Lecture Notes*
CSCI 688 – Analysis of Stochastic Networks

Daniela Hurtado-Lange
College of William and Mary

Spring 2022

*This note is based on: Introduction to Probability Models, Sheldon Ross. Academic Press (2014).

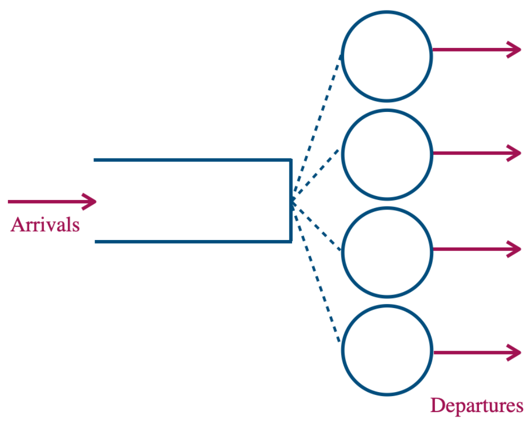
Contents

1 Introduction

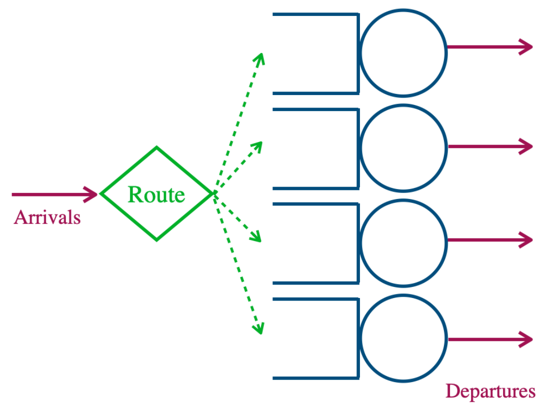
In this course we will learn how to analyze and compare performance measures of stochastic processing networks (SPN). An SPN is a system that receives job requests and has servers that complete them. Typically, the interarrival and processing times are random variables (thus the name stochastic). The most simple example of an SPN is a single-server queue. In a single-server queue, jobs arrive and they wait in line until the only server can process them. After being processed, they leave the system. In the diagram of Figure 1, the open rectangle represents the waiting area and the circle represents the server.

Figure 1: A single-server queue.

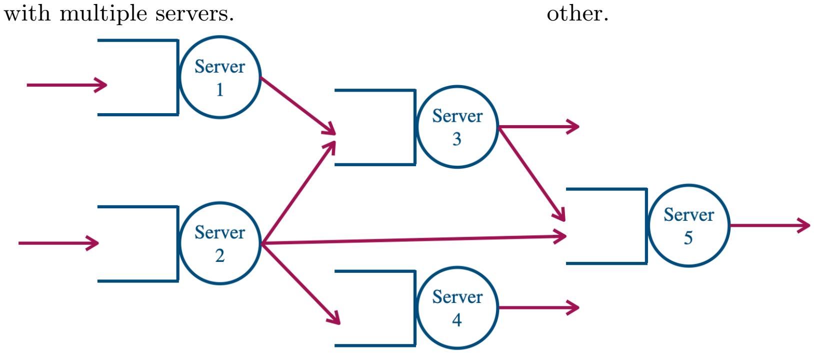
More complex systems involve a single queue with multiple servers, multiple single-server queues that depend on each other, and networks of queues. Diagrams of such systems are shown on Figure 2.



(a) A single queue with multiple servers.



(b) Multiple single queues that depend on each other.



(c) A network of single-server queues.

Figure 2: Examples of queues.

In Figure 2a there is a single stream of arrivals, and all the jobs wait in the same line. There are four servers, and each of them takes the next job in line whenever is idle. In Figure 2b there are four queues, and there is a single stream of arrivals. Upon arrival, the jobs are routed to one of the queues, where they wait in line until the corresponding server processes them. The routing decision can be made by the jobs or a centralized router. This system is also known as the supermarket-checkout system. In Figure 2c we have multiple single-server queues. Some of them receive external jobs, and some of them receive the jobs that another server finished processing. For example, Server 1 receives external arrivals and, whenever a job is completed, it goes to Server 3. The arrivals to Server 3 come from Servers 1 and 2, and after being processed, some jobs leave the system and some join the queue of Server 5. I invite you to think of situations where you have been in the waiting room of each of these examples.

Some classical settings where one uses SPN's analysis are healthcare systems, manufacturing, data centers, and call centers.

Our ultimate goal this semester is to analyze the waiting time of these systems, and many more. We are interested in performance measures such as mean queue length, mean waiting time, their variance and distribution. Understanding these measures will allow us to answer questions such as, how many servers we need to always keep a queue of less than 2 people? What server is the bottleneck? What can we do to decrease the waiting times?

It is very hard to model the inter-arrival and service times with a single function if we want to keep the analysis realistic. Hence, we adopt stochastic (i.e. probabilistic) models for the arrivals and departure processes. In other words, the inter-arrival and service times are random variables. Due to the probabilistic behavior of the arrivals and departures, the queue lengths and waiting times are complex objects. To analyze them properly, we will start the semester reviewing some basic concepts on probability and then we will spend some time learning about some models that can be used to represent queues. Towards the second half of the semester we will be in good shape to analyze the SPNs described above, and compare our theoretical analysis with simulations.

2 Review of Probability

[This section is based on Chapters 1, 2 and 3 of the textbook]

Since we are dealing with random variables, we will start with a quick review of probability. We start with the most basic definitions, and we will build our way up to conditional probability and expectation.

2.1 Basic definitions

Dealing with randomness means that we do not know certainly what outcome we will get from an experiment. However, most of the time we know the set of possible outcomes.

Definition 2.1. *Consider any probabilistic experiment.*

(i) The set of all possible outcomes of an experiment is called sample space, and we denote it with $\mathcal{S}$.

(ii) Any subset of the sample space $\mathcal{S}$ is known as an event. We usually denote events by capital letters from the first half of the alphabet.

For example, if we are flipping a coin, the state space is $\mathcal{S} = \{\text{Heads}, \ \text{Tails}\}$ and an event could be $E = \{\text{Heads}\}$, i.e., that we get heads after flipping it.

Now, if we are flipping two coins, we need to consider both coins when we define the sample space. Then, using H for heads and T for tails, we obtain $\mathcal{S} = \{HH, HT, TH, TT\}$. The events can be described as an explicit subset of $\mathcal{S}$, such as $E = \{HT, TH\}$ or in words. For example, we can say that E is the event where the outcomes of the coins are different. They are equivalent ways to describe events. In the next example, we present an alternative way to define the sample space.

Example 2.1. *Suppose you are rolling 2 dice, and you want to see if the outcomes are the same. Define the sample space and the event described.*

Solution. The sample space is the set of two-dimensional vectors/arrays where each element is an integer number from 1 to 6. Listing all the possibilities can be boring, so let's describe this set as follows:

$$\mathcal{S} = \{ij : \ i, j \in \{1, 2, 3, 4, 5, 6\}\}.$$

Observe that the sample space has $6^2 = 36$ elements, and each of them has the same probability, which equals 1/36.

The event "all the outcomes are the same" can be described as follows:

$$E = \{ij \in \mathcal{S} : \ i = j\} = \{11, 22, 33, 44, 55, 66\}.$$

□

Intuitively, we already know that the probability of an event E corresponds to the proportion of time E occurs if we repeat the experiment over and over again. Below we give a more formal definition, that captures the same intuition.

Definition 2.2. *For each event E of the sample space $\mathcal{S}$, consider a function $\mathbb{P}[E]$ that satisfies the following properties:*

(i) $0 \le \mathbb{P}[E] \le 1$

(ii) $\mathbb{P}[\mathcal{S}] = 1$

(iii) For any sequence of events $E_1, E_2, \ldots$ that are mutually exclusive, that is, that satisfy $E_i \cap E_j = \emptyset$ for any $i \neq j$, we have

$$\mathbb{P}\left[\bigcup_{i=1}^{\infty} E_i\right] = \sum_{i=1}^{\infty} \mathbb{P}[E_i]$$

We refer to $\mathbb{P}[E]$ as the probability of the event E.

We will now list some important properties of probability without proof.

Proposition 2.1. *Consider two events $E, F \subset \mathcal{S}$. Then,*

(i) $\mathbb{P}[\emptyset] = 0$

(ii) Let E^c be the complement of E, i.e., $E^c = \mathcal{S} \setminus E$. Then, $\mathbb{P}[E^c] = 1 - \mathbb{P}[E]$

(iii) $\mathbb{P}[E \cup F] = \mathbb{P}[E] + \mathbb{P}[F] - \mathbb{P}[E \cap F]$

Actually, there are known formulas for the probability of the union of finitely many events, but the expression is not nice to see, so we will skip it. You can find it in Eq. 1.4 of the textbook.

Sometimes, computing the probability of an event is very difficult. However, the computation might be simpler if we have access to some partial information, or to some information that is related to the event. To make use of this extra information, we condition. For example, suppose we are rolling a two dice and we are interested in the probability that the sum of the outcomes is 6. Call A the event that the sum is 6, B_i the event that the outcome of the first die is i and C_i the event that the outcome of the second die is i, for $i \in \{1, 2, 3, 4, 5, 6\}$. Then, it is easy to see that

$$\begin{aligned}\mathbb{P}[A] &= \mathbb{P}[B_1 \cap C_5] + \mathbb{P}[B_2 \cap C_4] + \mathbb{P}[B_3 \cap C_3] + \mathbb{P}[B_4 \cap C_2] + \mathbb{P}[B_5 \cap C_1] \\ &= \frac{5}{36} \qquad \text{(since all the events above have probability 1/36)}\end{aligned}$$

Suppose now that we already rolled the first die and we got a 4. What is the probability that the sum of both outcomes is 6? In this case, since we already know that the outcome of the first die is 4, we have that $\mathbb{P}[A'] = \mathbb{P}[C_2] = 1/6$, where A' denotes that the sum is 6 given that the first die gave a 4.

The difference between both experiments is the amount of certain information that we have. In the first case, we did not know anything about the outcome of the dice, whereas in the second case we already know the outcome of the first die. The formal way to deal with these situations is with conditional probability, which is one of the most important we will see in this class. We start with the definition.

Definition 2.3. *Consider the events $E, F \subseteq \mathcal{S}$, where $\mathbb{P}[F] \neq 0$. Then, the probability of E conditioned on F, denoted by $\mathbb{P}[E|F]$, is defined as*

$$\mathbb{P}[E|F] = \frac{\mathbb{P}[E \cap F]}{\mathbb{P}[F]}.$$

We may also say that $\mathbb{P}[E|F]$ is the probability of E given F.

Using this definition in the second case of the example above, we have that if the outcome of the first die is 4, the probability of getting the sum of the outcomes equal 6 is

$$\mathbb{P}[A|B_4] = \frac{\mathbb{P}[A \cap B_4]}{\mathbb{P}[B_4]} = \frac{\mathbb{P}[B_4 \cap C_2]}{\mathbb{P}[B_4]} = \frac{1/36}{1/6} = \frac{1}{6}.$$

Before we go to the next topic, let's do some examples.

Example 2.2. *[Example 1.7 from the textbook] Suppose an urn contains 7 black balls and 5 while balls. We draw 2 balls from the urn without replacement. Assuming that each ball in the urn is equally likely to be drawn, what is the probability that both balls are black?*

Solution. Since we are drawing the balls without replacement, the color of the first ball influences the probability of the second ball being black. Hence, we condition. Let A, B be the events of the first and second ball being black, respectively. Then, we want to compute $\mathbb{P}[A \cap B]$. We use conditional probability as follows:

$$\mathbb{P}[A \cap B] = \mathbb{P}[A]\,\mathbb{P}[B|A] = \frac{7}{12} \times \frac{6}{11} = \frac{42}{132}.$$

□

Example 2.3 (Example 1.8 from the textbook). *Suppose that each of 3 people at a party throws their hat into the center of the room. The hats are first mixed up and then each person randomly selects a hat. What is the probability that none of them gets their own hat?*

Solution. In this example we see that, sometimes, it is easier to compute the probability of the complement of the desired event. None of them getting their hat is very difficult because we need to condition on each of them getting a hat and making sure that the last hat is not owned by the last person to pick. Nonetheless, the complementary event to "none of them getting their hat" is "at least one of them gets their hat", and it is easier to deal with. Let A_i be the event that the i^{th} person gets their own hat. Then, we do

$$\mathbb{P}\left[\text{none of them getting their hat}\right] = 1 - \mathbb{P}\left[A_1 \cup A_2 \cup A_3\right].$$

Using the definition of the probability of the union of two events and grouping $A_1 \cup A_2$ as a single event, we obtain

$$\begin{aligned}
\mathbb{P}\left[A_1 \cup A_2 \cup A_3\right] &= \mathbb{P}\left[(A_1 \cup A_2) \cup A_3\right] \\
&= \mathbb{P}\left[A_1 \cup A_2\right] + \mathbb{P}\left[A_3\right] - \mathbb{P}\left[(A_1 \cup A_2) \cap A_3\right] \\
&\overset{(a)}{=} \mathbb{P}\left[A_1\right] + \mathbb{P}\left[A_2\right] - \mathbb{P}\left[A_1 \cap A_2\right] + \mathbb{P}\left[A_3\right] - \mathbb{P}\left[(A_1 \cap A_3) \cup (A_2 \cap A_3)\right] \\
&\overset{(b)}{=} \mathbb{P}\left[A_1\right] + \mathbb{P}\left[A_2\right] - \mathbb{P}\left[A_1 \cap A_2\right] + \mathbb{P}\left[A_3\right] - \left(\mathbb{P}\left[A_1 \cap A_3\right] + \mathbb{P}\left[A_2 \cap A_3\right] - \mathbb{P}\left[A_1 \cap A_2 \cap A_3\right]\right) \\
&\overset{(c)}{=} \mathbb{P}\left[A_1\right] + \mathbb{P}\left[A_2\right] + \mathbb{P}\left[A_3\right] - \mathbb{P}\left[A_1 \cap A_2\right] - \mathbb{P}\left[A_1 \cap A_3\right] - \mathbb{P}\left[A_2 \cap A_3\right] + \mathbb{P}\left[A_1 \cap A_2 \cap A_3\right]
\end{aligned}$$

where (a) holds after expanding $\mathbb{P}\left[A_1 \cup A_2\right]$ and using the distribution of intersection and union on the last term;(b) holds after using the probability of the union formula on the last term; and (c) holds after reorganizing terms. We now compute each of the terms.

