating Function

In this section we will go through the details of the probability generating function. We start with the following definition.

Definition 1.8 — Probability Generating Function. Let X be a random variable with state space $S = \mathbb{Z}_+$. Then the probability generating function for this random variable is a function define as

$$G_X(s) = \mathbb{E}\left[s^X\right] = \sum_{x \in S} s^x \mathbb{P}(X = x).$$

In different areas of mathematics, we often can define something algebraic that is very easy to handle (like differentiation, etc) and carries the important information of the object under study. One of these algebraic symbolic objects is the Tutte polynomial, Chromatic polynomial, matching polynomial, etc. These polynomials are kind of modeling the object under study with tools that are easy to handle. The probability generating function is one of those symbolic objects. Because of the way that is crafted, it carries most of the information about the random variable, while the actual object as a function might have poor properties. This will be more clear in the following proposition. In a nutshell, the probability generating function is more of a symbolic thing rather than actual function with meaning full properties. That is why we generally evaluate this function (and its derivatives) at point 0 or 1.

Proposition 1.3 — Properties of the probability generating function. Let X be a random variable, and $G_X(s)$ its probability generating function. Then we have

(i) $G_X(1) = 1$.

(ii) $\mathbb{E}\left[X\right] = G'_X(1)$.

(iii) $\delta X = G''_X(1) - G'_X(1)^2 + G'_X(1)$

(iv) Let X, Y be independent random variables. Then we have

$$G_{X+Y}(s) = G_X(s) G_Y(s).$$

(v) Let $X_1, X_2, \cdots$ be iid random variables, and N be a random variable taking values in $\mathbb{Z}_+$. Define $T = X_1 + X_2 + \cdots + X_N$. Then we have

$$G_T(s) = (G_N \circ G_{X_1})(s).$$

Proof. The proof for part i,ii, iii, and iv basically follows immediately from the definition. So we will only provide the proof for part iv.
T is the sum of N iid random variables where N is itself a random variable. We can make it a normal variable by using the law of total probabilities.

$$G_T(s) = G_{\sum_i^N X_i}(s) = \sum_{n=0}^{\infty} G_{\sum_i^n X_i}(s)\mathbb{P}(N = n) = \sum_{n=0}^{\infty}(G_{X_1})^n \mathbb{P}(N = n) = G_N(G_{X_1})(s)$$

and this completes the proof. □

The item (iv) in the proposition above is very important, as it makes the hard calculations easy to do. See the following example for more details.

■ **Example 1.9** We select a number N from $\{1, 2, 3, \cdots, 100\}$ randomly and then generate N random numbers $X_1, X_2, \cdots X_N$ from the distribution Unif$[0, 1]$. Then we compute $T = X_1 + X_2 + \cdots + X_N$. What is the average of T?

Solution We know that

$$\mathbb{E}\Big[T\Big] = G'_T(1).$$

Thus we need to calculate the probability generating function $G_T(s)$. From part (iv) of the proposition above we know that $G_T = G_N \circ G_{X_1}$. Thus we will have

$$G'_T = G'_{X_1} G'_N \circ G_{X_1}.$$

Thus evaluating at $s = 1$ we will have

$$G'_T(1) = G'_{X_1}(1) G'_N(\underbrace{G_{X_1}(1)}_{1}) = \mathbb{E}\Big[X_1\Big]\mathbb{E}\Big[N\Big].$$

On the other hand we have $\mathbb{E}\Big[N\Big] = 50$ and $\mathbb{E}\Big[X_1\Big] = 1/2$. Then

$$\mathbb{E}\Big[T\Big] = 25.$$

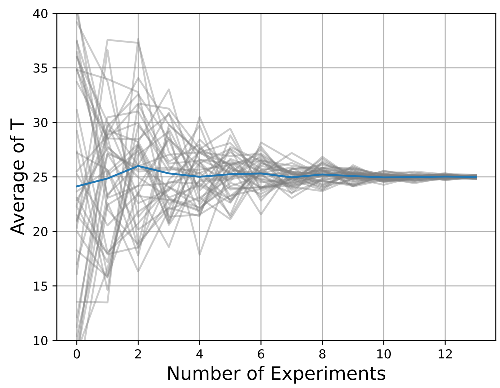
The following figure shows this fact (i.e. convergence of the average value of T to 25 when we increase the number of experiments.)

■

1.4 Solved Problems

■ **Problem 1.1 — From Ross.** Ben can talk a course in compute science or chemistry. If she takes the computer science course, then she will get A grade with probability $\frac{1}{2}$. If she takes the chemistry course, then she will get A grade with probability $\frac{1}{3}$. She decides to base her decision on the flip of a fair coin. What is the probability that she gets an A in chemistry?

Solution We define the following events



A: she will get an A grade.
CO: she will take the computer science course.
CH: she will take the chemistry course.

Then the question is basically asking for $\mathbb{P}(A \cap CH)$. We can compute it by

$$\mathbb{P}(A \cap CH) = \mathbb{P}(A|CH)\mathbb{P}(CH) = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}.$$

■ **Problem 1.2** And urn contains seven black balls and five white balls. We draw two times from the urn. Given that the each ball has the same probability to be drawn, what is the probability that both balls drawn are black?

Solution This question nicely demonstrates the fact that there are many ways to define the event spaces, and not all of them are very useful in computing the desired probability. Define

E: two drawn balls are black.

The question is in fact asking $\mathbb{P}(E)$. But this even is not very useful in any progress with the solution. Thus we need to define some finer events

E_1: The first drawn ball is black.
E_2: The second drawn ball is black.

It is clear that $E = E_1 \cap E_2$. These two finer events allows us to compute the probability of interest given the data we have in our hand.

$$\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_2|E_1)\mathbb{P}(E_1) = \frac{6}{11} \cdot \frac{7}{12}$$

■ **Problem 1.3 — From Ross.** Three men at a party through their hats into the center of the room, and then, after mixing the hats, each pick a hat randomly. What is the probability if non of them get their own hat back.

Solution There are a million ways to tack a probability problem. We can construct a suitable sample space and then compute the probabilities explicitly, or we can use the properties of the probability function to computer the desired probability without any need to construct the sample space. Here, we will demonstrate two ways.

Solving the problem by utilizing the properties of the probability function. First we need to define some suitable events. There are again many ways to define event sets and each have their own pros and cons. We proceed with the following definition.

E_i: The person i "selects" his own hat.

Also, with this particular construction of the event sets, it is much more easier to compute the complementary probability of the desired probability first and then compute the desired one by simply subtracting it from 1. The complement of the event "no men gets his own hat back" is "at least one man gets his hat back" which is $\mathbb{P}(E_1 \cup E_2 \cup E_3)$. To compute the terms of this we first need to calculate $\mathbb{P}(E_i)$, $\mathbb{P}(E_i \cap E_j)$ where $i \neq j$ and also $\mathbb{P}(E_1 \cap E_2 \cap E_3)$. We know that $\mathbb{P}(E_i) = 1/3$ for $i = 1, 2, 3$. That is because it is equally likely he selects any of the hats at the center. For $\mathbb{P}(E_i \cap E_j)$ we can write

$$\mathbb{P}(E_i \cap E_j) = \mathbb{P}(E_i | E_j)\mathbb{P}(E_j) = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}.$$

In which we used the fact that $\mathbb{P}(E_i|E_j)$ is $\frac{1}{2}$ for distinct i, j. That is because given person j selects his hat correctly, then there are two possibilities for E_i to select his hat (he can pick the correct one or the wrong one). Lastly for $\mathbb{P}(E_1 \cap E_2 \cap E_3)$ we write

$$\mathbb{P}(E_1 \cap E_2 \cap E_3) = \mathbb{P}(E_1 | E_2 \cap E_3)\mathbb{P}(E_2 \cap E_3) = \mathbb{P}(E_1 | E_2 \cap E_3)\mathbb{P}(E_2|E_3)\mathbb{P}(E_3) = 1 \cdot \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}.$$

Thus

$$\mathbb{P}(E_1 \cup E_2 \cup E_3) = (1) - (1/2) + (1/6) = \frac{4}{6}.$$

Then the probability of interest will be

$$\mathbb{P}(E) = 1 - \frac{4}{6} = \frac{1}{3}.$$

Solving by constructing a sample space. A suitable sample space for this problem can be the set of all permutations on three letters. This set is

$$\Omega = \{\begin{pmatrix} a & b & c \\ \boxed{a} & \boxed{b} & \boxed{c} \end{pmatrix}, \begin{pmatrix} a & b & c \\ \boxed{a} & c & b \end{pmatrix}, \begin{pmatrix} a & b & c \\ b & a & \boxed{c} \end{pmatrix}, \begin{pmatrix} a & b & c \\ b & c & a \end{pmatrix}, \begin{pmatrix} a & b & c \\ c & a & b \end{pmatrix}, \begin{pmatrix} a & b & c \\ c & \boxed{b} & a \end{pmatrix}\}.$$

Note that the elements in the box represents the fixed point of the permutation. The probability of interest is basically the number of permutations that has no fixed point. As it is clear from the set Ω, the probability is

$$\mathbb{P}(E) = \frac{2}{6} = \frac{1}{3}.$$

■ **Problem 1.4 — Conditional probability mass function (from Ross).** Let X, Y be two random variables with the joint probability mass function given as

$$P(1,1) = 0.5 \qquad P(1,2) = 0.1, \qquad P(2,1) = 0.1, \qquad P(2,2) = 0.3.$$

Calculate the conditional probability mass function of X given that $Y = 1$.

Solution We will use the following identity

$$P_{X|Y}(x|y) = \mathbb{P}(X = x | Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}.$$

Observe that

$$\mathbb{P}(Y = y) = \sum_x \mathbb{P}(Y = y, X = x)$$

thus $\mathbb{P}(Y = 1) = 0.5 + 0.1 = 0.6$. So we will have

$$P_{X|Y}(1|1) = \frac{0.5}{0.6} = \frac{5}{6}, \qquad P_{X|Y}(2|1) = \frac{0.1}{0.6} = \frac{1}{6}.$$

■ **Problem 1.5 — Conditional probability mass function for geometric random variables (from Ross).** Let X_1, X_2 be two independent random variables with geometric distributions with parameters (n_1, p) and (n_2, p). Calculate the conditional probability mass function of X_1 given that $X_1 + X_2 = m$.