Since each of the three people are equal, and the first person selects each hat with the same probability, we obtain that for each $i \in \{1, 2, 3\}$

$$\mathbb{P}\left[A_i\right] = \frac{1}{3}.$$

To compute the intersections we use the definition of conditional probability. Again, since the three people are indistinguishable, we obtain that for any pair $i, j \in \{1, 2, 3\}$ with $i \neq j$ we have

$$\mathbb{P}\left[A_i \cap A_j\right] = \mathbb{P}\left[A_i\right] \mathbb{P}\left[A_j | A_i\right] = \frac{1}{3} \times \frac{1}{2} = \frac{1}{6},$$

where the $1/2$ comes because the condition on A_i implies that the first person who picked a hat got theirs. Then, out of the two remaining hats, one corresponds to person j and one does not. Since they pick hats at random, the probability of the second person getting the right hat given that the first person did, is $1/2$.

Similarly to the intersection of two events, we compute the intersection of the three events using conditional probability. In this case, we obtain

$$\mathbb{P}\left[A_1 \cap A_2 \cap A_3\right] = \mathbb{P}\left[A_1 \cap A_2\right] \mathbb{P}\left[A_3 | A_1 \cap A_2\right] = \frac{1}{6} \times 1 = \frac{1}{6},$$

where we know that $\mathbb{P}\left[A_3 | A_1 \cap A_2\right] = 1$ because there are only 3 hats. Then, if the first two people got their own hat, the only remaining hat is the one corresponding to the third person.

Then,

$$\mathbb{P}\left[A_1 \cup A_2 \cup A_3\right] = 3 \times \frac{1}{3} - 3 \times \frac{1}{6} + \frac{1}{6} = \frac{4}{6} = \frac{2}{3}.$$

Hence, the probability that nobody gets their hat is

$$\mathbb{P}\left[\text{none of them getting their hat}\right] = \frac{1}{3}$$

□

We saw two examples where we directly applied the definition of conditional probability to facilitate computing the probability of certain events. Let's get back to our example of rolling two dice and consider a small variation. If now we know that the sum of the outcomes is 6, what is the probability that the outcome of the first die was a 4? Observe that, now, we know that the sum is 6 and the random variable is the outcome of the first die, whereas before we knew the outcome of the first die and the random variable was the sum of both outcomes. In some sense, we flipped the situation. For these cases, we use the following proposition, known as Bayes' formula.

Proposition 2.2. *[Bayes' formula] Let $E, F \subset \mathcal{S}$ be events with $\mathbb{P}[E] \neq 0$ and $\mathbb{P}[F] \neq 0$. Then,*

$$\mathbb{P}[F|E] = \frac{\mathbb{P}[E|F]\,\mathbb{P}[F]}{\mathbb{P}[E]}$$

Observe that Bayes' formula requires that we know how to compute $\mathbb{P}[E]$. We may know its value from the context, or we may need to condition on another event. In the second case, we use the law of total probability, which we state in the next proposition.

Proposition 2.3. *Suppose $F_1, F_2, \ldots, F_n$ are mutually exclusive events, that is, for every $i \neq j$ we have $F_i \cap F_j = \emptyset$. Additionally, $\bigcup_{i=1}^{n} F_i = \mathcal{S}$. In other words, exactly one of the events $F_1, F_2, \ldots, F_n$ will occur. Then,*

$$\mathbb{P}[E] = \sum_{i=1}^{n} \mathbb{P}[E \cap F_i] = \sum_{i=1}^{n} \mathbb{P}[E|F_i]\,\mathbb{P}[F_i].$$

Let's get back to our example. Using both propositions we can compute the probability that the outcome of the first die was a 4 given that the sum was 6? Recall that B_4 represents that the outcome of the first die is 4, and A represents that the sum of both outcomes is 6. Then, we have

$$\begin{aligned}
\mathbb{P}[B_4|A] &= \frac{\mathbb{P}[A|B_4]\,\mathbb{P}[B_4]}{\mathbb{P}[A]} && \text{(using Bayes' formula)} \\
&= \frac{1/6 \times 1/6}{5/36} && \text{(using our previous computations)} \\
&= \frac{1}{5}.
\end{aligned}$$

In this case, a simple application of Bayes' formula gave us the answer because we already knew the value of $\mathbb{P}[A]$ from before. Let's see an example where we compute the denominator from Bayes' formula using the law of total probability.

Example 2.4 (Example 1.13 from the textbook)**.** *In answering a question on a multiple-choice test, Eric either knows the answer or guesses. Let p be the probability that he knows the answer and $1 - p$ the probability that he guesses. Assume that a student who guesses at the answer will be correct with probability $1/m$, where m is the number of multiple-choice alternatives. What is the conditional probability that Eric knew the answer to a question given that he answered it correctly?*

Solution. In this example we also use the "flipping" intuition behind Bayes' formula because we know the probability of getting the right answer given that we know the answer (which is 1), and we are asked to compute the probability that we knew given that we have the right answer.

Let K be the event that Eric knew the answer and C the event that he answered correctly. Then, we obtain

$$\begin{aligned}
\mathbb{P}[K|C] &= \frac{\mathbb{P}[C|K]\,\mathbb{P}[K]}{\mathbb{P}[C]} && \text{(using Bayes' formula)} \\
&= \frac{1 \times p}{\mathbb{P}[C]} && \text{(since he knows the answer with probability } p) \\
&= \frac{p}{\mathbb{P}[C|K]\,\mathbb{P}[K] + \mathbb{P}[C|K^c]\,\mathbb{P}[K^c]} && \text{(using the law of total probability)} \\
&= \frac{p}{1 \times p + (1/m) \times (1-p)} \\
&= \frac{mp}{1 + (m-1)p}.
\end{aligned}$$

□

Bayes' formula and the law of total probability are extremely important results because they help us compute probabilities using information from other events. In other words, they help us to deal with randomness by adding some partial information that we may have. Additionally, they help us to deal with events that involve multiple random variables, as we shall see later in the semester.

The last concept we will review before moving on to studying random variables is independence, which we define below.

Definition 2.4. *Two events E and F are said to be* <u>*independent*</u> *if*

$$\mathbb{P}[E \cap F] = \mathbb{P}[E]\,\mathbb{P}[F].$$

Equivalently, we can say that two events are independent if

$$\mathbb{P}[E|F] = \mathbb{P}[E].$$

Intuitively, the definition of independence means that one event does not affect the other one. From the definition using conditional probability (second definition above), we intuitively can say that the occurrence of the event F does not affect the probability of E.

Let's go back to our example of rolling two dice and computing the probability of the sum of the outcomes being 6. Recall that the probability that the sum of the outcomes being 6 changes if we know that the first outcome is 4. Then, intuitively, the two events should not be independent. Let's see what happens with the first definition of independence above.

Recall that A is the event that the sum of the outcomes is 6 and B_4 is the event that the outcome of the first die is 4. Then,

$$\mathbb{P}[A \cap B_4] = \frac{1}{36}$$

and

$$\mathbb{P}[A]\,\mathbb{P}[B_4] = \frac{5}{36} \times \frac{1}{6} \neq \frac{1}{36}.$$

Hence, A and B_4 are not independent events.

Now let's see a small variation of the problem. Let A_7 be the event that the sum of the outcomes of the dice is 7. Is A_7 independent of B_4? Let's see! We have

$$\mathbb{P}[A_7 \cap B_4] = \frac{1}{36}$$

because $A_7 \cap B_4$ implies that the first outcome is 4 and the second one is 3. Then, only 1 out of the 36 possibilities for the outcome of both dice is a success.

To compute the product of the probability of both events we start computing $\mathbb{P}[A_7]$. Recall that B_i, C_i are the events that the outcome of the first and second die is i, respectively. Then,

$$\begin{aligned}
\mathbb{P}[A_7] &= \sum_{i=1}^{6} \mathbb{P}[B_i \cap C_{7-i}] \\
&= \sum_{i=1}^{6} \mathbb{P}[B_i]\,\mathbb{P}[C_{7-i}] \qquad \text{(since the outcome of each die is independent of the other one)} \\
&= \sum_{i=1}^{6} \frac{1}{6} \times \frac{1}{6} \\
&= \frac{1}{6}.
\end{aligned}$$

Then,

$$\mathbb{P}[A_7]\,\mathbb{P}[B_4] = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36} = \mathbb{P}[A_7 \cap B_4].$$

Hence, the events are independent! Why? This result is very counter-intuitive! The main difference between these cases is that when the sum of the outcomes is 6, none of the outcomes can be 6. In other words, if the first die shows a 6 we know that we cannot get a success. However, when the sum has to be 7, all the outcomes of the first die can lead to a success.

2.2 Random variables and well-known distributions

Let's start with a definition of random variable.

Definition 2.5. *A random variable is a real-valued function of the outcomes of an experiment. We typically use capital letters from the last third of the alphabet to denote random variables, and lower case letters for their values.*

Then, we assign probabilities to their possible values and we can make tons of computations with them. An essential property of random variables is that the sum of the probabilities of all of their outcomes must always be 1. Let's see some examples.

Example 2.5 (Example 2.3 from the textbook)**.** *Suppose that we repeteadly toss a coin having probability p of coming up heads until the first head appears. Let N be the number of flips required assuming that the outcome of successive flips are independent. Then, N is a random variable taking on one of the values $1, 2, 3, \ldots$, i.e., positive integer values.*

The event $\{N = n\}$ means that the random variable N takes the value n, and for different values of n has the following probability:

$$\begin{aligned}
&\mathbb{P}[N=1] = p \qquad \textit{(heads on the first flip)} \\
&\mathbb{P}[N=2] = (1-p)p \qquad \textit{(first flip tails and second, heads)} \\
&\mathbb{P}[N=3] = (1-p)^2 p \qquad \textit{(first two flips tails, and last one, heads)} \\
&\qquad\vdots \\
&\mathbb{P}[N=n] = (1-p)^{n-1} p \qquad \textit{(first } n-1 \textit{ flips tails, and last one, heads)}
\end{aligned}$$

Observe that

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathbb{P}[N=n] &= \sum_{i=1}^{\infty} (1-p)^{n-1} p \\
&= \frac{p}{1-(1-p)} \qquad \textit{(solving the geometric series)} \\
&= 1.
\end{aligned}$$

Example 2.6 (Example 2.4 from the textbook)**.** *Suppose that we are interested in knowing whether or not a battery will last two more years. We can define a random variable I as follows:*

$$I = \begin{cases} 1 & \textit{if the battery lasts at least two more years} \\ 0 & \textit{otherwise.} \end{cases}$$

Then, we may compute $\mathbb{P}[I = 1]$. If we use E to denote the event that the battery lasts two more years, the random variable I is known as indicator function of E and we will denote it as $\mathbb{1}_{\{E\}}$.

All the examples we have studied include random variables taking nonnegative integer values. However, in general, random variables can take any real number. Depending on the possible values that random variables take, we can classify them in two.

Definition 2.6. *Consider a random variable X.*

(a) If X takes values on a finite or countable set, such as $\mathbb{Z}$ (the set of integer numbers) or a subset, we say that X is a discrete random variable.

(b) If X takes values on a continuum of possible values, such as $\mathbb{R}$, we say that X is a continuous random variable.

For example, the variable I defined above, which represents whether a battery lasts for at least two more years, is a discrete random variable. The actual lifetime of the battery is a continuous random variables.

For both, discrete and continuous random variables, we can define the cumulative distribution function.

Definition 2.7. *The cumulative distribution function (cdf) of a random variable X is defined as a function $F : \mathbb{R} \to [0, 1]$ and for any number $x \in \mathbb{R}$ takes the value*

$$F(x) \triangleq \mathbb{P}[X \leq x],$$

and denotes the probability that the random variable X takes a value that is less than or equal to x.

We list some properties of the cdf below.

Proposition 2.4. *For any value $x \in \mathbb{R}$ the cdf satisfies the following properties:*

(i) $F(x)$ *is nondecreasing with respect to* x

(ii) $\lim_{x\to\infty} F(x) = F(\infty) = 1$

(iii) $\lim_{x\to-\infty} F(x) = F(-\infty) = 0$

(iv) For any $a, b \in \mathbb{R}$ *with* $a \leq b$ *we have*

$$\mathbb{P}[a \leq X \leq b] = F(b) - F(a)$$

The cdf gives us the probability that $X \leq x$ or $X \in [a, b]$, but how can we compute the probability that a random variable takes a specific value x? In other words, how do we compute $\mathbb{P}[X = x]$? The truth is that the event $\{X = x\}$ only makes sense for discrete random variables. If X is a continuous random variable, we always have $\mathbb{P}[X = x] = 0$. For example, if we are thinking of the lifetime of a battery, what is the probability that it exactly lasts 2 years? Not 2 years and one second; exactly 2 years. It is actually 0! Even though we cannot compute $\mathbb{P}[X = x]$ for continuous random variables, we can still associate a function that tells us the probability that the random variable takes a value in a small interval $(x - \epsilon, x + \epsilon)$.

The discussion above motivates the following definitions.