Solution First, observe that $Y = X_1 + X_2$ is a binomial distribution with parameter $(n_1 + n_2, p)$. Thus we can write

$$P_{X_1|Y}(k|m) = \mathbb{P}(X_1 = k|Y = m) = \frac{\mathbb{P}(X_1 = k, X_1 + X_2 = m)}{\mathbb{P}(X_1 + X_2 = m)} = \frac{\mathbb{P}(X_1 = k, X_2 = m - k)}{\mathbb{P}(Y = m)}$$

Since the random variables X_1 and X_2 are independent, we can write

$$P_{X_1|Y}(k|m) = \frac{\mathbb{P}(X_1 = k)\mathbb{P}(X_2 = m - k)}{\mathbb{P}(Y = m)} = \frac{\binom{n_1}{k}\binom{n_2}{m-k}}{\binom{n_1+n_2}{m}}.$$

■ **Problem 1.6 — Conditional probability mass function for Poisson random variables (from Ross).** Let X, Y be two independent Poisson random variables with parameters λ_1 and λ_2 respectively. Calculate the conditional probability mass function for X given that $X_1 + X_2 = n$.

Solution First observe that $Z = X + Y$ is a Poisson random variable with parameter $\lambda_1 + \lambda_2$. Thus we will have

$$P_{X|X+Y}(m|n) = \frac{\mathbb{P}(X = m|X + Y = n)}{\mathbb{P}(X + Y = n)} = \frac{\mathbb{P}(X = m, Y = n - m)}{\mathbb{P}(X + Y = n)}$$

Given that X, Y are independent random variables then we can write

$$P_{X|X+Y}(m, n) = \frac{\mathbb{P}(X = m)\mathbb{P}(Y = n - m)}{\mathbb{P}(X + Y = n)} = \frac{\lambda_1^n \lambda_2^{n-m} n!}{m!(n-m)!(\lambda_1 + \lambda_2)^n} = \binom{n}{m}(\frac{\lambda_1}{\lambda_1 + \lambda_2})^m(\frac{\lambda_2}{\lambda_1 + \lambda_2})^{n-m}$$

Thus the conditional probability mass function of X given that $X + Y = n$ will be a binomial random variable with parameter $(n, \lambda_1/(\lambda_1 + \lambda_2))$. We can now easily compute the conditional expectation value as

$$\mathbb{E}\Big[X|X + Y = n\Big] = \frac{n\lambda_1}{\lambda_1 + \lambda_2}$$

■ **Problem 1.7** Let X, Y be two discrete random variables. Prove that

$$\mathbb{E}\Big[\mathbb{E}\big[X|Y\big]\Big] = \mathbb{E}\big[X\big].$$

Solution We start with the definition of the expectation of a discrete random variable.

$$\begin{aligned}
\mathbb{E}\Big[\mathbb{E}\big[X|Y\big]\Big] &= \sum_y \mathbb{E}\big[X|Y = y\big]\mathbb{P}(Y = y) = \sum_y \sum_x x\mathbb{P}(X = x|Y = y)\mathbb{P}(Y = y) \\
&= \sum_{x,y} x\mathbb{P}(X = x, Y = y) = \sum_x x \sum_y \mathbb{P}(X = x, Y = y) = \sum_x x\mathbb{P}(X = x) = \mathbb{E}\big[X\big]
\end{aligned}$$

■ **Problem 1.8 — The expectation of a random number of random variables (from Ross).** Let the expected number of injures in an industrial field be 4 per week. Also, assume that the number of workers injured at each incidence are independent random variables with average 2. Then what is the expected number of injuries in one week?

Solution Let $X_1, X_2, \cdots$ be i.i.d random variables representing the number of workers injured at each incidence. We are interested in

$$\mathbb{E}\Big[X_1 + \cdots + X_N\Big]$$

where N is a random variable representing the number of incidences occurred in a week. By the law of conditional expectation we can write

$$\mathbb{E}\Big[X_1 + \cdots + X_N\Big] = \sum_n \mathbb{E}\Big[X_1 + \cdots + X_n\Big]\mathbb{P}(N = n) = \sum_n n\mathbb{E}\Big[X\Big]\mathbb{P}(N = n) = \mathbb{E}\Big[X\Big]\mathbb{E}\Big[N\Big].$$

Thus the average number of workers injured in a week will be 8.

■ **Problem 1.9 — An alternative way to compute the expectation of a geometric random variable.** Consider a coin with probability p to fall heads. What is the expectation value of the number of tosses required until we get the first head?

Solution Let $X_1, X_2, \cdots$ be Bernoulli random variables with parameter p. Let N be a random variable denoting the number of tosses required until we get the first heads. We can condition the expected value of E to the first outcome.

$$\mathbb{E}\Big[N\Big] = \mathbb{E}\Big[N|X_1 = H\Big]\underbrace{\mathbb{P}(X_1 = H)}_{=p} + \mathbb{E}\Big[N|X_1 = T\Big]\underbrace{\mathbb{P}(X_1 = T)}_{=1-p}$$

Observe that

$$\mathbb{E}\Big[N|X_1 = H\Big] = 1, \qquad \mathbb{E}\Big[N|X_1 = T\Big] = 1 + \mathbb{E}\Big[N\Big].$$

Thus we will have

$$\mathbb{E}\Big[N\Big] = \frac{1}{p}.$$

■ **Problem 1.10 — Trapped miner (from Ross).** A miner is trapped in the mine and has three doors in front of him. He is equally likely to choose any of the three. The first door will take him to safety after 2 hours of walking, the second door will take him to the mine again after 3 hours of walking, and the third door will take him to the mine again after 5 hours of walking. What is the expected time that the miner will arrive to safety?

Solution Let $X_1, X_2, \cdots$ be random variables denoting the doors that the miner choose at each time that he attempts to escape. Furthermore, let T be a random variable showing the the time it takes for the miner to escape. To calculate $\mathbb{E}\Big[T\Big]$ we can condition it on the first door choice. I.e.

$$\mathbb{E}\Big[T\Big] = \mathbb{E}\Big[T|X_1 = 1\Big]\mathbb{P}(X_1 = 1) + \mathbb{E}\Big[T|X_1 = 2\Big]\mathbb{P}(X_1 = 2) + \mathbb{E}\Big[T|X_1 = 3\Big]\mathbb{P}(X_1 = 3)$$

Observe that

$$\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = 2) = \mathbb{P}(X_1 = 3) = 1/3.$$

Also

$$\mathbb{E}\Big[T|X_1 = 1\Big] = 2, \qquad \mathbb{E}\Big[T|X_1 = 2\Big] = 3 + \mathbb{E}\Big[T\Big], \qquad \mathbb{E}\Big[T|X_1 = 3\Big] = 3 + \mathbb{E}\Big[T\Big].$$

Thus we will have

$$\mathbb{E}\Big[T\Big] = 10.$$

So on average it will take the miner to exit the mine in 10 hours. Note that this does not guarantee that the miner will eventually escape. It is possible that we will get in trap by repeatedly choosing the door number 3.

2. Basics and Definitions

3. Markov Chain

NOTE TO MYSELF: I prefer to develop whole theory of discrete Markov chains by defining the state space to be the set of symbols Σ (at most countable). This is beneficial, because then the sample space of any Markov chain will be the set of all infinite sequences (strings) from the symbols from Σ. At some point in the future, I might rewrite this chapter, working consistently with Σ as the state space.

We start with the definition of a Markov Chain.

Notation Let $(X_n)_{n\geq 0}$ be a Markov chain on the state space S, $x \in S$, and let E be an event. Then

$$\mathbb{P}_x(E) = \mathbb{P}(E|X_0 = x).$$

The following proposition will be one of our main tools throughout the chapter.

Proposition 3.1 — Conditional expansion. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\mathfrak{F}$ be a finite collection of events $\mathfrak{F} = \{F_1, F_2, \cdots, F_n\}$ that partitions Ω. I.e.

(i) $F_i \cap F_j = \varnothing \qquad i \neq j$,

(ii) $\bigcap_i F_i = \Omega$.

Let $E \in \mathcal{F}$ be any nonempty event. Then we can write

$$\mathbb{P}(E) = \sum_i \mathbb{P}(E|F_i)\mathbb{P}(F_i).$$

Proof. Since $\mathfrak{F}$ partitions Ω and $E \neq \varnothing$, then $\{E \cap F_i\}_i$ is a partition of E. Thus

$$\mathbb{P}(E) = \mathbb{P}(\bigcup_i (E \cap F_i)) = \sum_i \mathbb{P}(E \cap F_i) = \sum_i \mathbb{P}(E|F_i)\mathbb{P}(F_i).$$

This completes the proof. □

Proposition 3.2 — First step argument. Let $(X_n)_{n\geq 0}$ be a Markov chain on the state space S.

Let $x \in S$, and $W, Z \subset S$. Let B be any event. Then

$$\mathbb{P}_x(B) = \sum_{y:x\sim y} \mathbb{P}_y(B)P(x,y).$$

Proof. To prove the proposition above, we Let $E_i = \{X_0 = x, X_1 = y_i\}$ where $y_i \sim x$. So, in words, we say that the event E_i has occurred if $X_1 = y_i$. It is clear that $E_i \cap E_j = \varnothing$ where $i \neq j$. Thus $\bigcup_i (B \cap E_i) = B$. Thus

$$\mathbb{P}_x(B) = \sum_i \mathbb{P}_x(B \cap E_i) = \sum_i \mathbb{P}_x(B|E_i)\mathbb{P}_x(E_i).$$

In which $\mathbb{P}_x(E_i) = \mathbb{P}(E_i|X_0 = x) = \mathbb{P}(X_1 = y_i|X_0 = x) = P(x, y_i)$. Also

$$\mathbb{P}_x(B|E_i) = \mathbb{P}(B|X_1 = y_i, X_0 = x) = \mathbb{P}(B|X_1 = y_i) = \mathbb{P}_{y_i}(B),$$

in which we have used the Markov property. Thus we can write

$$P_x(B) = \sum_i \mathbb{P}_{y_i}(B)P(x, y_i).$$

□

3.1 Dissecting an Experiment

The idea of the Markov chain, random variables, probability spaces, etc. might be quite confusing when the setting of a particular random experiment becomes large. Here in this section, we are going to explain the details of a random experiment explicitly. The random experiment is the following

> Assume we have 6 urns, and we put ball at each urn successively. What is the probability that there will be exactly 3 non-empty urns after 9 balls have been distributed?

Sample Space

First, note that there are two things happening, that we can call experiments. First is that we are successively doing something, throwing dice and putting a ball inside the urn, and the second thing is that we can consider the whole thing to be a giant experiment by its own. Our convention, from now on, will be that we will call the whole thing as experiment, and we will consider each sub-experiment as sub-steps of the process. Intuitively, an experiment is something that we can repeat to observe different outcomes. It is true that the whole experiment is actually successive repetition of throwing dice, but we actually consider the largest meaningful setup to be our experiment and call the sub-experiment as the steps of the process. It is kind of confusing at first glance, but becomes more natural after a while. By the definition, the sample space of a random experiment, is the set of all possible outcomes. But what are the possible outcomes in our experiment? If we perform an experiment, then we can get an outcome like

$$134234521133456665342431255553214532145123453124563412314566654 35\dots.$$

This outcome basically is saying that the outcome of the first dice throw was 1, the second dice throw was 3, the third was 4, and etc. If we repeat the experiment, we will get other outcomes.

So the sample space is the set of all sequence of numbers consisting of $\{1, 2, 3, 4, 5, 6, 7, 8\}$. Thus, we can write

$$\Omega = \{13424\ldots, 54321\ldots, 65432\ldots, 12345\ldots, \cdots\}.$$

So, each outcome, consists of sequence of random variables $\{Y_t\}$, that is defined on the sample space $\Omega_{\text{dice}} = [1, 2, 3, 4, 5, 6]$. That is Y_{10} means the random variable associated with the dice throwing experiment at the time t of our main experiment. These random variables are all independent and identically distributed (i.e. they are iid).

Markov Chain

Let X_t be a random variable defined on Ω that represents the number of full urns at time t. Let $\omega = 23123456234213245342 3\ldots$ and $t = 10$. Then $X_1 0(\Omega) = 6$, since by the time t, we have put at least one ball at each urn and all of the urns are full. Since this is a Markov chain with state space $S = \{0, 1, 2, 3, 4, 5, 6\}$, we can draw a transition diagram, and analyze the question more carefully. This is what we have done in Example 3.3.

3.2 Solved Problems

■ **Example 3.1** An urn always contains 2 balls. Ball colors are red and blue. At each stage a ball is randomly chosen and then replaced by a new ball, which with probability 0.8 is the same color, and with probability 0.2 is the opposite color, as the ball it replaces. If initially both balls are red, find the probability that the fifth ball selected is red. [This question is from Ross]

Solution First, we need to translate this problem to a suitable Markov chain. There are many ways we can do so, each with its own pros and cons. The difference between all of these formulations come down to our choice for the state space (i.e. the co-domain of the random variable). For instance, we can assume that the state space is $S = \{RR, RB, BB\}$ that is the content of the Urn, or we can simply say that the state space is $S = \{0, 1, 2\}$ that is the number of red ball inside the Urn. Since these two sets are isomorphic (as there is a bijection between these two sets), but the actual choice depends on personal preference. Let's proceed with $S = \{0, 1, 2\}$. Then, we need to determine the transition matrix. We can do so by doing the first step argument. We start with $P(0, 0)$.

$$P(0,0) = \mathbb{P}(X_1 = 0|X_0 = 0) = \mathbb{P}(X_1 = 0|X_0 = 0, E_R)\underbrace{\mathbb{P}(E_R|X_0 = 0)}_{0} + \underbrace{\mathbb{P}(X_1 = 0|X_0 = 0, E_B)}_{0.8}\underbrace{\mathbb{P}(E_B|X_0 = 0)}_{1},$$

where E_R is the event at which a red balls is drawn from the Urn, while E_B is the event where a blue ball is drawn. The reason behind the values for the term above are very straight forward. For instance $\mathbb{P}(E_R|X_0 = 0) = 0$ because given the fact that number of red balls in the Urn is zero ($X_0 = 0$), then the probability that we draw a red ball is zero (as there is not red balls in the Urn). For the term $\mathbb{P}(X_1 = 0|X_0 = 0, E_B) = 0.8$, because given there is no red balls inside the urn, and also given the fact that the drawn ball is blue, the probability of ending up at the state $X_1 = 0$ (i.e. still no red balls) is that probability is that we replaced the drawn ball with a blue ball (same color) which has the probability 0.8. Similarly, we can calculate the first step transition probabilities.

$$P(0,1) = \mathbb{P}(X_1 = 1|X_0 = 0) = \mathbb{P}_0(X_1 = 1) = \mathbb{P}_0(X_1 = 1|E_R)\underbrace{\mathbb{P}_0(E_R)}_{0} + \underbrace{\mathbb{P}_0(X_1 = 1|E_B)}_{0.2}\underbrace{\mathbb{P}_0(E_B)}_{1} = 0.2,$$

$$P(0,2) = \mathbb{P}(X_1 = 2|X_0 = 0) = \mathbb{P}_0(X_1 = 2) = \mathbb{P}_0(X_1 = 2|E_R)\underbrace{\mathbb{P}_0(E_R)}_{0} + \underbrace{\mathbb{P}_0(X_1 = 2|E_B)}_{0}\underbrace{\mathbb{P}_0(E_B)}_{1} = 0,$$

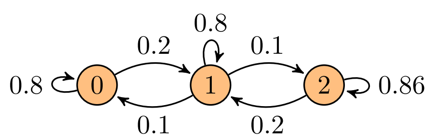
$$P(1,0) = \mathbb{P}(X_1 = 0|X_0 = 1) = \mathbb{P}_1(X_1 = 0) = \underbrace{\mathbb{P}_1(X_1 = 0|E_R)}_{0.2}\underbrace{\mathbb{P}_1(E_R)}_{0.5} + \underbrace{\mathbb{P}_1(X_1 = 0|E_B)}_{0}\underbrace{\mathbb{P}_1(E_B)}_{0.5} = 0.1.$$

$$P(1,1) = \mathbb{P}(X_1 = 1|X_0 = 1) = \mathbb{P}_1(X_1 = 1) = \underbrace{\mathbb{P}_1(X_1 = 1|E_R)}_{0.8}\underbrace{\mathbb{P}_1(E_R)}_{0.5} + \underbrace{\mathbb{P}_1(X_1 = 1|E_B)}_{0.8}\underbrace{\mathbb{P}_1(E_B)}_{0.5} = 0.8.$$

and so on. Then we will have the following transition matrix for this problem.

$$M = \begin{pmatrix} 0.8 & 0.2 & 0 \\ 0.1 & 0.8 & 0.1 \\ 0 & 0.2 & 0.8 \end{pmatrix}$$

with the following graph



Now, we need to compute the probability that the fifth ball drawn is red. This means that we have already drawn four balls, and now we want to draw the fifth one. So, we need to consider the 4 step transition matrix, i.e. M^4. Then

$$M^4 = \begin{pmatrix} 0.4872 & 0.4352 & 0.0776 \\ 0.2176 & 0.5648 & 0.2176 \\ 0.0776 & 0.4352 & 0.4872 \end{pmatrix}$$

Given that we have started with 2 red balls, then the probability of finding the Urn with 0 red balls is 0.0776, with 1 red ball is 0.4352, and with 2 red balls is 0.4872. So the probability that the next drawn balls is red is

$$\mathbb{P}(E_R) = \underbrace{\mathbb{P}(E_R|X_4 = 0)}_{0}\underbrace{\mathbb{P}(X_4 = 0)}_{0.0776} + \underbrace{\mathbb{P}(E_R|X_4 = 1)}_{0.5}\underbrace{\mathbb{P}(X_4 = 1)}_{0.4352} + \underbrace{\mathbb{P}(E_R|X_4 = 2)}_{1}\underbrace{\mathbb{P}(X_4 = 2)}_{0.4872} = 0.7048.$$

■

■ **Example 3.2 — Turning non-Markov processes to Markov-chain.** Suppose that whether or not it rains today depends on previous weather conditions through the last two days. Specifically, suppose that if it has rained for the past two days, then it will rain tomorrow with probability 0.7; if it rained today but not yesterday, then it will rain tomorrow with probability 0.5; if it rained yesterday but not today, then it will rain tomorrow with probability 0.4; if it has not rained in the past two days, then it will rain tomorrow with probability 0.2. [This question is from Ross]. Given that it rained on Monday and Tuesday, what is the probability that it will rain on Thursday?

Solution This random process is not a Markov chain, the value of the random variable at the next state, depends on two previous states. However, we can turn this into a Markov chain. Define the following states

RR: Rained yesterday and today.
$R\overline{R}$: Rained yesterday, but not today.
$\overline{R}R$: Not rained yesterday, but rained today.
$\overline{RR}$: Not rained yesterday and today.

Suppose that we are at state RR. Suppose that it rained yesterday and also today. Thus we are at state RR. If it rains tomorrow, then we will be still at state RR. That is because, That is because the yesterday of tomorrow is today! So if it rains tomorrow, since today (yesterday of tomorrow) was also rainy, thus if it rains tomorrow then we will stay at state RR. If it does not rain tomorrow, then we will get to state $\overline{R}R$. The following matrix is the transition matrix for this Markov chain

$$M = \begin{pmatrix} 0,7 & 0.3 & 0 & 0 \\ 0 & 0 & 0.4 & 0.6 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.2 & 0.8 \end{pmatrix}$$

Now, to calculate the probability of raining on Thursday, given it rained on Monday and Tuesday, we first need to calculate the two step transition probability.

$$M^2 = \begin{pmatrix} \boxed{0.49} & 0.21 & \boxed{0.12} & 0.18 \\ 0.2 & 0.2 & 0.12 & 0.48 \\ 0.35 & 0.15 & 0.2 & 0.3 \\ 0.1 & 0.1 & 0.16 & 0.64 \end{pmatrix}$$

The probability to rain on Thursday is the sum of the boxed elements in the matrix above. So the desired probability is

$$p = 0.61.$$

■

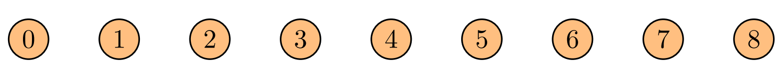
■ **Example 3.3** Suppose that balls are successively distributed among 8 urns, with each ball being equally likely to be put in any of these urns. What is the probability that there will be exactly 3 nonempty urns after 9 balls have been distributed? [Question from Ross]

Solution Before going through the solution, it might be more informative to explicitly write down what is the sample space Ω. At each time step, we basically throwing a 8 sided dice, and then put a ball at the urn number i if the output of the dice is i. So, each time we repeat the experiment, we will get a sequence of number each of which is one of $1, 2, \cdots, 8$. So the sample space will be the set of all sequences consisting of number $1, \cdots, 8$.

$$\Omega = \{21342\ldots, 44513\ldots, 11234\ldots, 88432\ldots, \cdots\}.$$

So, the outputs of the throwing dice at different steps are independent and identically distributed random variables. I.e. for a fixed $\omega \in \Omega$, The t-th element of the sequence is a random variable Y_t and all of the random variables $\{Y_t\}_t$ are independent and identically distributed. Note that the sample space associated with these random variables is $\{1, 2, 3, 4, 5, 6, 7, 8\}$ (i.e. the sample space of a 8 sided dice experiment).w

Let the random variable X_n be the number of filled (non-empty) urns at step n. So the state space will be $S = \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$, which is represented in the following graph.