Definition 2.8. *Let* X *be a random variable with cdf* $F(x)$.

(a) If X *is a discrete random variable, the probability mass function (pmf) of* X *is* $p(x) \triangleq \mathbb{P}[X = x]$, *and if* X *takes values on the countable set* $\mathcal{X}$ *we have*

$$F(x) = \sum_{y \in \mathcal{X}: y \leq x} p(x)$$

(b) For continuous random variables, there exists a nonnegative function $f : \mathbb{R} \to [0, 1]$ *known as the probability density function (pdf), having the property that for any set* $B \subset \mathbb{R}$,

$$\mathbb{P}[X \in B] = \int_B f(x)\, dx.$$

The relationship between the pdf and cdf of a continuous random variable is expressed by

$$F(x) = \int_{-\infty}^{x} f(t)\, dt \qquad \& \qquad f(x) = \frac{d}{dx} F(x).$$

Random variables are often described by their distribution. Later we list some of the most frequently used distributions.

2.3 Expected value, variance and covariance

Both the cdf and pdf/pmf have all the information about the random variable. If we know any of these functions we can completely characterize our random variable. However, sometimes the distribution is difficult to interpret. Instead, we would like to know some "summarizing information", such as the average behavior or the variability. Below we define the expected value and variance, which provide this information.

Definition 2.9. *The expected value, or mean, of a random variable* X *is denoted by* $\mathbb{E}[X]$ *and is defined as follows.*

(a) If X *is a discrete random variable taking values on the countable set* $\mathcal{X}$ *with pmf* $p(x)$, *we have*

$$\mathbb{E}[X] \triangleq \sum_{x \in \mathcal{X}} x p(x).$$

(b) If X is a continuous random variable with pdf $f(x)$, we have

$$\mathbb{E}[X] \triangleq \int_{\mathbb{R}} x f(x)\, dx.$$

Sometimes, the pdf/pmf of a random variable has a complicated expression, and the cdf is easier to compute. In such cases, we may compute the expected value using the following proposition.

Proposition 2.5. *Let X be a random variable with cdf $F(x)$. Then,*

(a) If X is a discrete random variable taking values on $\mathcal{X}$, we have

$$\mathbb{E}[X] = \sum_{x\in\mathcal{X}\cap\mathbb{R}_+} \big[1 - F(x)\big] - \sum_{x\in\mathcal{X}\cap\mathbb{R}_-} F(x)$$

(b) If X is a continuous random variable, then

$$\mathbb{E}[X] = \int_0^{\infty} \big[1 - F(x)\big]\, dx - \int_{-\infty}^{0} F(x)\, dx.$$

A very cool property of the expected value is that it is a linear function. Below we state this property formally.

Proposition 2.6. *Consider two numbers (not random) $a, b \in \mathbb{R}$. Then,*

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b.$$

Linearity of expectation is a very simple property, but it is extremely useful. Let's solve an example to show it.

Example 2.7 (Example 2.31 from the textbook)**.** *Suppose there are 25 types of coupons and suppose that each time one obtains a coupon, it is equally likely to be any of the 25 types. Compute the expected number of different types of coupons that are contained in a set of 10 coupons.*

Solution. Let X denote the number of different types of coupons in the set of 10 coupons. We want to compute $\mathbb{E}[X]$. To do that, define

$$X_i = \begin{cases} 1 & \text{if at least one type } i \text{ coupon is in the set of 10} \\ 0 & \text{if not} \end{cases}$$

In other words, X_i is the indicator function of the coupon i being present in the set of 10. Then,

$$X = \sum_{i=1}^{25} X_i.$$

Therefore,

$$\mathbb{E}[X] = \sum_{i=1}^{25} \mathbb{E}[X_i].$$

We compute $\mathbb{E}[X_i]$. Since X_i is a discrete random variable that takes 0-1 values, we have

$$\begin{aligned} \mathbb{E}[X_i] &= \mathbb{P}[X_i = 1] \\ &= 1 - \mathbb{P}[X_i = 0] \\ &\overset{(*)}{=} 1 - \left(\frac{24}{25}\right)^{10}, \end{aligned}$$

where $(*)$ holds because $X_i = 0$ if none of the 10 coupons are of type i. Since there are 25 coupons and every coupon has the same probability to appear, the probability that the first coupon is not type i is $24/25$. Similarly for the second, third, ..., up to the tenth coupon. Since the coupons are independent, we obtain that the probability that coupon i did not show up in the 10 coupons selected is $(24/25)^{10}$. □

We may also be interested in the expected value of a function of the random variable X, say $g(X)$. Since $g(X)$ is also a random variable, we may compute its cdf or pmf/pdf and then compute $\mathbb{E}[g(X)]$ using the definition or Proposition 2.5. Luckily, there is an easier way to compute $\mathbb{E}[g(X)]$.

The property described in Proposition 2.7 is called law of the unconscious statistician (LOTUS), and it is extremely useful. So keep it in mind (and in your cheat sheet!).

Proposition 2.7. *Consider a random variable X and a real-valued function $g(X)$ defined on the set of possible outcomes of X.*

(a) If X is a discrete random variable taking values on $\mathcal{X}$, with pmf $p(x)$, we have

$$\mathbb{E}[g(X)] = \sum_{x \in \mathcal{X}} g(x)p(x).$$

(b) If X is a continuous random variable with pdf $f(x)$, then

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x)f(x)\,dx.$$

A very important function of the random variables to consider is $g(X) = (X - \mathbb{E}[X])$ because it measures the deviation from the mean of the random variable X. The expected value of this particular function of X is called the variance, and we formally define it below.

Definition 2.10. *Consider a random variable X. The variance of X is defined as*

$$Var[X] \triangleq \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2.$$

Even though the variance is not a linear function of the random variable, it is still easy to deal with the variance of linear combinations of the random variable.

Proposition 2.8. *Let $a, b \in \mathbb{R}$. Then,*

$$Var[aX + b] = a^2\, Var[X].$$

To remember this proposition, recall that the variance is related to the square of the random variable (hence, we pull out an a^2) and to the deviation of the random variable with respect to the mean. Hence, since constants do not deviate from their mean ($\mathbb{E}[b] = b$), the variance of a constant is 0.

Some discrete distributions

Name	Example	Support	pmf $p(x)$	cdf $F(x)$	Mean	Variance
Bernoulli(p)	Success/Failure experiment	$\{0,1\}$	$p(1) = p$ (success) and $p(0) = 1-p$ (failure)	$\begin{cases} 0 & \text{if } x < 0 \\ 1-p & \text{if } 0 \le x < 1 \\ 1 & \text{if } x \ge 1. \end{cases}$	p	$p(1-p)$
Binomial(n,p)	Number of successes in n experiments	$\{0,1,\ldots,n\}$	$\binom{n}{x}p^x(1-p)^{n-x}$	Difficult expression	np	$np(1-p)$
Geometric(p)	Number of failures (or trials) until first success	$\{0,1,2,\ldots\}$ $\{1,2,3,\ldots\}$	$(1-p)^x p$ $(1-p)^{x-1}p$	$1-(1-p)^{x+1}$ $1-(1-p)^x$	$\frac{1-p}{p}$ $\frac{1}{p}$	$\frac{1-p}{p^2}$
Poisson(λ)	Number of arrivals in an interval of time	$\{0,1,2,\ldots\}$	$\frac{e^{-\lambda}\lambda^x}{x!}$	Difficult expression	λ	λ

Some continuous distributions

Name	Example	Support	pdf $f(x)$	cdf $F(x)$	Mean	Variance
Uniform(a,b)	Any symmetric experiment	$[a,b]$	$f(x)=\frac{1}{b-a}$ for $x\in[a,b]$ and $f(x)=0$ otherwise	$\begin{cases}0 & \text{if } x<a\\ \frac{x-a}{b-a} & \text{if } a\le x\le b\\ 1 & \text{if } x>b\end{cases}$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$
Exponential(λ)	Lifetime of a bulb	$\mathbb{R}_+$	$\lambda e^{-\lambda x}$	$1-e^{-\lambda x}$	$1/\lambda$	$1/\lambda^2$
Normal(μ,σ^2)	Average of independent random variables (Central Limit Theorem)	$\mathbb{R}$	$\dfrac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$	Difficult expression. There are tables.	μ	σ^2

The last topic we review before going a bit deeper into conditioning, is multiple random variables. There are only a few cases where we deal with a single random variable. Most of the time, two or more random variables interact and we need to analyze them. In such case, we define the pdf/pmf and cdf of both variables together, and we call them the joint pdf/pmf or joint cdf, respectively. The extension of the definition is immediate, so we skip it. Instead, we focus on the covariance of the random variables, which indicates how related (or dependent) the random variables are from each other.

Definition 2.11. *The covariance of any pair of random variables X, Y is defined as*

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

The covariance is a bilinear function, and satisfies the following properties.

Proposition 2.9. *Consider the random variables X, Y, Z, and the constants $a, b \in \mathbb{R}$. Then,*

(i) $Cov[X,X] = Var[X]$

(ii) $Cov[X,Y] = Cov[Y,X]$

(iii) $Cov[aX,Y] = a\,Cov[X,Y]$

(iv) $Cov[X,Y+Z] = Cov[X,Y] + Cov[X,Z]$

(v) $Var[aX+bY] = a^2\,Var[X] + b^2\,Var[Y] + 2ab\,Cov[X,Y]$

(vi) Properties (iv) and (v) can be also concluded from the following property. Let $X_1,\dots,X_n$ and $Y_1,\dots,Y_m$ be random variables. Then,

$$Cov\left[\sum_{i=1}^{n} X_i,\ \sum_{j=1}^{m} Y_j\right] = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov[X_i, Y_j]$$

Considering the properties listed above, we may think of the covariance as some sort of "dot product" between random variables and the variance as some sort of "norm". This is just an intuitive explanation to remember the properties, because the random variables X and Y are not vectors.

Example 2.8 (Example 2.34 from the textbook)**.** *Compute the variance of a binomial random variable X with parameters n and p.*

Solution. There are multiple ways to solve this question, but let's try to use the properties of variance and covariance stated in Proposition 2.9. Recall that a binomial random variable is the sum of n independent Bernoulli random variables. Formally, for $i \in \{1,\dots,n\}$, let X_i be a Bernoulli random variable, i.e.

$$X_i = \begin{cases} 1 & \text{if we get a success} \\ 0 & \text{otherwise} \end{cases}$$

Then,

$$X = \sum_{i=1}^{n} X_i.$$

We obtain

$$\begin{aligned}
\text{Var}[X] &= \text{Cov}\left[\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\right] \qquad \text{(definition of } X_i\text{'s and property (i))} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \text{Cov}[X_i, X_j] \qquad \text{(property (vi))} \\
&= \sum_{i=1}^{n} \text{Cov}[X_i, X_i] + \sum_{i=1}^{n}\sum_{j=1,\, j\neq i}^{n} \text{Cov}[X_i, X_j] \\
&= \sum_{i=1}^{n} \text{Var}[X_i] + \sum_{i=1}^{n}\sum_{j=1,\, j\neq i}^{n} 0 \qquad \text{(property (i) and since } X_i\text{'s are independent)} \\
&= \sum_{i=1}^{n} p(1-p) \qquad \text{(since the variance of a Bernoulli random variable is } p(1-p)) \\
&= np(1-p)
\end{aligned}$$

□

2.4 Conditioning on random variables

Recall that we can define events with random variables. For example, if X is a discrete random variable, $\{X = 10\}$ is an event. Then, we can use Bayes' rule and the law of total probability to make computations easier. Let's do an example.

Example 2.9. *Suppose that we have a printer that is not working very well. Every time it prints a sheet, there is a probability p that it is well done and $1-p$ that it's not. When the printout is not well done, we discard it and try again. Since the machine is not working very well, the time it takes to finalize printing a sheet is exponentially distributed with rate μ. The sheets printed by the machine are independent from each other (in terms of the time and success/failure to do it well). What is the distribution of the time until you get a well-printed sheet?*

Solution. First notice that, if we forget for a minute about the time the printer takes to print, and we only look at the number of printed sheets until we get a good one, we have a geometric random variable with parameter p. Let N denote this number. Then, if we use X_i to denote the time it takes to print the i^{th} sheet, notice that the time until we get a well-printed sheet (call it X) is

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

Here we need to resist the temptation to say that, since X is the sum of independent exponential random variables, it has Gamma distribution. The problem is that we are adding up a random number of exponential random variables. Then, we need to compute the distribution.

Since we are dealing with multiple random variables, we condition as follows:

$$\begin{aligned}
\mathbb{P}\left[\sum_{i=1}^{N} X_i \le x\right] &= \sum_{n=1}^{\infty} \mathbb{P}\left[\sum_{i=1}^{N} X_i \le x \,\middle|\, N = n\right] \mathbb{P}[N = n] \qquad \text{(law of total probability)} \\
&= \sum_{n=1}^{\infty} \mathbb{P}\left[\sum_{i=1}^{n} X_i \le x\right] \mathbb{P}[N = n] \qquad \text{(since } X_i\text{'s are } N \text{ are independent)} \\
&\overset{(*)}{=} \sum_{n=1}^{\infty} \left(\int_0^x \frac{\mu^n}{(n-1)!} t^{n-1} e^{-\mu t}\, dt\right) (1-p)^{n-1} p \qquad \text{(replacing the Gamma cdf and Geometric pmf)}
\end{aligned}$$

$$
\begin{aligned}
&= \mu p \int_0^x e^{-\mu t} \left(\sum_{n=1}^{\infty} \frac{(\mu t p)^{n-1}}{(n-1)!} \right) dt \qquad \text{(reorganizing terms)} \\
&= \mu p \int_0^x e^{-\mu t} e^{\mu t p} \, dt \qquad \text{(solving the exponential series)} \\
&= \mu p \int_0^x e^{-\mu p t}, \, dt \\
&= \left[e^{-\mu p t} \right]_{t=0}^{t=x} \\
&= 1 - e^{-\mu p x}.
\end{aligned}
$$

In step $(*)$ observe that we have a finite sum of exponential random variables. When we conditioned on $N = n$, we got rid of the randomness at the upper limit of the sum. Hence, we do have a Gamma distribution.