This picture is not yet complete and we need to include the transition probabilities. We will do so by the first step argument. First, observe that $P(0, 0) = 0$, because if we start with all of the urns empty, then after one step, we have put a ball somewhere, thus it is impossible to end up with zero filled urn. Similarly, $P(8,8) = 1$, that is because if all of the urns are filled, then adding any new ball somewhere to any of the urns will keep the number of filled urns at 8. Then for $X_0 = n$, i.e. starting with n filled urns, we have

$$\mathbb{P}_n(X_1 = n-1) = 0.$$

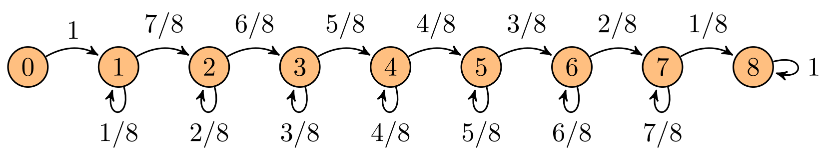
That is because starting with n filled urns, after doing one step, it is not possible to have less urns filled. I.e. after each step, we can either end up with more filled urns or the same number of filled urns. For $P(n,n)$, define the event E be the event of putting the ball in any of the filled urns. Thus E^c will be the probability of putting the ball at one of the empty urns.

$$\mathbb{P}_n(X_1 = n) = \underbrace{\mathbb{P}_n(X_1 = n|E)}_{1}\underbrace{\mathbb{P}_n(E)}_{n/8} + \underbrace{\mathbb{P}_n(X_1 = n|E^c)}_{0}\underbrace{\mathbb{P}_n(E^c)}_{(8-n)/8} = \frac{n}{8}.$$

Now for $\mathbb{P}(n, n+1)$ we can write

$$\mathbb{P}_n(X_1 = n+1) = \underbrace{\mathbb{P}_n(X_1 = n+1|E)}_{0}\underbrace{\mathbb{P}_n(E)}_{n/8} + \underbrace{\mathbb{P}_n(X_1 = n+1|E^c)}_{1}\underbrace{\mathbb{P}_n(E^c)}_{(8-n)/8} = 1 - \frac{n}{8}.$$

Thus the completed graph will be



The corresponding transition matrix will be

$$M = \begin{pmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1/8 & 7/8 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1/8 & 6/8 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 3/8 & 5/8 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 4/8 & 4/8 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 5/8 & 3/8 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 6/8 & 2/8 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 7/8 & 1/8 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}$$

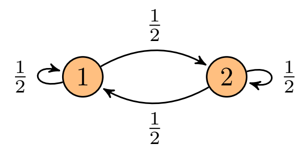
The probability that after 9 steps, there are exactly three empty urns is $(M^9)_{(0,3)}$, which is

$$p = (M^9)_{(0,3)} \approx 0.007572.$$

■

■ **Example 3.4** It is a good practice to derive the value of the transition probability of a simple Markov chain using the first principles. Consider the Markov chain representing a lamp that turns on with probability 1/2 and turns off with probability 1/2, and stays at the old state with probability 1/2. Thus we will have the following diagram for this Markov chain.
In this example, the state space is $S = \{0, 1\}$, and the sample space is

$$\Omega = \{(x_1, x_2, \cdots) : x_i \in S\}$$



which is basically the set of all sequences of one's and zero's. Given this, the random variables $(X_n)_n$ defined t be

$$X_n(\omega) = x_n,$$

where $\omega \in \Omega$ and x_n is the n-th letter in ω. Intuitively speaking, we know that

$$P(1,0) = \mathbb{P}(X_{n+1} = 1|X_n = 0) = \frac{1}{2}.$$

However, here we want to derive that number more explicitly by working directly with the elements of the probability space. First, we need to determine the event associated with $X_{n+1} = 1$. This is the event that has elements where the $n+1$-th position is 1. I.e.

$$E = \{(x_1, x_2, \cdots, x_n, 1, x_{n+2}, \cdots) : x_i \in S\}.$$

Similarly, we have

$$F = \{(x_1, x_2, \cdots, x_{n-1}, 0, x_{n+1}, \cdots) : x_i \in S\}.$$

So we have

$$\mathbb{P}(X_{n+1} = 1|X_n = 0) = \mathbb{P}(E|F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)} = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F \cap E) + \mathbb{P}(F \cap E^c)} = \frac{\frac{1}{|\Omega|}}{\frac{1}{|\Omega|} + \frac{1}{|\Omega|}} = \frac{1}{2}.$$

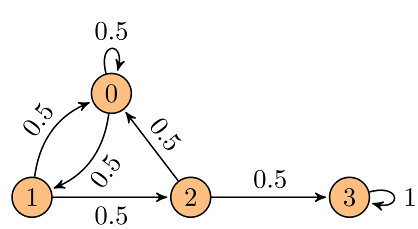
Note that $\mathbb{P}(E \cap F) = \frac{1}{|\Omega|}$, since out of many combinations of the sequence of zeros and ones, there is one one sequence whose n-th place is 0 and $n+1$-th place is 1. Furthermore, $\mathbb{P}(F \cap E^c) = \frac{1}{|\Omega|}$ as there is only one string where its n-th and $(n+1)$-th string are both zero. ■

■ **Example 3.5** In a sequence of independent flips of a fair coin, let N denote the number of flips until there is a run of three consecutive heads. Find

(a) $\mathbb{P}(N \leq 8)$,

(b) $\mathbb{P}(N = 8)$.

Solution Let X_n denote the number of consecutive heads at step n. For instance for the outcome $\omega \in \Omega$ where $\omega = HTTHTTHHHTTHT\ldots$, $X_2(\omega) = 0$ since the second symbol is T thus there is no consecutive heads. But $X_4(\omega) = 1$, as there is one consecutive heads at step 4. Lastly $X_9(\omega) = 3$, since there is three consecutive heads at step 9. This Markov chain will have the following transition diagram.



The transition probabilities are simply computed by the first step argument. For instance, for $P(0,1)$ we have

$$\mathbb{P}_0(X_1 = 1) = \underbrace{\mathbb{P}_0(X_1 = 1|H)}_{1}\underbrace{\mathbb{P}_0(H)}_{1/2} + \underbrace{\mathbb{P}_0(X_1 = 1|T)}_{0}\underbrace{\mathbb{P}_0(T)}_{1/2},$$

where H is the event that the flipped coin is heads and $H^c = T$. The transition matrix for this Markov chain will be

$$M = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

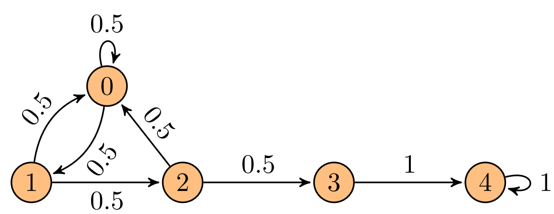
Since the state 3 is an absorbing state, then if we get there we will be there for the rest of our life! Thus the probability that the random walker has got there for $N \leq 8$ is simply $(M^8)_{(0,3)}$. Then

$$\mathbb{P}(N \leq 8) = 0.4180.$$

Now for part (b), the probability that the random walker has arrived at the state 3 right at the step 8, is

$$\mathbb{P}(N = 8) = \mathbb{P}(N \leq 8) - \mathbb{P}(N \leq 7) = 0.0508.$$

There is yet another approach that we can compute the probability $\mathbb{P}(N = 8)$. To do this, we need to consider 4 states $S = \{0, 1, 2, 3, 4\}$ where the state 4 is of when 3 consecutive heads has occurred at the past. So when the random walker enters the state 3 at some time, it moves to the state 4 at the next time and remains there forever. The state diagram will be



Then the transition matrix will be

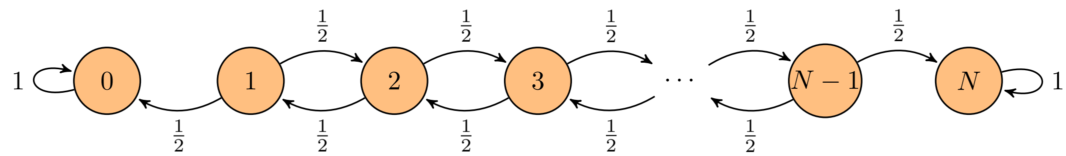
$$M = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0.5 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Then the probability $\mathbb{P}(N = 8) = (M^8)_{(0,3)} = 0.05080$.

■

■ **Example 3.6 — Gambler's Ruin.** Suppose Alice and Bob have in total of N coins. Alice and Bob play a game with a fair coin. When Alice wins, gets a coin from Bop, and vise versa. What is the probability that Alice wins if she starts with $0 \leq a \leq N$ coins.

Solution There are many ways to tackle a probability problem like this and the solution presented here is not the only way to find the solution to this problem. We want to model this with Markov chain whose state space is $\{0, 1, 2, \cdots, N\}$. Thus X_n represents the fortune of Alice after playing the games for n times.



Let p_a be the probability of Alice wining if she starts with a coins. First, observe that $p_0 = 0$ and $p_N = 1$. Let E denote that event of Alice wining the whole game. Also, let F_1 be the event in which she looses the first game and F_2 the event in which she wins the first game. Then

$$p_a = \mathbb{P}_a(E) = \underbrace{\mathbb{P}_a(E|F_1)}_{\mathbb{P}(E|F_1, X_0=a)} \mathbb{P}(F_1) + \underbrace{\mathbb{P}_a(E|F_1^c)}_{\mathbb{P}(E|F_1^c, X_0=a)} \mathbb{P}(F_1^c)$$

(note that this identity is actually true for any set F_1, but here F_1 is the specific event explained above). The probability that she looses or wins the first game is $\frac{1}{2}$. Also, observe that $\mathbb{P}_a(E|F_1) = p_{a+1}$ (since if she wins the first game she will have one more coin) and $\mathbb{P}_a(E|F_1^c) = p_{a-1}$. Thus

$$p_a = \frac{1}{2}p_{a+1} + \frac{1}{2}p_{a-1}.$$

Now we can solve this recurrent equation with the characterization polynomial which is $2 = X + 1/X$ or $X^2 - 2X + 1 = (X-1)^2 = 0$. Thus the characteristic polynomial has a double root. Thus

$$p_a = (Aa + B)(1)^a = Aa + B.$$

Since $p_0 = 0, \ p_N = 1$, then it turns out that

$$p_a = \frac{a}{N}.$$

■

■ **Example 3.7 — Gambler's Ruin with Draw.** Let Alice and Bob play Rock-Paper-Scissors. If Alice and Bob has a total of N coins, and at each play, the winner gets one coin from the loser, what is the probability that Alice will win the game if he starts with a coins. When they draw, then they repeat the game (or equivalently, they play another game without any coins exchange).

Solution We need to do a first step analysis similar to what we did before. Let E be the event that Alice wins the whole game, and the event $F = F_{-1} \cup F_0 \cup F_1$ where

F_{-1}: Alice loses the first game,
F_0: Alice draws the first game,
F_1: Alice wins the first game.

It is clear that $\mathbb{P}(F) = 1$, since the components are mutually disjoint. Thus $E \cap F_{-1}, \ E \cap F_0, \ E \cap F_1$ are also mutually disjoints where. Thus we can write

$$\mathbb{P}_a(E) = \mathbb{P}_a(E \cap F_{-1}) + \mathbb{P}_a(E \cap F_0) + \mathbb{P}_a(E \cap F_1) = \mathbb{P}_a(E|F_{-1})\mathbb{P}_a(F_{-1}) + \mathbb{P}_a(E|F_0)\mathbb{P}_a(F_0) + \mathbb{P}_a(E|F_1)\mathbb{P}_a(F_1).$$

Since the game is fair we know

$$\mathbb{P}_a(F_{-1}) = \mathbb{P}_a(F_0) = \mathbb{P}_a(F_1) = \frac{1}{3}.$$

Furthermore, we know

$$\mathbb{P}_a(E|F_{-1}) = p_{a-1}, \qquad \mathbb{P}_a(E|F_0) = p_a, \qquad \mathbb{P}_a(E|F_1) = p_{a+1}.$$

Thus the first step analysis will lead to the following identity.

$$\mathbb{P}_a(E) = p_a = \frac{1}{3}(p_{a-1} + p_a + p_{a+1}),$$

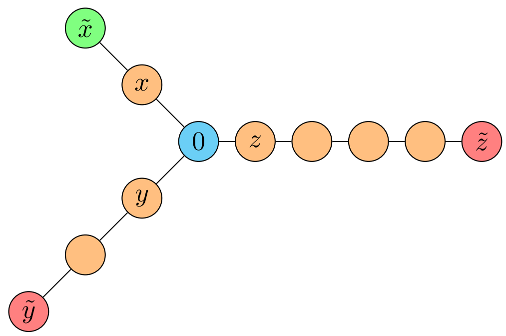
which after simplification becomes

$$2p_a = p_{a-1} + p_{a+1},$$

which is the same recursive formula we got in the previous example. So the possibility of the draw, will not change the behaviour of the system.

■

■ **Example 3.8** Consider the a simple random walker on the following graph. Let $B = \{T_{\tilde{x}} < T_{\{\tilde{z},\tilde{y}\}}\}$. Compute the probability $\mathbb{P}_0(B)$.



Solution This problem is simply asking what is the probability that we hit $\tilde{x}$ state before hitting any of $\tilde{y}$ or $\tilde{z}$ states, given the fact that the random walker starts from the state 0. To keep unnecessary details out of the way, we have only labeled the vertices that we will use in our analysis. We will have the following notation to simplify the solution

$$p_v = \mathbb{P}_v(B),$$

where v is any vertex in the graph. Note that starting at 0, i.e. $X_0 = 0$, then going to any of the states x, y, or z, are mutually disjoint events, and the probability of the union of these events is one. With our first time step analysis (see Proposition 3.2) we can write

$$\mathbb{P}_0(B) = \frac{1}{3}(p_x + p_y + p_z).$$

Now we need to analyze each of terms in the RHS. Let's start with p_z. Consider two events $\{T_0 < T_{\tilde{z}}\}$ and $\{T_0 > T_{\tilde{z}}\}$, where the first time is the event where the random walker hits the 0 state before hitting the $\tilde{z}$ step first, and the second one is the vice versa. These two events are disjoint and the probability of the union is 1. Thus we write the conditional expansion of p_z based on these events

$$p_z = \mathbb{P}_z(B) = \mathbb{P}_z(B|T_0 < T_{\tilde{z}})\mathbb{P}_z(T_0 < T_{\tilde{z}}) + \mathbb{P}_z(B|T_0 > T_{\tilde{z}})\mathbb{P}_z(T_0 > T_{\tilde{z}}).$$

We know that $\mathbb{P}_z(B|T_0 > T_{\tilde{z}}) = \mathbb{P}(B|X_0 = z, X_i = \tilde{z})$ for some $i > 0$. From Markov property it follows that

$$\mathbb{P}(B|X_0 = z, X_i = \tilde{z}) = \mathbb{P}(B|X_i = \tilde{z}) = \mathbb{P}(B|X_0 = \tilde{z}) = p_{\tilde{z}}.$$

Also $\mathbb{P}_z(B|T_0 < T_{\tilde{z}}) = \mathbb{P}_0(B) = p_0$ by the Markov property. Lastly, $\mathbb{P}_z(T_0 < T_{\tilde{z}})$ is determined by the Gambler's ruin method we say before, which is basically

$$\mathbb{P}_z(T_0 < T_{\tilde{z}}) = \frac{5}{4}, \qquad \mathbb{P}_z(T_0 > T_{\tilde{z}}) = \frac{1}{5}.$$

By doing the same kind of analysis for p_x as well as p_y we will get

$$p_z = \frac{4}{5}p_0, \qquad p_y = \frac{2}{3}p_0, \qquad p_x = \frac{1}{2}p_0 + \frac{1}{2}.$$

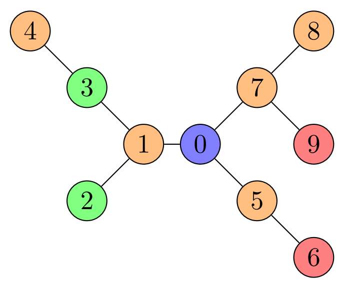
Now by substituting in the identity we got from the first time step argument, we can fine that

$$p_0 = \frac{15}{31},$$

And this completes our solution for the problem.

■

■ **Example 3.9** Consider the graph $\gamma = (V, E)$ drawn below. Set $Z = \{2, 3\}$, and $W = \{6, 9\}$. Compute $\mathbb{P}_0(T_Z < T_W)$. In colors: we start at blue, win if we reach green, and lose of we reach red.



■

Solution As always, we start with our powerful tool in hand, which is the first step argument (which is basically a special form of the more general conditional expansion). We start with first step argument at state 0. We will get

$$\mathbb{P}_0(B) = \frac{1}{3}(\mathbb{P}_1(B) + \mathbb{P}_7(B) + \mathbb{P}_5(B)),$$

and now we need to analyze each of the terms in the right hand side. We start with $\mathbb{P}_5(B)$ which is the most straight forward one. As we saw in the last example, we can analyze this state with a conditional expansion on the two disjoint events, whose union probability is 1. Let those two events be $\{T_6 < T_0\}$ (where the random walker hits the state 6 before hitting the state 0), and $\{T_6 > T_0\}$, where the random walker hits the state 0 before hitting the state 6. Thus the expansion will be

$$\mathbb{P}_5(B) = \mathbb{P}_5(B|T_6 < T_0)\mathbb{P}_5(T_6 < T_0) + \mathbb{P}_5(B|T_6 > T_0)\mathbb{P}_5(T_6 > T_0).$$

We know that if we hit the state 6 before 0, we have no chance to hit any of the green states (we will lose). Thus

$$\mathbb{P}_5(B|T_6 < T_0) = 0.$$

And from the Gambler's ruin we know that $\mathbb{P}_5(T_6 > T_0) = 1/2$, and from the Markov property we know that $\mathbb{P}_5(B|T_6 > T_0) = \mathbb{P}_0(B)$, because the conditional probability $\mathbb{P}_5(B|T_6 > T_0)$ is basically stating what is the probability of B happening, if we start from 5 and $X_i = 0$ for some i in the future. Thus

$$\mathbb{P}_5(B) = \frac{\mathbb{P}_0(B)}{2}.$$

Now, we need to analyze the term $\mathbb{P}_1(B)$. Again, at this step, we do another first step analysis.

$$\mathbb{P}_1(B) = \frac{1}{3}(\underbrace{\mathbb{P}_3(B)}_{=1} + \underbrace{\mathbb{P}_2(B)}_{=1} + \mathbb{P}_0(B)) = \frac{2 + \mathbb{P}_0(B)}{3}.$$

Note that from the assumption, we know that if we reach any of green states, then we are declared winner, that is why we have $\mathbb{P}_3(B) = \mathbb{P}_2(B) = 1$. Now it only remains to analyze the term $\mathbb{P}_7(B)$. Again, similar to the case above, we do a first time step argument

$$\mathbb{P}_7(B) = \frac{1}{3}(\mathbb{P}_0(B) + \underbrace{\mathbb{P}_8(B)}_{=\mathbb{P}_7(B)} + \underbrace{\mathbb{P}_9(B)}_{=0}) \implies \mathbb{P}_7(B) = \frac{\mathbb{P}_0(B)}{2}.$$

Note that $\mathbb{P}_8(B) = \mathbb{P}_7(B)$ by a first stem analysis when starting at the state 8. Putting all of these terms back to the original identity we derived the first, we can conclude that

$$p_0 = \mathbb{P}_0(B) = \frac{2}{5}.$$

3.3 Solved Problems

■ **Problem 3.1** The French roulette game has slots numbered from 0 to 36. The slot 0 is green, Among the slots from 1 to 36, 18 are black and 18 are red. Alex goes to a casino to play roulette. Their strategy is to always bet "red". They start with 50 coins, play 1 coin each turn, and stop when reaching 100 or getting broke.