The last expression is exactly the cdf of an exponential random variable with rate μp. Hence, we will have to wait an exponential time with rate μp until the printer gives us a good sheet. Observe that the distribution of the time until getting the job done did not change, but the rate decreased. Intuitively, since the printer can make mistakes, it will take a bit longer to finish a well-done job. □

2.5 Independence of random variables

Now that we can deal with multiple random variables, we may wonder if they are independent. Before we discussed about independence of events, and we can expand the same concept to random variables. Recall that $\{X \leq x\}$ is also an event. Then, we have the following definition for independence of random variables.

Definition 2.12. *The random variables X and Y are independent if for all constants x, y we have*

$$\mathbb{P}[X \leq x, \, Y \leq y] = \mathbb{P}[X \leq x] \, \mathbb{P}[Y \leq y] \, .$$

Observe that our definition of independence involves the joint cdf and the individual cdf's. Indeed, the expression above can be restated as

$$F_{X,Y}(x, y) = F_X(x) F_Y(y).$$

We can actually use the pdf/pmf to equivalently define independence, as shown in the following proposition.

Proposition 2.10. *The random variables X and Y are independent if and only if for all x, y we have*

(i) $p_{X,Y}(x, y) = p_X(x) p_Y(y)$ *if X and Y are discrete random variables*

(ii) $f_{X,Y}(x, y) = f_X(x) f_Y(y)$ *if X and Y are continuous random variables*

We can also use the expected value and the notion of covariance to talk about independence. However, in this case we don't have an "if and only if" relationship. If X and Y are independent random variables and $g(X)$ and $h(Y)$ are real-valued functions, then

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

However, $\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)] \, \mathbb{E}[h(Y)]$ does not necessarily imply that X and Y are independent. In other words, independence implies uncorrelated (covariance=0), but uncorrelated does not imply independence. To see this counter-intuitive property, let's see the following example.

Example 2.10. *Let X be a discrete random variable with $\mathbb{P}[X = 1] = \mathbb{P}[X = 0] = \mathbb{P}[X = -1] = 1/3$, and let $Y \triangleq X^2$. Since the random variable Y is defined in terms of the random variable X, they are clearly not independent. Let's compute their covariance. By definition,*

$$Cov[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X] \, \mathbb{E}[Y] \, ,$$

where

$$\mathbb{E}[X] = 1 \times \frac{1}{3} + 0 \times \frac{1}{3} + (-1) \times \frac{1}{3} = 0$$

$$\mathbb{E}[Y] = 1 \times \frac{2}{3} + 0 \times \frac{1}{3} = \frac{2}{3}$$
$$\mathbb{E}[XY] = 1 \times \frac{1}{3} + 0 \times \frac{1}{3} + (-1) \times \frac{1}{3} = 0 \qquad \textit{(because } XY = X\textit{)}$$

Therefore,

$$Cov[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0 - 0 \times \frac{2}{3} = 0.$$

2.6 Conditional expectation

The last topic we will cover before starting to review Markov chains is conditional expectation. We have been talking about conditioning on events and conditioning on the value of random variables, using the idea that $\{X = x\}$, for example, is an event. We will now focus on using this idea of conditioning to compute expectations. We will split the analysis in discrete and continuous random variables.

Discrete case

We already know that the pmf and the cdf of X given $Y = y$ are, respectively,

$$p_{X|Y}(x|y) \triangleq \frac{\mathbb{P}[X = x,\, Y = y]}{\mathbb{P}[Y = y]} = \frac{p_{X,Y}(x, y)}{p_Y(y)}$$
$$\text{and} \quad F_{X|Y}(x|y) \triangleq \mathbb{P}[X \le x \,|\, Y = y] = \sum_{i \le x} \mathbb{P}[X = i \,|\, Y = y].$$

The conditional expectation is defined similarly, as specified below.

Definition 2.13. *If X is a discrete random variable, the* <u>*conditional expectation of X given $Y = y$*</u> *is defined by*

$$\mathbb{E}[X \,|\, Y = y] \triangleq \sum_x x\, p_{X|Y}(x, y) = \sum_x x \mathbb{P}[X = x \,|\, Y = y]$$

Let's use this definition in an example.

Example 2.11 (Example 3.3 from the textbook)**.** *If X and Y are independent Poisson random variables with rate λ_1 and λ_2, respectively, compute the conditional expected value of X given $X + Y = n$.*

Since Poisson random variables are good models for the arrival process to certain systems, we have the following interpretation of the result above. Consider the arrival process to a pharmacy. X represents the arrivals of people who come to get a prescription, and Y of people who come to get over-the-counter medicines. Let's assume that nobody wants to get a prescription and over-the-counter medicines. Then, we want to compute the expected number of people who have arrived to the pharmacy, given that n people entered the pharmacy. Now let's solve the problem.

Solution. Let's first compute the conditional pmf of X given $X + Y = n$. By definition, we obtain

$$\begin{aligned}
\mathbb{P}[X = k \,|\, X + Y = n] &= \frac{\mathbb{P}[X = k,\, X + Y = n]}{\mathbb{P}[X + Y = n]} \\
&= \frac{\mathbb{P}[X = k,\, Y = n - k]}{\mathbb{P}[X + Y = n]} \\
&= \frac{\mathbb{P}[X = k]\,\mathbb{P}[Y = n - k]}{\mathbb{P}[X + Y = n]} \qquad \text{(since } X \text{ and } Y \text{ are independent)}
\end{aligned}$$

You may have forgotten by now, but let me remind you that the sum of independent Poisson random variables is also Poisson and the rate is simply the sum of the rates of the original random variables. Then, $X + Y$ is a Poisson random variable with rate $\lambda_1 + \lambda_2$. Hence, we obtain

$$\mathbb{P}[X = k \,|\, X + Y = n] = \frac{\mathbb{P}[X = k]\mathbb{P}[Y = n - k]}{\mathbb{P}[X + Y = n]}$$

$$
\begin{aligned}
&= \left(\frac{\lambda_1^k e^{-\lambda_1}}{k!}\right)\left(\frac{\lambda_2^{n-k} e^{-\lambda_2}}{(n-k)!}\right)\left(\frac{n!}{(\lambda_1+\lambda_2)^n e^{-(\lambda_1+\lambda_2)}}\right) \\
&= \frac{n!}{k!(n-k)!}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k} \qquad \text{(canceling } e^{-(\lambda_1+\lambda_2)} \text{ and rearranging terms)} \\
&= \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(1-\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{n-k}.
\end{aligned}
$$

Then, the conditional distribution of X given $X+Y=n$ is Binomial with parameters n and $p=\frac{\lambda_1}{\lambda_1+\lambda_2}$. Then, the expected value of X given $Y=y$ is

$$
\mathbb{E}\left[X \mid X+Y=n\right] = np = \frac{n\lambda_1}{\lambda_1+\lambda_2}
$$

□

From the result, observe that given that we have n people in the pharmacy, we can interpret that getting a prescription is a success, and getting an over-the-counter medicine is not. Then, the number of people that get a prescription is equivalent to obtaining k successes in n independent trials. Even though X and $X+Y$ are not independent random variables, they behave as if each of the n people in the pharmacy were independent. This is one of the nice properties of Poisson random variables.

Continuous case

For continuous random variables we replace the conditional pmf by the conditional pdf, as follows. Recall that for any y such that $f_Y(y)>0$,

$$
f_{X|Y}(x|y) \triangleq \frac{f_{X,Y}(x,y)}{f_Y(y)}.
$$

Then, we obtain the following definition of expectation conditioned on $Y=y$.

Definition 2.14. *If X and Y are continuous random variables, the* <u>*conditional expectation of X given $Y=y$*</u> *is defined for values of y such that $f_Y(y)>0$ as*

$$
\mathbb{E}\left[X \mid Y=y\right] \triangleq \int_{\mathbb{R}} x f_{X|Y}(x|y)\,dx
$$

Before discussing some properties of the conditional expectation, let's do a continuous example.

Example 2.12 (Example 3.7 from the textbook)**.** *The joint density of X and Y is given by*

$$
f_{X,Y}(x,y) = \begin{cases} \frac{1}{2}ye^{-xy} & , 0<x<\infty,\ 0<y<2 \\ 0 & \textit{otherwise} \end{cases}
$$

Compute $\mathbb{E}\left[e^{\frac{X}{2}} \mid Y=1\right]$.

Solution. We first compute the conditional density of X given $Y=1$. We have

$$
\begin{aligned}
f_{X|Y}(x,1) &= \frac{f_{X,Y}(x,1)}{f_Y(1)} \\
&= \frac{\frac{1}{2}e^{-x}}{\int_0^\infty f_{X,Y}(t,1)\,dt} \\
&= e^{-x}
\end{aligned}
$$

Now, since the expectation conditional on $Y=y$ is an expected value, it satisfies LOTUS. Hence,

$$
\mathbb{E}\left[e^{\frac{X}{2}} \mid Y=1\right] = \int_0^\infty e^{\frac{x}{2}} f_{X|Y}(x,1)\,dx
$$

$$= \int_0^\infty e^{\frac{x}{2}} e^{-x}\, dx$$
$$= 2.$$

□

Let's get back to any random variable X (discrete or continuous).

The first observation we want to make is that $\mathbb{E}[X \mid Y = y]$ is an expected value at the end of the day and, therefore, it satisfies **linearity** and **LOTUS**.

The last topic we will cover is how to use conditional expectation to compute expectations. To do that, let's observe that $\mathbb{E}[X \mid Y = y]$ is a function of y. Let's call it $g(y) \overset{\triangle}{=} \mathbb{E}[X \mid Y = y]$. That said, we can generalize $\mathbb{E}[X \mid Y = y]$ to a general value of Y and define the random variable $g(Y) \overset{\triangle}{=} \mathbb{E}[X \mid Y]$. Then, the random variable $g(Y) = \mathbb{E}[X \mid Y]$ takes the value $g(y) = \mathbb{E}[X \mid Y = y]$ if the random variable $Y = y$.

In the previous sections, we were using the law of total probability to condition on the value of a random variable and obtain the pmf and cdf of another random variable. The analogous to this law in terms of the expectation is the **tower property**, that we state below.

Proposition 2.11 (Tower property)**.** *For any random variables X and Y,*

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]].$$

For discrete random variables, the tower property translates to

$$\mathbb{E}[X] = \sum_y \mathbb{E}[X \mid Y = y]\, \mathbb{P}[Y = y]$$

and for continuous random variables, it translates to

$$\mathbb{E}[X] = \int_{\mathbb{R}} \mathbb{E}[X \mid Y = y]\ f_Y(y)\, dy$$

Example 2.13 (Example 3.10 from the textbook)**.** *Suppose that the expected number of accidents per week at an industrial plant is 4. Suppose also that the number of workers injured in each accident are independent random variables with a common mean 2. Assume also that the number of workers injured in each accident is independent of the number of accidents that occur. What is the expected number of injuries during a week?*

Solution. Observe that there are two random variables playing here: the number of accidents and the number of injuries in each accident. Let N be the number of accidents in a week, and X_i be the number of injuries in the i^{th} accident. Then, the total number of injuries in a week X is

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

Observe that, exactly as in the printers example, we are adding up a random number of random variables.

Then, we compute $\mathbb{E}[X]$ as follows.

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}\left[\sum_{i=1}^{N} X_i\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^{N} X_i \,\middle|\, N\right]\right] \qquad \text{(tower property)} \\
&= \mathbb{E}[N\mathbb{E}[X_1 \mid N]] \qquad \text{(by definition of conditional expectation and since the } X_i\text{'s have the same mean)} \\
&= \mathbb{E}[N\mathbb{E}[X_1]] \qquad \text{(since } N \text{ and } X_i\text{'s are independent)} \\
&= \mathbb{E}[N]\,\mathbb{E}[X_1] \qquad \text{(by linearity of expectation because } \mathbb{E}[X_1] \text{ is a constant)} \\
&= 4 \times 2 = 8
\end{aligned}$$

□

We have used conditioning to compute probabilities and expected values. We can also use it to compute variances. To do it, we use the law of total variance, that we present below.

Proposition 2.12 (Proposition 3.1 from the textbook)**.** *Consider two random variables X and Y. Then,*

$$Var[X] = \mathbb{E}[Var[X|Y]] + Var[\mathbb{E}[X|Y]]$$

Let's end the review of probability chapter with an example. There are many many more examples in the book.