(a) What is the probability that Alex reaches 100?

(b) How many coins should Alex start with to have about 50% chance to reach 100?

Solution (a) Let B be the event $B = \{T_{100} < T_0\}$ and we are looking for $\mathbb{P}_a(B)$ where $0 \leq a \leq 100$ and indicates the number of coins we are starting with. First observe that

- $p_0 = 0$: Since if we start with zero coins we are already broken and the game is over.
- $p_{100} = 1$: Since if we start with 100 coins then we won the game and the game is finished.

To compute the probability for intermediate values of a, we do the first step argument. Let WF be the event where Alex wins the first bet, and LF the event where Alex loses the first bet. Then we can write

$$p_a = \mathbb{P}_a(B) = \mathbb{P}_a(B|WF)\mathbb{P}_a(WF) + \mathbb{P}_a(B|LF)\mathbb{P}_a(LF).$$

Since there are 18 red spots, then the chance to win the first bet is

$$\mathbb{P}_a(WF) = \frac{18}{37}.$$

and since there are 19 non-red spots in total, then the chance to win is

$$\mathbb{P}_a(LF) = \frac{19}{37}.$$

Also, from Markov property, we know that

$$\mathbb{P}_a(B|WF) = p_{a+1}, \qquad \mathbb{P}_a(B|LF) = p_a.$$

Thus the first step argument formula will be

$$p_a = \frac{18}{37} p_{a+1} + \frac{19}{37} p_{a-1} \implies \boxed{37 p_a = 18 p_{a+1} + 19 p_{a-1}}.$$

The characteristic equation for the recursive equation is

$$37 = 18x + \frac{19}{x} \implies \boxed{18x^2 - 37x + 19 = 0}.$$

We can write it as $(x-1)(18x-19) = 0$. Thus the roots will be

$$r_1 = 1, \qquad r_2 = \frac{19}{18}.$$

So

$$p_a = A + B r_2^a.$$

To fine A and B we use the fact $p_0 = 0$, and $p_{100} = 1$. Then $A = -B$, and $A = 1/(1 - r_2^{100})$. Thus

$$\boxed{p_a = \frac{1 - r_2^a}{1 - r_2^{100}}}.$$

(b) We basically need to compute find a for which $p_a = 1/2$. Thus we need to solve for a

$$\frac{1 - r_2^a}{1 - r_2^{100}} = \frac{1}{2}.$$

After some algebra we will find

$$\boxed{a = \frac{\ln\left(\frac{1 + r_2^{100}}{2}\right)}{\ln(r_2)} \approx 87.26}.$$

Thus we need to start with at least 88 coins to have a 50% chance of winning.

□

■ **Problem 3.2** There are 6 coins on a table, each showing heads (H) or tails (T). In each step we

- Select uniformly one of the coins.
- If it is heads, toss it and replace on the table (with random side).
- If it sis tails, toss it. If it comes up heads, leave it at that. If it comes up tails, toss it a second time, and leave the result as it is. Let X_n be the number of heads showing after n such steps. Answer the following questions
  (a) Determine the transition probabilities for this Markov chain.
  (b) Draw the transition diagram and write the transition matrix.
  (c) What is $\mathbb{P}(X_2 = 4 | X_0 = 5)$?

Solution (a) To compute the transition probabilities, we need to perform the first step analysis. Let the events

$$I = \{X_1 = a+1\}, \qquad S = \{X_1 = a\}, \qquad D = \{X_1 = a-1\},$$

where $0 \leq a \leq 6$ is the number of heads. So to compute the transition probabilities, we need to compute

$$P(a,a+1) = \mathbb{P}_a(I), \qquad P(a,a) = \mathbb{P}_a(S), \qquad P(a,a-1) = \mathbb{P}_a(D).$$

We start with $\mathbb{P}_a(I)$. Let ST be the event where the selected coin is tails, and SH be the event where the selected coin is heads. These two events are disjoint and the probability of their union is 1, thus

$$\mathbb{P}_a(I) = \underbrace{\mathbb{P}_a(I|SH)}_{\text{see Eq (2.I.1)}} \underbrace{\mathbb{P}_a(SH)}_{\frac{a}{6}} + \underbrace{\mathbb{P}_a(I|ST)}_{\text{see Eq (2.I.2)}} \underbrace{\mathbb{P}_a(ST)}_{\frac{6-a}{6}}. \tag{2.I}$$

Note that if we start with a coins heads, then the chance we choose a random coin from the table and find it heads is $\frac{a}{6}$, hence $\mathbb{P}_a(SH) = \frac{a}{6}$, and $\mathbb{P}_a(ST) = \frac{6-a}{6}$. Now we need to expand the remaining terms with appropriate conditioning. Let TT be the event where we toss a coin and find it tails and TH be the event where we toss a coin and find it heads. Thus we can write

$$\mathbb{P}_a(I|SH) = \underbrace{\mathbb{P}_a(I|SH,TH)}_{0}\underbrace{\mathbb{P}_a(TS)}_{\frac{1}{2}} + \underbrace{\mathbb{P}_a(I|SH,TT)}_{0}\underbrace{\mathbb{P}_a(TT)}_{\frac{1}{2}}. \tag{2.I.1}$$

Note that $\mathbb{P}_a(TT) = \mathbb{P}_a(TH) = \frac{1}{2}$, since the coin tossing is fair. Also, note that $\mathbb{P}_a(I|SH,TH) = \mathbb{P}_a(I|SH,TT) = 0$ since if we select a heads, and then toss it, finding it either heads or tails will not increase the total number of heads on the table. Similarly, for the other term in (2.1) we have

$$\mathbb{P}_a(I|ST) = \underbrace{\mathbb{P}_a(I|ST,TH)}_{1}\underbrace{\mathbb{P}_a(TH)}_{\frac{1}{2}} + \underbrace{\mathbb{P}_a(I|ST,TT)}_{\text{see Eq (2.I.3)}}\underbrace{\mathbb{P}_a(TT)}_{\frac{1}{2}}. \tag{2.I.2}$$

Now we need to expand the remaining terms in the equation above.

$$\mathbb{P}_a(I|ST,TT) = \underbrace{\mathbb{P}_a(I|ST,TT,TH)}_{1}\mathbb{P}_a(TH) + \underbrace{\mathbb{P}_a(I|ST,TT,TT)}_{0}\mathbb{P}_a(TT) = \frac{1}{2}. \tag{2.I.3}$$

Putting all together we can write

$$\boxed{P(a,a+1) = \mathbb{P}_a(I) = \frac{6-a}{8}}.$$

Similarly, we can compute other transition probabilities. For instance for $\mathbb{P}_a(S)$ we can write

$$\mathbb{P}_a(S) = \underbrace{\mathbb{P}_a(S|SH)}_{\text{see Eq (2.S.1)}} \underbrace{\mathbb{P}_a(SH)}_{\frac{a}{6}} + \underbrace{\mathbb{P}_a(S|ST)}_{\text{see Eq (2.S.2)}} \underbrace{\mathbb{P}_a(ST)}_{\frac{6-a}{6}}. \tag{2.S}$$

and for the remaining terms we can write

$$\mathbb{P}_a(S|SH) = \underbrace{\mathbb{P}_a(S|SH,TH)}_{1}\underbrace{\mathbb{P}_a(TS)}_{\frac{1}{2}} + \underbrace{\mathbb{P}_a(S|SH,TT)}_{0}\underbrace{\mathbb{P}_a(TT)}_{\frac{1}{2}}, \tag{2.S.1}$$

and

$$\mathbb{P}_a(S|ST) = \underbrace{\mathbb{P}_a(S|ST,TH)}_{0}\underbrace{\mathbb{P}_a(TH)}_{\frac{1}{2}} + \underbrace{\mathbb{P}_a(S|ST,TT)}_{\text{see Eq (2.S.3)}}\underbrace{\mathbb{P}_a(TT)}_{\frac{1}{2}}. \tag{2.S.2}$$

And for the remaining term above

$$\mathbb{P}_a(S|ST,TT) = \underbrace{\mathbb{P}_a(S|ST,TT,TH)}_{0}\mathbb{P}_a(TH) + \underbrace{\mathbb{P}_a(S|ST,TT,TT)}_{1}\mathbb{P}_a(TT) = \frac{1}{2}. \qquad (2.S.3)$$

and by putting all together we will get

$$\boxed{P(a,a) = \mathbb{P}_a(S) = \frac{6+a}{24}}.$$

Finally, since $\mathbb{P}_a(I \cup S \cup D) = 1$, and I, S, D are mutually disjoint, we can write

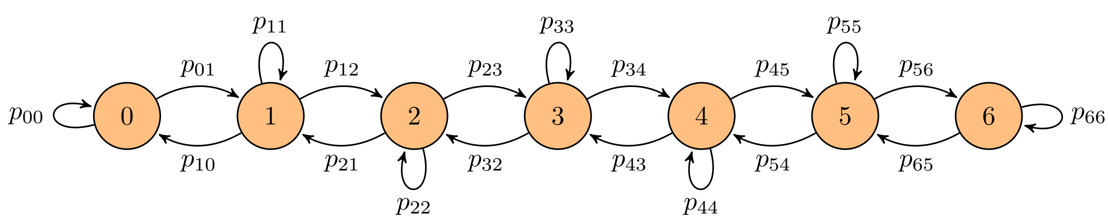
$$\mathbb{P}_a(D) = 1 - (\mathbb{P}_a(I) + \mathbb{P}_a(S)),$$

hence

$$\boxed{P(a,a-1) = \mathbb{P}_a(D) = \frac{a}{12}}.$$

so the transition probabilities are as calculated.