Example 2.14 (Chapter 3, Exercise 97 from the textbook)**.** *Use the law of total variance to compute the variance of a geometric random variable.*

Solution. Let X be a geometric random variable with parameter p. For simplicity, let's assume that X represents the number of **trials** until the first success. We want to use the conditional variance formula, so we need to find another random variable Y to condition on. Since a geometric random variable involves **independent** trials of a success/failure experiment, let's split the possible values of X based on the first outcome of the experiment. Then, define Y as the indicator function that the first trial is a success. Then,

$$\mathbb{P}[Y = 1] = p \qquad \text{and} \qquad \mathbb{P}[Y = 0] = 1 - p.$$

We first compute $\mathbb{E}[X|Y]$ and $\text{Var}[X|Y]$ for the possible values of Y. If $Y = 1$, the variable X can only take the value 1. In other words, $X|Y = 1$ is deterministic. Therefore,

$$\mathbb{E}[X|Y = 1] = 1 \qquad \text{and} \qquad \text{Var}[X|Y = 1] = 0.$$

Now, if $Y = 0$, it means that the first trial was a failure and, hence, we need to keep trying. Since different trials are independent and have the same probability of success, starting to count trials after the first failure is probabilistically the same as counting the total number of trials. Therefore, the random variables $X|Y = 0$ and $1 + X$ have the same distribution and, hence,

$$\mathbb{E}[X|Y = 0] = 1 + \mathbb{E}[X] = 1 + \frac{1}{p} \qquad \text{and} \qquad \text{Var}[X|Y = 0] = \text{Var}[X].$$

Then, we obtain

$$\begin{aligned} \mathbb{E}[\text{Var}[X|Y]] &= \text{Var}[X|Y = 1]\,\mathbb{P}[Y = 1] + \text{Var}[X|Y = 0]\,\mathbb{P}[Y = 0] \\ &= 0 \times p + \text{Var}[X](1 - p) \\ &= (1 - p)\text{Var}[X]. \end{aligned} \tag{1}$$

Now we compute $\text{Var}[\mathbb{E}[X|Y]]$. First observe that

$$\mathbb{E}[X|Y] = \begin{cases} 1 & \text{if } Y = 1 \\ 1 + 1/p & \text{if } Y = 0 \end{cases}$$

Then, we can write

$$\mathbb{E}[X|Y] = 1 + \frac{1}{p}(1 - Y).$$

Then, using properties of variance, we have

$$\begin{aligned} \text{Var}[\mathbb{E}[X|Y]] &= \text{Var}\left[1 + \frac{1}{p}(1 - Y)\right] \\ &= \text{Var}[1] + \text{Var}\left[\frac{1}{p}\right] + \text{Var}\left[\frac{-1}{p}Y\right] \qquad \text{(since constants are independent of any random variable)} \\ &= \frac{1}{p^2}\text{Var}[Y] \qquad \text{(since variance of a constant is 0 and } \text{Var}[aX] = a^2\text{Var}[X]) \\ &= \frac{1}{p^2}(p(1 - p)) \qquad \text{(since } Y \text{ is a Bernoulli random variable with parameter } p) \\ &= \frac{1 - p}{p}. \end{aligned} \tag{2}$$

Finally, using Equation (1) and Equation (2) in the law of total variance, we obtain

$$\mathrm{Var}\left[X\right] = (1-p)\mathrm{Var}\left[X\right] + \frac{1-p}{p}$$

and rearranging terms, we obtain

$$\mathrm{Var}\left[X\right] = \frac{1-p}{p^2}.$$

Note: If we wanted to do the number of failures until the first success the process is equivalent. All we need to modify is that $\mathbb{E}\left[X|Y=1\right] = 0$. However, we are interested in computing $\mathrm{Var}\left[\mathbb{E}\left[X|Y\right]\right]$ and 0 is a constant, so the variance is not affected. □

In probability and stochastic processes, a very popular technique is conditioning on the first outcome. In the example above we saw that, by doing that, we obtain a recursive equation because consecutive experiments are independent. Then, we simply computed the variance by solving the equation.

An important observation above is that we computed the variance of X using only limited information about its pmf. Observe that we only used the expected value of X and the probability that X takes the value 1.

3 Discrete-Time Markov Chains (DTMC)

[This section is based on Chapter 4 of the textbook]

3.1 Definition and main properties of DTMC

We have been discussing random variables, and the possible outcomes they may have. However, life is dynamic and we usually observe multiple outcomes of a family of random variables. Then, we also need to consider time when we analyze random events. We start with a definition that will allow us to think of a sequence of random variables, where the index of the sequence is usually associated with time.

Definition 3.1 (Section 2.9 of the textbook)**.** *A stochastic process $\{X_t : t \in T\}$ is a collection of random variables indexed by t. That is, for each $t \in T$, X_t is a random variable. The index t often represents time, so we refer to X_t as the state of the process at time t. The indices t take values in the set T, which we call index set, and it is usually $\mathbb{Z}_+$ (discrete-time process) or $\mathbb{R}_+$ (continuous-time processes).*

The state space of a stochastic process is the set of all possible outcomes of the random variables X_t for all values of $t \in T$. We usually denote it by $\mathcal{X}$.

Examples of stochastic processes and their state spaces are:

- Number of students in the classroom, $\mathcal{X} = \mathbb{Z}_+$
- Number of customers who enter a store, $\mathcal{X} = \mathbb{Z}_+$
- Number of customers waiting in line, $\mathcal{X} = \mathbb{Z}_+$
- Number of cars that cross a specific intersection, $\mathcal{X} = \mathbb{Z}_+$
- Total amount of sales at a store, $\mathcal{X} = \mathbb{R}_+$
- Stock price of Amazon, $\mathcal{X} = \mathbb{R}_+$
- Position of a particle with respect to a fixed point, $\mathcal{X} = \mathbb{R}^3$

In this chapter we are interested in developing a probability model for stochastic processes. Using the knowledge we already have, we can always describe a sequence of random variables $X_1, X_2, \ldots, X_n$ as a stochastic process. We just need to consider $T = \{1, \ldots, n\}$, and the sequence $\{X_n : n \in T\}$ is clearly a stochastic process. However, we have only studied sequences of independent random variables. In this chapter, we develop probability models that allow dependence among the elements of the sequence.

Definition 3.2. *A stochastic process $\{X_n : n \in \mathbb{Z}_+\}$ over a countable state space $\mathcal{X}$ is a Discrete-Time Markov Chain (DTMC) if for any pair $i, j \in \mathcal{X}$ there exists a constant p_{ij} such that*

$$p_{ij} = \mathbb{P}\left[X_{n+1} = j \,|\, X_n = i,\, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, X_0 = i_0\right]$$

for any choice of $i_0, i_1, \ldots, i_{n-1} \in \mathcal{X}$ and all $n \in \mathbb{Z}_+$.

The property defining DTMC in Definition 3.2 can be interpreted in English as:

> The outcome of the future state X_{n+1} only depends on the past states $X_0, X_1, \ldots, X_n$ through the outcome of the present state X_n.

Observe that we are not only absorbing in X_n the information about the **outcomes** of the past states, but we are also absorbing the information about how "big" is the history. In other words, p_{ij} only depends on i and j, and it does not depend on $i_0, \ldots, i_{n-1}$ and n. In English, we are saying that in DTMCs,

> The future depends on the past only through the present.

Let's see some examples.

Example 3.1. *[Example 4.2 from the textbook] Consider a communication system that transmits the digits 0 and 1. Each digit transmitted must pass through several stages, at each of which there is a probability p that the digit entered will be unchanged when it leaves.*

Let X_n be the digit entering the n^{th} stage. Then, $\{X_n : n \in \mathbb{Z}_+\}$ is a DTMC.

Example 3.2. *[Example 4.4 from the textbook] 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.*

Define a DTMC that reflects the situation described above.

Solution. Observe that if we define X_n as 1 if it rained on day n and 0 if not, the process $\{X_n : n \in \mathbb{Z}_+\}$ is **not** a DTMC because X_{n+1} depends on X_n and X_{n-1}.

Instead, we define states that represent whether it rained or not in two consecutive days, as follows.

State	Yesterday	Today
0	Rain	Rain
1	Rain	Dry
2	Dry	Rain
3	Dry	Dry

□

If we go back to the definition of a DTMC, observe that all the probabilistic information about the sequence $\{X_n : n \in \mathbb{Z}_+\}$ is in the values of p_{ij} for $i, j \in \mathcal{X}$. We start studying these numbers with the following key observations:

(1) The value p_{ij} is the probability that, starting from state i, the sequence goes to state j. Then, since p_{ij} is a probability, and the sequence must go somewhere, we have the following properties:

$$p_{ij} \geq 0 \qquad \forall i, j \in \mathcal{X}$$

$$\text{and} \quad \sum_{j \in \mathcal{X}} p_{ij} = 1 \qquad \forall i \in \mathcal{X}$$

(2) The numbers p_{ij} can be written in a square matrix, that we call transition probability matrix. If the state space is $\mathcal{X} = \{1, \dots, n\}$, we have

$$P = \begin{bmatrix} p_{00} & p_{01} & \cdots & \cdots & p_{0n} \\ p_{10} & p_{11} & \cdots & \cdots & p_{1n} \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ p_{n0} & p_{n1} & \cdots & \cdots & p_{nn} \end{bmatrix}.$$

To build the matrix P, we think of the rows as the current state, and the columns as the future state.

Let's see some examples.

Example 3.3. *Compute the probability transition matrix for the communication system from Example 3.1.*

Solution. First observe that the state space in this case is $\mathcal{X} = \{0, 1\}$. Then, the transition matrix is of dimension 2×2 and has entries:

$$P = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}$$

□

Example 3.4. *Compute the transition probability matrix of Example 3.2.*

Solution. To construct the probability transition matrix, observe that we have some transitions that do not make physical sense because the "future state" includes days tomorrow and today, and the "present state" includes today and yesterday. For example, starting from state 0 we can only go to state 2 because state 0 says that today is raining and states 1 and 3 say that today is not raining. With this in mind, the transition probability matrix is

$$P = \begin{bmatrix} 0.7 & 0.3 & 0 & 0 \\ 0 & 0 & 0.4 & 0.6 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.2 & 0.8 \end{bmatrix}$$

□

Example 3.5 (Inventory (s, S) problem)**.** *Suppose you want to sell muffins at school. You have done some surveys, and you are pretty much sure of the pmf of the demand you will have. Let d_i be the probability that the demand on a given day is i, for $i \in \mathbb{Z}_+$.*

Considering the size of your bag, the maximum inventory you can carry is S and, to avoid cooking every day, you have decided that you will only make more muffins if you arrive home with s or less at the end of the day. When this happens, you make enough muffins to go to school with an inventory of S the next morning.

Suppose $S = 5$ and $s = 2$, and assume that the demand is independent of the inventory you have. Model the situation described as a Markov chain and construct the transition probability matrix.

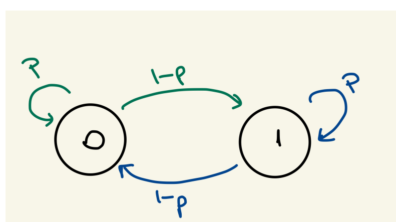
Solution. Let X_n be the inventory of muffins at the beginning of day n. Then, the state space is $\mathcal{X} = \{3, 4, 5\}$ because you replenish your inventory overnight whenever you arrive home with 0, 1 or 2 muffins.

Then, we obtain the following transition matrix

$$P = \begin{bmatrix} d_0 & 0 & \sum_{i\geq 1} d_i \\ d_1 & d_0 & \sum_{i\geq 2} d_i \\ d_2 & d_1 & d_0 + \sum_{i\geq 3} d_i \end{bmatrix}$$

□

The transition probability matrix has all the information that we need to study a DTMC. However, sometimes it is helpful to have a visual representation of the chain. The standard representation is a graph where the nodes are the states and the edges represent the nonzero elements of the transition probability matrix. For example, for Example 3.1, the graph is given by



where the green arrows represent the first row of the transition probability matrix, and the blue arrow represent the second row.

Before we move on let's see an example inspired by stochastic networks.

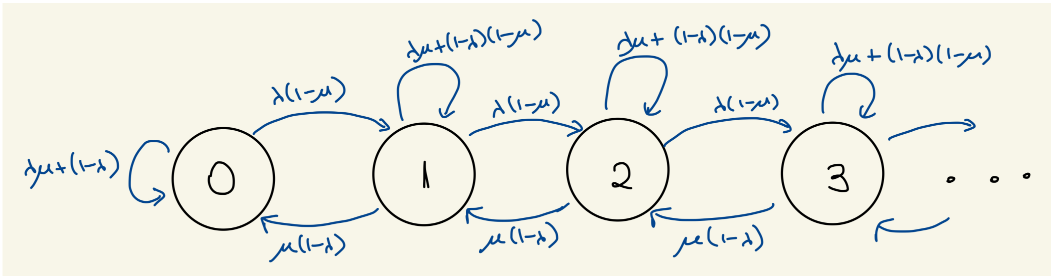
Example 3.6. *Consider a single-server queue modeled in discrete time, that is, we observe the system at the end of pre-determined time slots of, say, one second. The number of arrivals to the system in one time slot follows a Bernoulli distribution with parameter λ, and, when the system is not empty, the number of potential departures follows a Bernoulli distribution with parameter μ. Assume that the number of arrivals per time slot is independent of the number of processed jobs, and these two are independent of the state of the system.*

(a) Let N_k be the number of jobs in the system (considering the job in service, if any) at the beginning of time slot k. Is $\{N_k : k \in \mathbb{Z}_+\}$ a DTMC? If yes, compute the transition probability matrix and draw the diagram.

(b) Let Q_k be the number of jobs in the queue (without considering the job in service, if any) at the beginning of time slot k. Is $\{Q_k : k \in \mathbb{Z}_+\}$ a DTMC? If yes, compute the transition probability matrix and draw the diagram.

Solution.

(a) Yes, $\{N_k : k \in \mathbb{Z}_+\}$ is a DTMC. The diagram is:



The transition probability matrix is defined by

$$p_{0,0} = \lambda\mu + (1-\lambda)$$
$$p_{0,1} = \lambda(1-\mu)$$

and for $i \geq 1$ we have

$$p_{i,i} = \lambda\mu + (1-\lambda)(1-\mu)$$
$$p_{i,i+1} = \lambda(1-\mu)$$
$$p_{i,i-1} = \mu(1-\lambda)$$

and $p_{i,j} = 0$ otherwise.

(b) No. When $Q_k = 0$ we actually do not know the probability to stay in 0 and go to state 1 because $Q_k = 0$ means that either $N_k = 0$ or $N_k = 1$. In other words, if we do not look at the server (to see if it's busy or not) we are losing too much information and we obtain a stochastic process that is no longer a Markov chain.

□

3.2 Chapman-Kolmogorov equations

The matrix P gives us the one-step transition probabilities. But how can we compute the multiple-step transition probabilities? For example, if we have a single-server queue that has 2 customers, then what is the pmf of the number of customers in 5 time slots?

For $i, j \in \mathcal{X}$ and $n \in \mathbb{Z}_+$, define the n-step transition probabilities as

$$p_{i,j}^{(n)} = \mathbb{P}\left[X_{n+k} = j \,|\, X_k = i\right].$$

To compute these quantities, we use the Chapman-Kolmogorov equations, which establish that

$$p_{i,j}^{(n+m)} = \sum_{\ell \in \mathcal{X}} p_{i,\ell}^{(n)} p_{\ell,j}^{(m)} \qquad \forall n, m \in \mathbb{Z}_+,\ i, j \in \mathcal{X}.$$

In words, if we want to compute the probability of going from state i to j in $n+m$ steps, we can condition on where the chain goes in the first n steps. If it goes to state ℓ, then it must go from ℓ to j in m steps. Formally,

$$\begin{aligned}
p_{i,j}^{(n+m)} &= \mathbb{P}\left[X_{n+m} = j \,|\, X_0 = i\right] \\
&= \sum_{\ell \in \mathcal{X}} \mathbb{P}\left[X_{n+m} = j \,|\, X_n = \ell,\ X_0 = i\right] \mathbb{P}\left[X_n = \ell \,|\, X_0 = i\right] \\
&= \sum_{\ell \in \mathcal{X}} \mathbb{P}\left[X_{n+m} = j \,|\, X_n = \ell\right] \mathbb{P}\left[X_n = \ell \,|\, X_0 = i\right] \qquad \text{(because } \{X_k\}_k \text{ is a DTMC)} \\
&= \sum_{\ell \in \mathcal{X}} \mathbb{P}\left[X_m = j \,|\, X_0 = \ell\right] \mathbb{P}\left[X_n = \ell \,|\, X_0 = i\right]
\end{aligned}$$

$$= \sum_{\ell \in \mathcal{X}} p_{\ell,j}^{(m)} p_{i,\ell}^{(n)}$$

In the next theorem, we confirm the intuition we get from Chapman-Kolmogorov equations.

Theorem 3.1. *Consider a DTMC with transition probability matrix P and let $P^{(n)}$ denote the n-step transition probability matrix. Then,*

$$P^{(n)} = P^n,$$

i.e., the n-step transition probability matrix is exactly the one-step transition probability matrix in the power of n.

Example 3.7. *Consider again Example 3.2.*

(a) If it rained on Monday and Tuesday, what is the probability that it will rain on Thursday?

(b) Now suppose we know it rained on Tuesday, but we don't remember if it rained on Monday so we assume that it rained with probability 0.5. What is the probability that it will rain on Thursday?

Solution. In both cases, we need to compute the two-step transition probability matrix. We obtain

$$P^2 = \begin{bmatrix} 0.7 & 0.3 & 0 & 0 \\ 0 & 0 & 0.4 & 0.6 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.2 & 0.8 \end{bmatrix} \begin{bmatrix} 0.7 & 0.3 & 0 & 0 \\ 0 & 0 & 0.4 & 0.6 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.2 & 0.8 \end{bmatrix} = \begin{bmatrix} 0.49 & 0.21 & 0.12 & 0.18 \\ 0.20 & 0.20 & 0.12 & 0.48 \\ 0.35 & 0.15 & 0.20 & 0.30 \\ 0.10 & 0.10 & 0.16 & 0.64 \end{bmatrix}$$

(a) If it rained on Monday and Tuesday, our initial state is 0, so we look at the first row of P^2. Now, rain on Thursday means that we are either in state 0 or 2, so we have

$$\begin{aligned} &\mathbb{P}\left[\text{Rain on Thursday} \,|\, \text{Rain on Monday and Tuesday}\right] \\ &= \mathbb{P}\left[X_2 \in \{0,1\} \,|\, X_0 = 0\right] \\ &= \mathbb{P}\left[X_2 = 0 \,|\, X_0 = 0\right] + \mathbb{P}\left[X_2 = 2 \,|\, X_0 = 0\right] \qquad \text{(using the formula for probability of the union)} \\ &= p_{0,0}^{(2)} + p_{0,2}^{(2)} \\ &= 0.49 + 0.12 \\ &= 0.61 \end{aligned}$$

(b) Now we don't know the initial state, but we have a pmf for it. Let $\alpha_i = \mathbb{P}\left[X_0 = i\right]$ for $i \in \{0,1,2,3\}$. We can write the pmf as a vector as follows:

$$\begin{aligned} \boldsymbol{\alpha} &= \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} \mathbb{P}\left[X_0 = 0\right] \\ \mathbb{P}\left[X_0 = 1\right] \\ \mathbb{P}\left[X_0 = 2\right] \\ \mathbb{P}\left[X_0 = 3\right] \end{pmatrix} \\ &= \begin{pmatrix} 0.5 \\ 0.5 \\ 0 \\ 0 \end{pmatrix} \end{aligned}$$

Then, using the intuition built above, we have that the pmf of the state in two days is

$$\begin{aligned} \boldsymbol{\alpha}^{(2)} &= (P^2)^T \boldsymbol{\alpha} \\ &= \begin{bmatrix} 0.49 & 0.21 & 0.12 & 0.18 \\ 0.20 & 0.20 & 0.12 & 0.48 \\ 0.35 & 0.15 & 0.20 & 0.30 \\ 0.10 & 0.10 & 0.16 & 0.64 \end{bmatrix} \begin{pmatrix} 0.5 \\ 0 \\ 0.5 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 0.42 \\ 0.18 \\ 0.16 \\ 0.24 \end{pmatrix} \end{aligned}$$

Then, the probability that it rains on Thursday is $\alpha_0^{(2)} + \alpha_2^{(2)} = 0.42 + 0.16 = 0.58$

□

To use the matrix multiplication method, let's remember that the i^{th} row of P^k represents the pmf of the state in step k given that $X_0 = i$. Then, if the initial state has pmf $\boldsymbol{\alpha}^{(0)}$, then by law of total probability we have that

$$\mathbb{P}[X_k = j] = \sum_{i \in \mathcal{X}} \mathbb{P}[X_k = j | X_0 = i] \, \alpha_i^{(0)} \qquad \forall j \in \mathcal{X}$$

If we rewrite these equations in matrix notation, we obtain that the probability mass function of the k^{th} step, denoted by $\boldsymbol{\alpha}^{(k)}$ is given by

$$\boldsymbol{\alpha}^{(k)} = (P^k)^T \boldsymbol{\alpha}^{(0)}$$

or, equivalently,

$$(\boldsymbol{\alpha}^{(k)})^T = (\boldsymbol{\alpha}^{(0)})^T P^k.$$

3.3 Classification of states

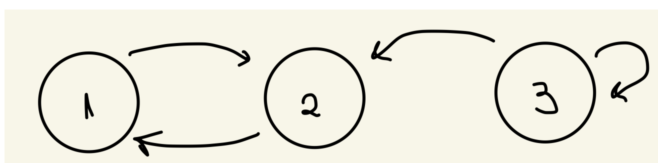
So far we have know that the behavior of the DTMC's after a finite number of steps can be determined by the transition matrix. We just need to compute the power corresponding to the number of steps we want to give. In this section, we will learn how to make sense of the transition matrix and try to predict the behavior of the DTMC without necessarily computing the powers of P.

We will define "types" of states and study properties of each type.

Definition 3.3. *State j is accessible from state i if there exists $n \in \mathbb{Z}_+$ such that $p_{i,j}^{(n)} > 0$. In words, if we can get to state i given that we started from j.*

Two states i and j that are accessible to each other are said to communicate, and we write $i \leftrightarrow j$.

Example 3.8. *Consider the following DTMC.*



In this example:

- *States 1 and 2 communicate ($1 \leftrightarrow 2$)*
- *State 2 is accessible from state 3*
- *State 3 is not accessible from state 2. Then, states 2 and 3 do not communicate.*

The relation of communication is an equivalence class and, hence, it satisfies the following properties:

(1) **Reflexivity:** $i \leftrightarrow i$ for all $i \in \mathcal{X}$. In words, state i always communicates with itself.

(2) **Symmetry:** If $i \leftrightarrow j$, then $j \leftarrow i$. In words, if i communicates with j, then j communicates with i

(3) **Transitivity:** If $i \leftrightarrow j$ and $j \leftrightarrow k$, then $i \leftrightarrow k$. In words, if i communicates with j and j communicates with k, then i communicates with k.

The first two properties immediately hold by definition, and the third property is a consequence of Chapman-Kolmogorov equations. Let's prove it!

Proof. By definition of communication, we need to show that $p_{i,k}^{(n)} > 0$ for some $n \in \mathbb{Z}_+$. Indeed, since $i \leftrightarrow j$ and $j \leftrightarrow k$, we know the existence of $n_1 > 0$ and $n_2 > 0$ such that

$$p_{i,j}^{(n_1)} > 0 \qquad \text{and} \qquad p_{j,k}^{(n_2)} > 0.$$

Now, we expect that $n = n_1 + n_2$ is a good number to our proof. Let's see.

$$\begin{aligned} p_{i,k}^{(n_1+n_2)} &= \sum_{\ell \in \mathcal{X}} p_{i,\ell}^{(n_1)} p_{\ell,k}^{(n_2)} \qquad \text{(by Chapman-Kolmogorov equations)} \\ &\geq p_{i,j}^{(n_1)} p_{j,k}^{(n_2)} > 0. \end{aligned}$$

This completes the proof. □

Definition 3.4. *States that communicate are said to be in the same class.*

The DTMC is said to be irreducible if there is only one class, that is, if all the states communicate with each other.

In Example 3.8 there are two classes: $\{1, 2\}$ and $\{3\}$. Hence, it is not irreducible. Before we work on an example, let's see additional definitions.

Definition 3.5. *For any state i, let f_i be the probability that, starting from state i, the DTMC will ever reenter state i. State i is recurrent if $f_i = 1$, and state i is transient if $f_i < 1$.*

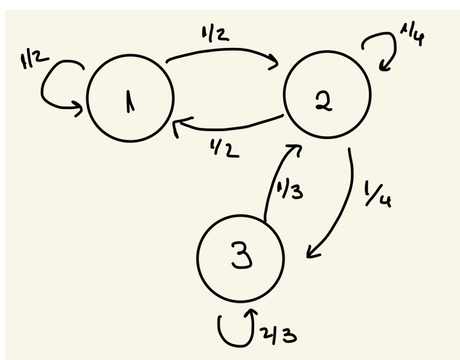
We will spend some time studying the properties of recurrent and transient states. But before we get there, one more definition to completely classify our DTMC's.

Definition 3.6. *State i is periodic with period $d > 1$ if the chain can only return to state i in a multiple of d steps. If $d = 1$ the state is said to be aperiodic.*

Example 3.9. *For the DTMC's defined by the following transition probability matrices, determine if they are irreducible. If not, determine how many classes they have. Are the states recurrent or transient? Periodic or aperiodic?*

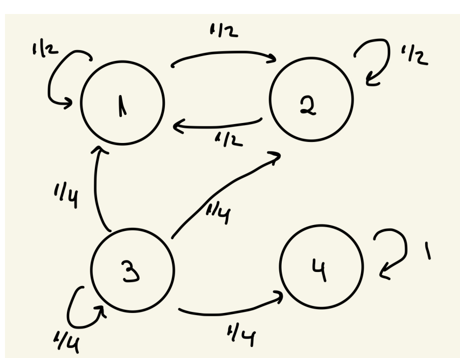
$$P = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/4 & 1/4 \\ 0 & 1/3 & 2/3 \end{bmatrix} \qquad Q = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Solution. We draw the diagrams. For matrix P we obtain



Observe that all the states communicate with each other. Then, the DTMC is irreducible. Since the chain is irreducible, the class is recurrent. Additionally, all the states are aperiodic.

For matrix Q we have the following diagram:



In this case we have three classes: $\{1,2\}$, $\{3\}$ and $\{4\}$. Class $\{1,2\}$ and $\{4\}$ are recurrent, and class $\{3\}$ is transient. Observe that state 4 is absorbing, i.e., if we enter there, we will never leave. However, all the states are aperiodic. □

Properties of recurrent and transient states

By definition of a Markov chain, every time we visit a state i we are restarting the process. Then, recurrent states are visited infinitely many times. Transient states, on the other hand, are visited a geometric number of times with parameter $p = f_i$.

Before we determined if the classes are recurrent or transient based on the diagram. However, when the state space is large we may not be able to do this. The next proposition gives an alternative way to verify.

Proposition 3.2 (Proposition 4.1 from the textbook)**.** *State i is recurrent if*

$$\sum_{n=0}^{\infty} p_{i,i}^{(n)} = \infty$$

and transient if

$$\sum_{n=0}^{\infty} p_{i,i}^{(n)} < \infty$$

Proof. To prove both results, let I_n be the indicator function that $X_n = i$. Then, the total number of times the chain visits state i is

$$\begin{aligned}\mathbb{E}\left[\sum_{n=0}^{\infty} I_n \,\middle|\, X_0 = i\right] &= \sum_{n=0}^{\infty} \mathbb{E}\left[I_n \,|\, X_0 = i\right] \\ &= \sum_{n=0}^{\infty} \mathbb{P}\left[X_n = i \,|\, X_0 = i\right] \\ &= \sum_{n=0}^{\infty} p_{i,i}^{(n)}.\end{aligned}$$

Then, since a recurrent state is visited infinitely many times and a transient state is visited finitely many times, we obtain the result. □

This representation of recurrent and transient states allows us to show the following result.