(b) The transition diagram is plotted below.



And the transition matrix is

$$M = \begin{pmatrix} 1/4 & 3/4 & 0 & 0 & 0 & 0 & 0 \\ 1/12 & 7/24 & 5/8 & 0 & 0 & 0 & 0 \\ 0 & 1/6 & 1/3 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/4 & 3/8 & 3/8 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 5/12 & 1/4 & 0 \\ 0 & 0 & 0 & 0 & 5/12 & 11/24 & 1/8 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix}$$

(c) $\mathbb{P}(X_2 = 4|X_0 = 5)$ is the second transition probability $P_2(5,4)$. To compute this, we need to fine the element in the 6-th row and 5-th column in the M^2 matrix, which is basically the inner product between the vectors formed by the 6-th row and the 5-th column.

$$P_2(5,4) = (\frac{5}{12})^2 + \frac{11}{24} \cdot \frac{5}{12} = \frac{35}{96}$$

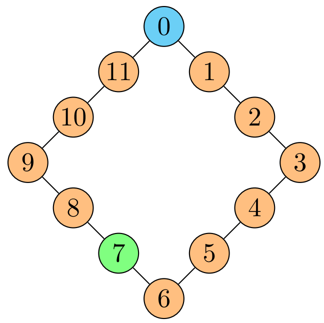
which after simplification becomes

$$\boxed{P_2(5,4) = \frac{35}{96}}.$$

$\square$

■ **Problem 3.3** A clock is broken. It has only one hand which moves every hour either clockwise with probability 1/2 or counter-clockwise with probability 1/2 (the numbers are from 0 to 11 and the hand moves by one full hour when it moves). Assume it starts at 0. What is the probability that it reaches 7 before coming back to 0 for the first time?

Solution First, let's draw the graph representing the state space of the random variable of interest.



Define the event B be $B = \{T_0^+ > T_7\}$. We are interested in finding $\mathbb{P}_0(B)$. Now we can perform the first step argument as follows

$$p_0 = \frac{1}{2}(p_1 + p_{11}). \tag{3.1}$$

Then we analyze each term in the right hand side of the equation above. For p_1 we have

$$\mathbb{P}_1(B) = \underbrace{\mathbb{P}_1(B|T_0 > T_7)}_{1}\underbrace{\mathbb{P}_1(T_0 > T_7)}_{1/5} + \underbrace{\mathbb{P}_1(B|T_0 < T_7)}_{0}\underbrace{\mathbb{P}_1(T_0 < T_7)}_{6/7} = \frac{1}{5}.$$

Note that $\mathbb{P}_1(B|T_0 > T_7) = 1$ since it literally means the random walker reaches 7 before 0. Also $\mathbb{P}_1(B|T_0 < T_7) = 0$ since the event B is conditioned on reaching 0 before 7, which is clearly 0. The term $\mathbb{P}_1(T_0 > T_7)$ is computed using the Gambler's ruin analysis. Similarly, for the p_{11} term we have

$$\mathbb{P}_{11}(B) = \underbrace{\mathbb{P}_{11}(B|T_0 > T_7)}_{1}\underbrace{\mathbb{P}_{11}(T_0 > T_7)}_{1/7} + \underbrace{\mathbb{P}_{11}(B|T_0 < T_7)}_{0}\mathbb{P}_{11}(T_0 < T_7) = \frac{1}{7}.$$

The rationale behind the values of the terms are the same as the ones discussed above. Now we can substitute everything in (3.1)

$$\boxed{p_0 = \frac{1}{2}(\frac{12}{35}) = \frac{6}{35}}.$$

■ **Problem 3.4** The Fibonacci sequence is the sequence $(F_n)_{n\geq 0}$ defined by $F_0 = 0, F_1 = 1$ and

$$F_{n+2} = F_{n+1} + F_n \quad \text{for } n \geq 0.$$

Find a general formula for F_n

Solution First, we construct the characteristic polynomial of the sequence. From the recursive formula we can write

$$X^2 = X + 1 \implies \boxed{X^2 - X - 1 = 0}.$$

The roots of the equation is

$$r_1, r_2 = \frac{1 \pm \sqrt{5}}{2}.$$

Now the general formula will be

$$F_n = Ar_1^n + Br_2^n.$$

To find the coefficients, we utilize the first two terms

$$0 = A + B, \qquad 1 = \frac{1}{2}(A+B) + \frac{\sqrt{5}}{2}(A-B).$$

This system of equations implies that

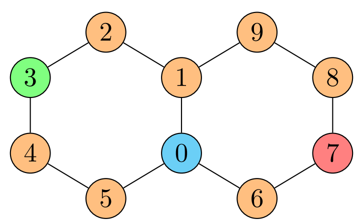
$$A = \frac{1}{\sqrt{5}}, \qquad B = \frac{-1}{\sqrt{5}}.$$

Thus the general formula will be

$$\boxed{F_n = \frac{1}{\sqrt{5}}((\frac{1+\sqrt{5}}{2})^n - \frac{1-\sqrt{5}}{2})^n)}.$$

□

■ **Problem 3.5** Let (X_n) be the simple random walk on the following graph. Compute $\mathbb{P}_0(T_3 < T_7)$.



Solution For a much more simpler solution, let's define the two following notations

$$B = \{T_3 < T_7\}, \qquad p_v = \mathbb{P}_v(B).$$

Then, by first step argument at state 0, we can write

$$p_0 = \frac{1}{3}(p_5 + p_6 + p_1). \tag{5.1}$$

Now we need to evaluate each of the terms in the right hand side. We start with p_5.

$$p_5 = \mathbb{P}_5(B) = \underbrace{\mathbb{P}_5(B|T_3 < T_0)}_{1}\underbrace{\mathbb{P}_5(T_3 < T_0)}_{1/3} + \underbrace{\mathbb{P}_5(B|T_3 > T_0)}_{p_0}\underbrace{\mathbb{P}_5(T_3 > T_0)}_{2/3} = \frac{1}{3} + \frac{2}{3}p_0.$$

note that $\mathbb{P}_5(B|T_3 < T_0) = 1$, since if we get to state 3, before getting to state 0, then it means that we have reached the state 3 before reaching the state 7, thus the event B occurs with probability 1. Also $\mathbb{P}_5(T_3 < T_0) = 1/3$ from the Gambler's ruin. Furthermore $\mathbb{P}_5(B|T_3 > T_0) = p_0$ by using the Markov property, and finally $\mathbb{P}_5(T_3 > T_0) = 2/3$ by the Gambler's ruin.
Now, we need to evaluate the term p_6. To analyze this term, we will do a first step argument starting at this point

$$p_6 = \mathbb{P}_6(B) = \frac{1}{2}(\underbrace{p_7}_{0} + p_0) = \frac{p_0}{2}.$$

Note that $p_7 = 0$, since then the event B has not occurred.
Finally, we need to analyze the term p_1. Again, by first step argument on this state we have

$$p_1 = \frac{1}{3}(p_0 + p_9 + p_2).$$

By doing a analysis on p_9 similar to the one we did for 5, we can write

$$p_9 = \mathbb{P}_9(B) = \underbrace{\mathbb{P}_9(B|T_7 < T_1)}_{0}\mathbb{P}_9(T_7 < T_1) + \underbrace{\mathbb{P}_9(B|T_7 > T_1)}_{p_1}\underbrace{\mathbb{P}_9(T_7 > T_1)}_{2/3} = \frac{2}{3}p_1.$$

The rationale behind the values for each term in the equation above, is exactly the same as in analyzing the terms of p_5.
Now, we analyze the term p_2 by performing another first step analysis, similar to the one we did for state 6.

$$p_2 = \frac{1}{2}(\underbrace{p_3}_{1} + p_1) = \frac{1}{2}(1 + p_1).$$

Now we can calculate p_1 in terms of p_0 which turns out to be

$$p_1 = \frac{6}{11}p_0 + \frac{3}{11}.$$

Now we insert all of the terms in the equation (5.1) to get

$$\begin{aligned}
3p_0 &= \frac{1}{3} + \frac{2}{3}p_0 + \frac{1}{2}p_0 + \frac{6}{11}p_0 + \frac{3}{11} \\
\implies& 3p_0 - \frac{113}{66}p_0 = \frac{40}{33} \\
\implies& p_0 = \frac{66}{85} \cdot \frac{40}{33} = \frac{16}{17} \\
\implies& \boxed{p_0 = \frac{16}{17}}.
\end{aligned}$$

□

4. Stationary Distributions of Markov Chains

4.1 Time Evolution of Distributions

Let $\{X_n\}_n$ be a discrete Markov with finite state space $S = \{1, 2, \cdots, N\}$. Then, we know that starting at a state $X_0 = 1$, then the probability to be at state j after one step is the $(1, j)$ element of the transition matrix. To be more concrete, let's assume that the Markov chain is defined on $\{1, 2, 3\}$, and assume $X_0 = 1$. Then $P(1, 3) = \mathbb{P}_1(X_1 = 3)$ is the element $(1, 3)$ of the transition matrix. Don't forget that $\{X_n\}$ are all random variables. Thus while we can talk about the cases that what will happen if, for example X_0, be $X_0 = 1$ and etc. We can also talk about the probability that the random variable has specific values, which is the idea of distribution. Let $\mu_{X_0}(i)$ for $i \in \{1, 2, 3\}$ be the distribution of X_0. In other words, we have

$$\mu_{X_0}(i) = \mathbb{P}(X_0 = i) \qquad \text{for } i \in \{1, 2, 3\}.$$

We can also introduce the vector notation for the distribution. Note that we can drop the subscript X_0 as shown in the following notation.

$$\mu_0 = \mu_{X_0} = (\mu_{X_0}(1), \mu_{X_0}(2), \mu_{X_0}(3)).$$

Now, suppose we want to find the distribution of X_1 given the distribution of X_0, i.e. we want to calculate the time evolution of the distribution after one step. Let's calculate what happens for μ_0 after one step:

$$\mu_1 = \mu_{X_1} = (\mu_1(1), \mu_1(2), \mu_1(3)).$$

To calculate $\mu_1(1)$ we can write

$$\mu_1(1) = \mathbb{P}(X_1 = 1) = \underbrace{\mathbb{P}_1(X_1 = 1)}_{P(1,1)} \underbrace{\mathbb{P}(X_0 = 1)}_{\mu_0(1)} + \underbrace{\mathbb{P}_2(X_1 = 1)}_{P(2,1)} \underbrace{\mathbb{P}(X_0 = 2)}_{\mu_0(2)} + \underbrace{\mathbb{P}_3(X_1 = 1)}_{P(3,1)} \underbrace{\mathbb{P}(X_0 = 3)}_{\mu_0(3)}.$$

In summary

$$\mu_1(1) = P(1, 1)\mu_0(1) + P(2, 1)\mu_0(2) + P(3, 1)\mu_0(3).$$

By a similar argument, we can quickly see that

$$\mu_1 = \mu_0 M$$

where M is the transition matrix.