Corollary 3.3. *If $i \leftrightarrow j$ and i is recurrent, then j is also recurrent.*

Proof. We're trying to show that state j is recurrent, i.e., that

$$\sum_{n=1}^{\infty} p_{j,j}^{(n)} = \infty.$$

It suffices to find a lower bound that grows to infinity.

Since $i \leftrightarrow j$, we know that there exist some $m_1, m_2 \in \mathbb{Z}_+$ such that $p_{i,j}^{(m_1)}$ and $p_{j,i}^{(m_2)}$. Now,

$$p_{j,j}^{(m_1+m_2+n)} \geq p_{j,i}^{(m_2)} p_{i,i}^{(n)} p_{i,j}^{(m_2)}.$$

Then, by Chapman-Kolmogorov equations, we obtain

$$\begin{aligned}\sum_{\ell=1}^{\infty} p_{j,j}^{(\ell)} &\geq \sum_{n=1}^{\infty} p_{j,i}^{(m_2)} p_{i,i}^{(n)} p_{i,j}^{(m_1)} \\ &= p_{i,j}^{(m_1)} p_{j,i}^{(m_2)} \sum_{n=1}^{\infty} p_{i,i}^{(n)} = \infty.\end{aligned}$$

This shows that $\sum_{n=1}^{\infty} p_{j,j}^{(n)} = \infty$ and, hence, j is a recurrent state. □

From the last two results, we can draw some conclusions:

(i) Transience is also a class property because states that are not recurrent are transient.

(ii) If the state space is finite, then not all states are transient.
Why? Since we are studying a DTMC that can evolve forever (i.e., infinitely many steps), the only way to keep visiting a finite number of states is to visit at least one state infinitely many times. Hence, not all the states can be transient.

(iii) If we have an irreducible DTMC with finitely many states, then all of them must be recurrent.
Why? This property relates to the previous property. In an irreducible Markov chain, all the states are either transient or recurrent. Since we have finitely many states, they must be recurrent (using the previous property)

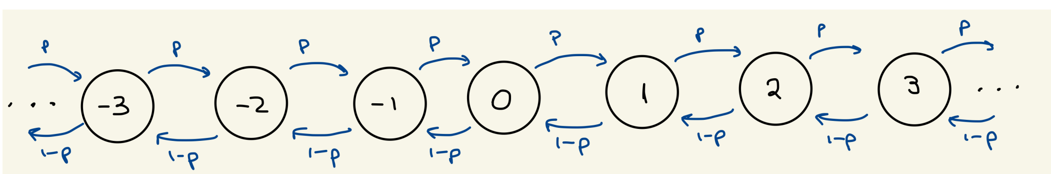
Below we study a random walk, which is one of the classical examples of Markov chains. Some applications are queueing theory (if we truncate at 0), and the earnings/losses of a gambler who bets $1 in each game. The random walk is also an example of the so-called birth-death process.

Example 3.10 (Example 4.19 from the textbook)**.** ***The random walk in 1D:*** *Consider a Markov chain where the state space is* $\mathcal{X} = \mathbb{Z}$ *and has transition probabilities:*

$$p_{i,j} = \begin{cases} p & , \text{ if } j = i+1 \\ 1-p & , \text{ if } j = i-1 \\ 0 & , \text{ otherwise} \end{cases}$$

where $p \in (0,1)$*. In other words, on each transition the chain moves one step forward or one step backwards. Find the classes of the Markov chain and classify them.*

Solution. Before adventuring into the classification of states, let's draw the diagram.



The first observation is that all the states communicate with each other. Hence, the chain is irreducible.

To decide about periodicity, let's look at state 0. Observe that if $X_0 = 0$, then $X_n = 0$ only has a positive probability if n is an even number. Then, state 0 has period 2. The period is also a class property, so all the states are periodic with period $d = 2$.

The last question to answer (and that will take us a while to answer) is whether the states are recurrent or transient. To answer this question we study

$$\sum_{n=1}^{\infty} p_{0,0}^{(n)}$$

and determine if it is finite or not.

As noted before, state 0 is periodic with period $d = 2$. Then, $p_{0,0}^{(2n-1)} = 0$ for all $n \in \mathbb{Z}_+$.

To compute $p_{0,0}^{(2n)}$ observe that if we are back to state 0 in $2n$ steps, we went n steps forward and n backwards, and the probability of doing that is $p^n(1-p)^n$. However, $p^n(1-p)^n$ represents the probability of a specific path, and there are many many possibilities. For example, if $n = 3$ and we represent a step forward by $+$ and backwards by $-$, we may do $+++---$, or $++-+--$, or $+-+-+-$, etc. Then, the probability of going from state 0 to 0 in $2n$ steps is $p^n(1-p)^n$ times the number of possible paths. One way to think of the number of paths is that, out of the $2n$ steps, we need to choose n to go forward. Hence, there are $\binom{2n}{n}$ possible paths. Putting everything together, we obtain

$$p_{0,0}^{(2n)} = \binom{2n}{n} p^n (1-p)^n$$

Therefore,

$$\begin{aligned}\sum_{n=1}^{\infty} p_{0,0}^{(n)} &= \sum_{n=1}^{\infty} p_{0,0}^{(2n)} \qquad \left(\text{since } p_{0,0}^{(2n-1)} = 0\right) \\ &= \sum_{n=1}^{\infty} \binom{2n}{n} p^n (1-p)^n \\ &= \sum_{n=1}^{\infty} \frac{(2n)!}{n!n!} p^n (1-p)^n\end{aligned}$$

We use Stirling's approximation to deal with the factorials.

Proposition 3.4. *For a positive integer number n, we have*

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

where $\sim$ represents that both quantities are asymptotically equal as $n \to \infty$.

Using Stirling's approximation, we obtain

$$\frac{(2n)!}{n!n!} \sim \frac{\sqrt{2\pi}(2n)^{2n+1/2}e^{-2n}}{\left(\sqrt{2\pi}n^{n+1/2}e^{-n}\right)^2} = \frac{2^{2n}}{\sqrt{\pi n}}$$

Therefore, we obtain

$$\sum_{n=1}^{\infty} p_{0,0}^{(2n)} \sim \sum_{n=1}^{\infty} \frac{1}{\sqrt{\pi}} \frac{(4p(1-p))^n}{\sqrt{n}}$$

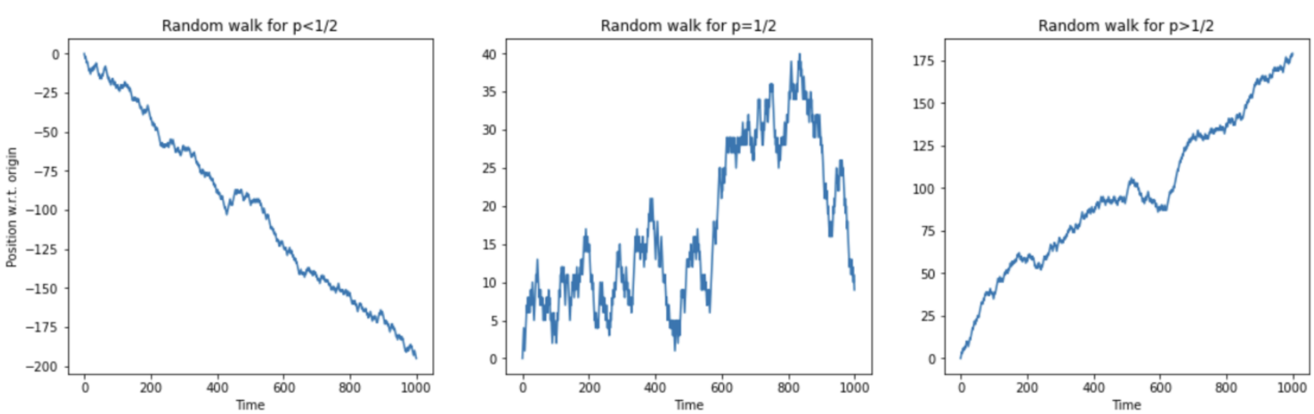
Does the sum converge? First observe that $\sum_{n=1}^{\infty} \frac{1}{\sqrt{n}}$ diverges. Then, the sum will converge if $4p(1-p) < 1$ and will diverge if $4p(1-p) \geq 1$. Let's study $4p(1-p) - 1$ and see its sign. Rearranging terms, we obtain

$$4p(1-p) - 1 = 4p - 4p^2 - 1 = -\left(4p^2 - 4p + 1\right) = -(2p-1)^2 \leq 0 \qquad \forall p$$

Hence, we obtained that the sum diverges if and only if $p = 1/2$. Interpreting the result, the random walk is a recurrent chain if and only if it's symmetric (i.e., if $p = 1/2$). If it's asymmetric, it is transient. □

A random walk being transient is a very counter-intuitive result. However, we have to remember that it is a infinite-state Markov chain. If $p > 1/2$, then we **always** have more probability of getting away from zero towards ∞. Then, after a long time of going forward with high probability, going back to state 0 becomes impossible. Similarly, if $p < 1/2$ we will probably end up in $-\infty$.

Below we present the results of a simulation.



Observe that when $p = 1/2$, the random walk stays close to the origin. However, when $p \neq 1/2$ it goes away very fast. For $p > 1/2$ it goes away to ∞ and for $p < 1/2$ to $-\infty$.

3.4 Long-run proportions and limiting probabilities

The first couple of steps of a Markov chain are usually messy to study because they highly depend on the probability mass function of the initial state (that we denoted $\boldsymbol{\alpha}^{(0)}$). Since many systems run for a long time, there is value in understanding what would happen if we wait for a long time. Will the system reach some sort of equilibrium? How would it look like?

In this section we will study the probabilistic behavior of DTMCs in the long run. We start with some notation.

Definition 3.7. *Consider a DTMC $\{X_n : n \in \mathbb{Z}_+\}$ with state space $\mathcal{X}$, and take two states $i, j \in \mathcal{X}$ with $i \neq j$. Then, the probability of ever visiting state j given that we started from i is denoted $f_{i,j}$. More formally,*

$$f_{i,j} \triangleq \mathbb{P}\left[X_n = j \text{ for some } n > 0 \,|\, X_0 = i\right]$$

The probability $f_{i,j}$ can be considered a generalization of f_i, which represented the probability of ever coming back to state i. Indeed, f_i and $f_{i,j}$ have a similar property for recurrent states in the same class, as we show in the next proposition.

Proposition 3.5 (Proposition 4.3 from the textbook)**.** *If i is recurrent and $i \leftrightarrow j$, then $f_{i,j} = 1$.*

The proof is simple, and it is based on the idea that recurrence is a class property. We will skip the details, but the main idea is that i is a recurrent state, so starting from state $X_0 = i$, we are sure that we will come back to state i infinitely many times. Additionally, $i \leftrightarrow j$. Then, there is a positive probability that j will be visited in n steps starting from i (for some finite n). Then, the number of visits to i before we visit j for the first time is a geometric random variable and, hence, we know that we will visit state j in a finite number of steps.

We will now focus on the time between visits to a state, in order to compute the proportion of time that we spend in each of them. We start with a definition.

Definition 3.8. *Define N_j as the number of transitions until the Markov chain makes a transition into state j, that is*

$$N_j \triangleq \min\{n > 0 : \; X_n = j\}$$

and let m_j be the expected number of transitions (or steps) between visits to state j, that is,

$$m_j \triangleq \mathbb{E}\left[N_j \,|\, X_0 = j\right].$$

As we saw in the examples when we were classifying states, recurrence is a property that might look differently in different DTMC's. In the next definition, we add a surname to distinguish two types of recurrent states.

Definition 3.9. *We say that a recurrent state j is positive recurrent if $m_j < \infty$ and that j is null recurrent if $m_j = \infty$.*

For example, in Example 3.9 we saw examples of positive recurrent classes, and in Example 3.10 we saw an example of null recurrent states when $p = 1/2$.

We now relate the time between visits to a state with the long-run proportion of time spent at that state.

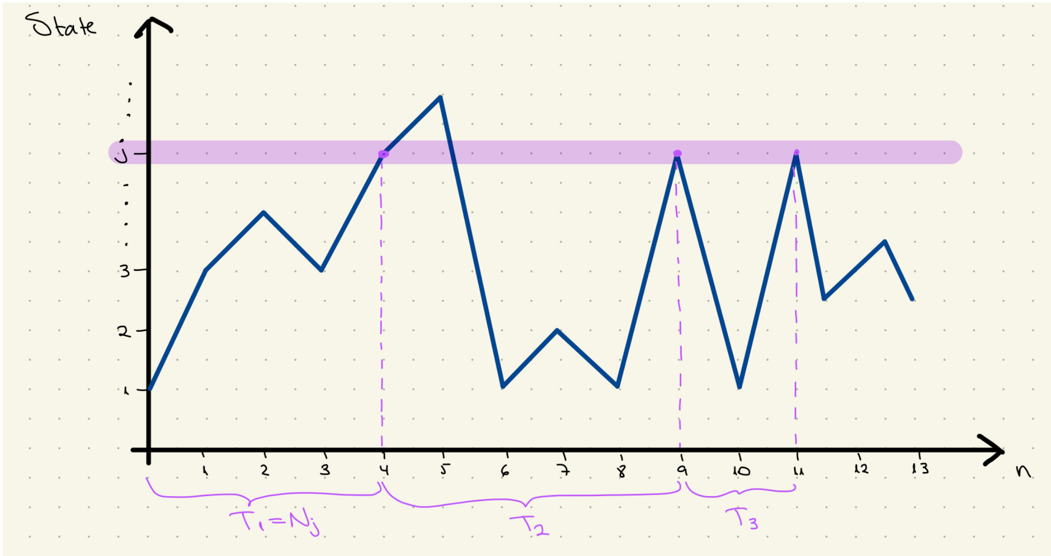
Proposition 3.6 (Proposition 4.4 from the textbook)**.** *Let π_j be the long-run proportion of time that a Markov chain spends at state j. If the Markov chain is irreducible and recurrent, then for any initial state we have:*

$$\pi_j = \frac{1}{m_j} \qquad \forall j \in \mathcal{X}$$

Proof. We define the time between visits as follows:

$$\begin{aligned} T_1 &= \min\{n \in \mathbb{Z}_+ : \; X_n = j\} \\ T_i &= \min\{n > T_{i-1} : \; X_n = j\} \qquad \forall i \geq 2 \end{aligned}$$

Observe that $T_1 = N_j$. In this case, we introduce more notation because we are interested in the time between visits. In the next figure, we plot time in the x axis and the state of the DTMC in the y axis to show what the T_i's represent pictorially.