Proposition 4.1 Let $X_0 \sim \mu_0$. Then $X_n \sim \mu_n$ where

$$\mu_n = \mu_0 M^n$$

where M is the transition matrix.

Now we can state the following important observation.

Observation 4.1.1 For a given discrete Markov chain $\{X_n\}$ defined on a *finite* state space $S = \{1, 2, \cdots, N\}$, the sequence of distributions of the random variables at each time step

$$\mu_{X_n} = \mu_n = (\mu_n(1), \mu_n(2), \cdots, \mu_n(N)),$$

defines a discrete time Markov chain with continuous state space $\mathbb{R}^{N-1}$. The transition matrix for $\{\mu_n\}_n$ will be the same as the original Markov chain. The state space will in fact be an affine hyperplane at $\mathbb{R}^N$, that intersects each axis at 1. That is because we require the distributions to sum up to 1. Thus we will have a discrete map

$$\mu_{n+1} = \mu_n M.$$

Observation 4.1.2 — Be careful here!. You need to be careful here and pay special attention to the notations and conventions here. We said that any Markov chain defines another Markov chain $\{\mu_n\}$ which is the distribution of the Markov chain random variable at each step. And also we said that this Markov chain has the same Transition matrix as the original Markov chain. However, you need to note that μ is defined to be a *row* vector

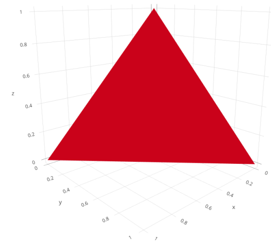
$$\mu_n = (\mu_n(1) \quad \mu_n(2) \quad \cdots \quad \mu_n(N)).$$

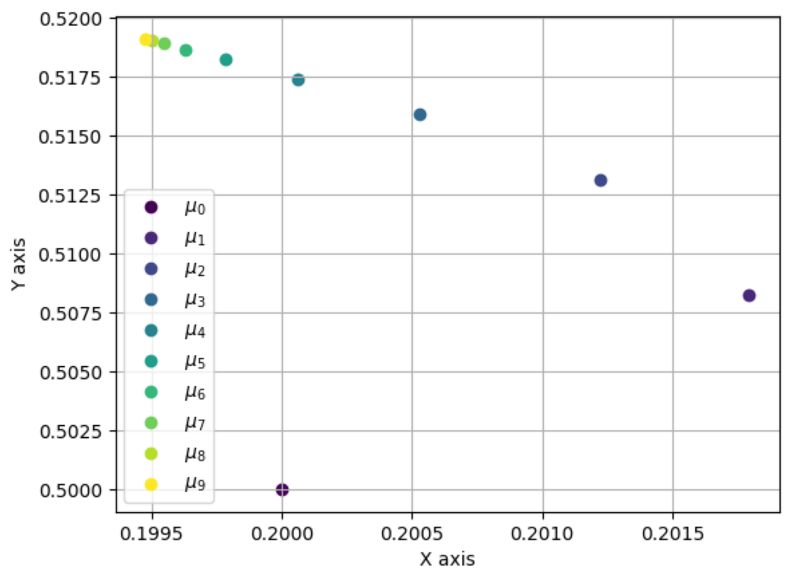
Thus value of μ_{n+1} will be μn multiplied by the transition matrix from *left*. We can develop the whole theory using the notion of transpose, but here we will keep this convention as it is more straight forward.

■ **Example 4.1** Consider the Markov chain $\{X_n\}_n$ defined on the state space $S = \{1, 2, 3\}$, with the following transition probability

$$M = \begin{bmatrix} 0.43 & 0.30 & 0.27 \\ 0.10 & 0.77 & 0.12 \\ 0.22 & 0.20 & 0.58 \end{bmatrix}$$

Now assume that the distribution of X_0 is $\mu_0 = (0.2, 0.5, 0.3)$. The $\{\mu_n\}$ is a Markov chain defined on $\mathbb{R}^3$. To be more precise, since all of the distributions should be normalized, the state space if $\{\mu_n\}$ is in fact the 2D plane that cuts through each axis at 1, as shown in the figure below. This is basically a two dimensional manifold that is embedded in 3 dimensional Euclidean space. We can consider the projection of μ_n on the $x - y$ plane as the 2d atlas (this projection is the diffeomorphism). The following is the time evolution of the distribution with $\mu_0 = (0.2, 0.5, 0.3)$. This example demonstrates that the time evolution of the distribution of a Markov chain is a Markov chain with the same transition matrix. ■





The following is the definition of the stationary distribution for a Markov chain.

Definition 4.1 Let $\{X_n\}$ be a Markov chain defined on the state space S. The distribution vector π is a stationary distribution if we have

$$\pi = \pi P.$$

■ **Remark** Given a distribution, we can do the following test to check if it is a stationary distribution. First, it needs to be a distribution, i.e.

$$\sum_{x \in S} \pi(x) = 1,$$

And secondly, it needs to satisfy the definition for a stationary distribution, i.e. for all $x \in S$ we have

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

Observation 4.1.3 A stationary distribution is a left eigenvector for the transition matrix with eigenvalue 1.

Proposition 4.2 Let $\Gamma = (E, V)$ be a connected graph with at least two vertices. Then the stationary distribution for a simple random walk on the graph is given as

$$\pi(x) = \frac{\deg(x)}{2|E|}$$

where $|E|$ is the number of edges of the graph.

■ **Remark** We can show that this is true by two simple checks of the stationary distributions. First, we need to check this is indeed distribution, i.e. sums up to 1.

$$\sum_{x\in S} \pi(x) = \frac{1}{2|E|} \sum_{x\in S} \deg(x) = \frac{2|E|}{2|E|} = 1.$$

Now, we need to check if it is a stationary distribution, i.e. $\forall x \in S$ we have

$$\sum_{y\in S} \pi(y)P(y,x) = \sum_{x\sim y} \frac{\pi(y)}{\deg(y)} = \sum_{x\sim y} \frac{\deg(y)}{2|E|} \frac{1}{\deg(y)} = \frac{\deg(x)}{2|E|}.$$

Observation 4.1.4 — Intuition behind the distribution. Intuitively speaking, the notion of distribution for a Markov chain on a graph, is intuitively speaking similar to the idea of considering the vertices as containers that has some sort of liquid in it, and the transition probability as the rate at which a flow moves between these vertices. Thus at each step, the liquid moves around and the stationary distribution is a distribution where the input and output flow of each vertex is just balanced, that we see no change in the liquid content of each vertex through time.

4.1.1 Uniqueness of the stationary distribution on the irreducible and finite Markov Chains

For this section, we will need the notion of a harmonic function on a graph, as defined below.

Definition 4.2 Let $\{X_n\}$ be a Markov chain defined on the *finite* state space S. Then a function $h : S \to \mathbb{R}$ is harmonic if it satisfies

$$h(x) = \sum_{y\sim x} P(x,y)h(y) \qquad \forall x \in S$$

The notion of a harmonic function on a graph has more intuitive characterization, as follows.

Observation 4.1.5 Let $\Gamma = (V, E)$ be a connected graph. Then a harmonic function defined on Γ is $h : V \to \mathbb{R}$ where $\forall x \in V$ we have

$$h(x) = \sum_{y\sim x} P(x,y)h(y) = \frac{1}{\deg(x)} \sum_{y\sim x} h(y).$$

Thus we can see that a function defined on a graph is harmonic if its value at a particular vertex x is the average of its values at the neighborhood vertices.

Lemma 4.1 Let $\{X_n\}$ be a Markov chain defined on a finite state space S, and denote the transition matrix of this Markov chain as P. Then h is a harmonic function on S if it satisfies

$$Ph = h,$$

i.e. h is the right eigenvector of P with eigenvalue 1.

The following proposition shows a deep connection between the harmonic functions and the first step analysis.

Proposition 4.3 Let $\{X_n\}$ be a Markov chain defined on a finite state space S. Let $W, Z \subset S$ disjoint. Then the following function

$$h(x) = \mathbb{P}_x(T_W < T_Z)$$

is harmonic on $S \backslash (W \cup Z)$.

Proof. This immediately follows from the first step argument. Define the event $E = T_W < T_Z$. Then by the first step analysis we have

$$\mathbb{P}_x(E) = \sum_{y \in S} P(x, y)\mathbb{P}_y(E).$$

Thus we can see that the function h satisfies

$$h(x) = \sum_{y \in S} P(x, y)h(y) \qquad \forall x \in S.$$

Thus we conclude that the function h is harmonic. $\square$

5. Poisson Processes

5.1 Solved Problems

■ **Problem 5.1** Smith enters a bank with two tellers who just started processing Allen and Yang. Assume that the processing time depends on the special kind of online authentication whose waiting time is a random variable that has a exponential distribution with some constant λ that is constant for all costumers. What is the probability that Smith will leave the bank the last?

Solution Smith should wait until either Allen or Yang is finished. After either is processed (without loss of generality, assume Allen got served and Yang is still waiting), then one of the tellers will give start serving Smith. Let T be a random variable showing the waiting time of Smith and S be a random variable representing the waiting time of Yang, both of which are i.i.d. random variables with exponential distribution. First note that since exponential random variables are memory less, then we have

$$\mathbb{P}(S > s + t | S > s) = \mathbb{P}(S > t).$$

We can formulate the Smith leaving the first as

$$\mathbb{P}(T < S).$$

By the law of total probabilities we have

$$\mathbb{P}(T < S) = \int_{-\infty}^{\infty} \mathbb{P}(T < s) f(s)\ ds = \int_{0}^{\infty} (1 - e^{\lambda s}) \lambda e^{-\lambda s}\ ds = 1 - \frac{1}{2} = \frac{1}{2}.$$