Now, since T_i's are the number of transitions between visits to a particular state of a DTMC, and the future of DTMC's only depends on the current state, we know that $T_1, T_2, \ldots$ is a sequence of independent random variables. Further, $T_2, T_3, \ldots$ are also identically distributed.

Additionally, observe that $T_1 + T_2 + \cdots + T_n$ represents the time at which the n^{th} visit to state j occurs. Then, the proportion of time we spend at state j by time $T_1 + \cdots + T_n$ is $\frac{n}{T_1+\cdots+T_n}$ because, out of the $T_1 + \cdots + T_n$ transitions that have happened, n have been to state j.

Therefore, since π_j represents the long-run proportion of time spent at state j, we have

$$\begin{aligned}
\pi_j &= \lim_{n\to\infty} \frac{n}{T_1 + T_2 + \cdots + T_n} \\
&= \lim_{n\to\infty} \frac{1}{\frac{T_1+T_2+\cdots+T_n}{n}} \\
&= \lim_{n\to\infty} \frac{1}{\frac{T_1}{n} + \left(\frac{n-1}{n}\right)\left(\frac{T_2+\cdots+T_n}{n-1}\right)}
\end{aligned}$$

To compute the limit observe:

- T_1 is finite with probability 1 because j is a recurrent state. Then,

$$\lim_{n\to\infty} \frac{T_1}{n} = 0.$$

- $T_2, T_3, \ldots$ are independent and identically distributed random variables. Then, by the strong law of large numbers,

$$\lim_{n\to\infty} \frac{T_2 + \ldots + T_n}{n-1} = \mathbb{E}\left[T_i\right] = m_j.$$

Hence,

$$\pi_j = \frac{1}{m_j}.$$

□

The previous proposition also shows that $\pi_j > 0$ is equivalent to $m_j < \infty$. Then, we can say that a state j is positive recurrent if and only if $\pi_j > 0$, i.e., if and only if in the long run we spent some time in it.

We now show that positive recurrence is a class property. The proof is very similar to the proof that recurrence is a class property, so we only sketch it.

Proposition 3.7 (Proposition 4.5 from the textbook)**.** *If state i is positive recurrent and $i \leftrightarrow j$, then j is positive recurrent.*

Proof sketch. We need to show that $\pi_j > 0$. By definition of π_i and because $i \leftrightarrow j$, we obtain

$$\pi_i p_{i,j}^{(n)} \leq \pi_j$$

for n such that $p_{i,j}^{(n)} > 0$. In words, the proportion of time we spend at state j is larger than the proportion of time that we visit state j after visiting state i.

This shows that $\pi_j > 0$ because $\pi_i > 0$. □

From the propositions proved above, we can draw the following conclusions:

(i) Null recurrence is also a class property

(ii) In an irreducible Markov chain with finite state space, all the states are positive recurrent. Otherwise, the proportion of time spent in every state would be 0 in the long run.

The next theorem gives us a nice way to compute the proportion of time spent in each state.

Theorem 3.8 (Theorem 4.1 from the textbook)**.** *Consider an irreducible DTMC. If the chain is positive recurent, then the long time proportions are the unique solution to the system of equations*

$$\boldsymbol{\pi} = P^T \boldsymbol{\pi}, \qquad \sum_{j \in \mathcal{X}} \pi_j = 1.$$

Moreover, if there is no solution to the preceding system of equations, then the DTMC is either transient or null recurrent and all $\pi_j = 0$.

We won't prove the result, but let's try to make sense of the system of equations. If we write the equations $\boldsymbol{\pi} = P^T \boldsymbol{\pi}$ component-wise, we obtain

$$\pi_j = \sum_{i \in \mathcal{X}} p_{i,j} \pi_i \qquad \forall j$$

In words, this equation is saying that the proportion of time we spend in state j equals the proportion of time we spend in every state and travel to state j. Since the chain must have been somewhere before going to state j, this equation is just a result from the law of total probability.

Now, $\sum_j \pi_j = 1$ just says that the DTMC must always be at some of the state.

Long-run proportions vs. limiting probabilities

When we just learned probability, we were told that the probability of an event E can be computed as the proportion of successful cases over the total size of the state space. In this section, we've been very careful in calling π_j proportions and not probabilities. But do π_js represent the probability that the chain is in state j in the long run?

Let's first clarify what we mean by limiting probabilities.

Definition 3.10. *The limiting probability of visiting state j is defined as*

$$\alpha_j \triangleq \lim_{n \to \infty} \mathbb{P}[X_n = j]$$

Observe that α_j does not depend on the initial state. If the values α_j exist, then $\alpha_j = \pi_j$ because $\boldsymbol{\alpha}$ satisfies the equations from Theorem 3.8. Further, by definition,

$$\boldsymbol{\alpha} = \lim_{n \to \infty} (P^n)^T \boldsymbol{\alpha}^{(0)},$$

where $\boldsymbol{\alpha}^{(0)}$ represents the probability mass function of the initial state. Then, $\boldsymbol{\alpha}$ is well defined when the limit $\lim_{n \to \infty} P^n$ exists. As we saw in the homework, this limit exists if the chain is aperiodic.

After all, one interpretation of the equation $\boldsymbol{\pi} = P^T \boldsymbol{\pi}$ is that the probability mass function before a transition is $\boldsymbol{\pi}$ and after one transition it is the same, and this is clearly untrue when we have periodicity.

Let's see a simple example.

Example 3.11. *Consider the DTMC with transition probability matrix*

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Then, the proportion of time spent in each state is clearly 1/2. However,

$$P^{2n} = P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad \text{and} \qquad P^{2n-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \forall n \in \mathbb{Z}_+$$

Therefore, $\lim_{n\to\infty} P^n$ *does not exist.*

Now let's see some examples where $\boldsymbol{\pi}$ represents the limiting probabilities.

Example 3.12 (Example 4.22 from the textbook)**.** *Suppose that the probability that it rains tomorrow is* α *if it rained today and* β *if it did not rain today. What is the proportion of days that it will rain in the long run?*

$$P = \begin{bmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{bmatrix}$$

Solution. We solve the system of equations from Theorem 3.8. We obtain:

$$\begin{aligned} \pi_0 &= \alpha\pi_0 + \beta\pi_1 \\ \pi_1 &= (1-\alpha)\pi_0 + (1-\beta)\pi_1 \\ \pi_0 + \pi_1 &= 1. \end{aligned}$$

Solving the system of equations we obtain

$$\pi_0 = \frac{\beta}{1+\beta-\alpha}, \quad \pi_1 = \frac{1-\alpha}{1+\beta-\alpha}$$

□

In the next examples we interpret the equations $\boldsymbol{\pi} = P^T\boldsymbol{\pi}$ to solve. In the next example we use the interpretation of π_i as the proportion of time spent in state i in the long run.

Example 3.13 (Example 4.26 from the textbook)**.** *The printer in front of Becca's office changes states in accordance with an irreducible, positive recurrent Markov chain with* n *states and transition probabilities determined by the matrix* P*. Suppose that certain states are considered acceptable and the remaining, unacceptable. Let* A *be the set of acceptable states and* A^c *the unacceptable. If the printer is said to be "up" when in an acceptable state and "down" when in an unacceptable state, determine:*

(a) The rate at which the printer goes from up to down, that is, the rate of breakdowns

(b) The average length of time the printer remains down when it goes down

(c) The average length of time the printer remains up when it goes up

Solution. First observe that $\pi_i P_{ij}$ is the rate at which the DTMC goes from state i to j in the long run. Then, the rate at which the DTMC enters state j from an acceptable state is

$$\sum_{i\in A} \pi_i P_{i,j}$$

Hence, the rate at which the printer enters an unacceptable state starting from an acceptable state (which is the rate at which breakdowns occur) is

$$\sum_{j\in A^c}\sum_{i\in A} \pi_i P_{i,j}$$

Now, if we let $\overline{U}$ and $\overline{D}$ be the average time the printer is up and down, respectively, we have that breakdowns occur once every $\overline{U}+\overline{D}$ time slots. Hence,

$$\frac{1}{\overline{U}+\overline{D}} = \sum_{j\in A^c}\sum_{i\in A} \pi_i P_{i,j}$$

Out of the $\overline{U} + \overline{D}$ time slots between breakdowns, the printer spends $\overline{U}$ time slots up. Then, we get

$$\frac{\overline{U}}{\overline{U} + \overline{D}} = \sum_{j \in A} \pi_j$$

From these two equations we obtain that the average time that the printer is up is

$$\overline{U} = \frac{\sum_{j \in A} \pi_j}{\sum_{j \in A^c} \sum_{i \in A} \pi_i P_{i,j}}$$

and the average time it is down is

$$\overline{D} = \frac{\sum_{j \in A^c} \pi_j}{\sum_{j \in A^c} \sum_{i \in A} \pi_i P_{i,j}}$$

If the states of the printer are $\{1, 2, 3, 4\}$ with $A = \{1, 2\}$ and

$$P = \begin{bmatrix} 1/4 & 1/4 & 1/2 & 0 \\ 0 & 1/4 & 1/2 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 0 & 1/2 \end{bmatrix}$$

then,

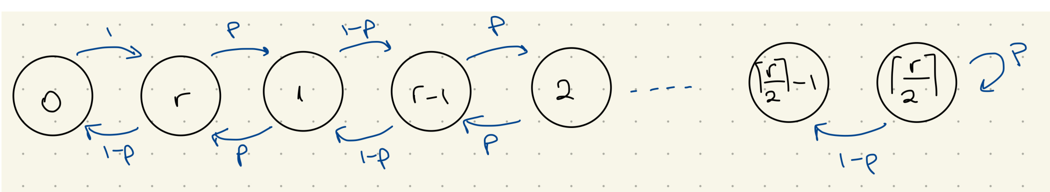
$$\boldsymbol{\pi} = \left(\frac{3}{16}, \frac{1}{4}, \frac{14}{48}, \frac{13}{48} \right)$$

and we obtain $\overline{U} = \frac{14}{9}$, $\overline{D} = 2$. □

In the next example we explicitly use the equations $\boldsymbol{\pi} = P^T \boldsymbol{\pi}$ to give an interpretation and solve.

Example 3.14 (Exercise 46 from Chapter 4 of the book)**.** *Ana possesses r umbrellas that she employs in going from her home to office, and vice versa. If she is at home/the office at the beginning/end of the day and it is raining, then she takes an umbrella with her to the office/home, provided there is one available at home/the office. If it is not raining, then she never takes an umbrella. Assume that, independent of the past, it rains at the beginning/end of the day with probability p. What fraction of the time does Ana get wet in the long run?*

Solution. We first define our state. For simplicity, let's assume r is an odd number. Let X_n be the number of umbrellas at the present location of Ana. Then, the diagram of the Markov chain is:



and the transition matrix is determined by

$$\begin{aligned} p_{0,r} &= 1 \\ p_{i,r-i+1} &= p \\ p_{i,r-i} &= 1 - p \end{aligned}$$

for all $i \in \{1, \ldots, r\}$.

To compute the probability that Ana gets wet, we need her to be at a location that does not have an umbrella and that it rains. Then, the probability that Ana gets wet is $p\pi_0$. We now compute $\boldsymbol{\pi}$.

One way to think of the equation $\boldsymbol{\pi} = P^T \boldsymbol{\pi}$ is that the out-flow to every state must equal the in-flow. Under this interpretation, we obtain the following system of equations:

$$\pi_0 = (1 - p)\pi_r$$

$$\pi_r = \pi_0 + p\pi_1$$
$$\pi_1 = p\pi_r + (1-p)\pi_{r-1}$$
$$\pi_{r-1} = (1-p)\pi_1 + p\pi_2$$
$$\vdots$$
$$\pi_i = p\pi_{r-i+1} + (1-p)\pi_{r-i}$$

Let's add the first two equations:

$$\pi_0 + \pi_r = (1-p)\pi_r + \pi_0 + p\pi_1 \qquad \Longrightarrow \qquad \pi_r = \pi_1$$

Using the last result in the third equation yields $\pi_1 = \pi_{r-1}$. Hence, we have

$$\pi_0 = (1-p)\pi_r \qquad \text{and} \qquad \pi_1 = \pi_2 = \cdots = \pi_r$$

Now we use that $\boldsymbol{\pi}$ is a probability mass function and obtain

$$1 = \sum_{i=0}^{r} \pi_i = \pi_0 \left(1 + \frac{r}{1-p}\right)$$
$$\Longrightarrow \pi_0 = \frac{1-p}{r+1-p}$$

Hence, Ana gets wet with probability $\frac{p(1-p)}{r+1-p}$ □

3.5 DTMC's with costs on each state

In this subsection we consider a cost associated to each state and we compute the long-run average cost.

Proposition 3.9 (Proposition 4.6 from the textbook)**.** *Let $\{X_n : n \in \mathbb{Z}_+\}$ be an irreducible Markov chain with stationary probabilities π_j for $j \in \mathcal{X}$. Let $r : \mathcal{X} \to \mathbb{R}$ be a bounded function on the state space. Then, with probability 1, we have*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} r(X_n) = \sum_{j\in\mathcal{X}} r(j)\pi_j$$

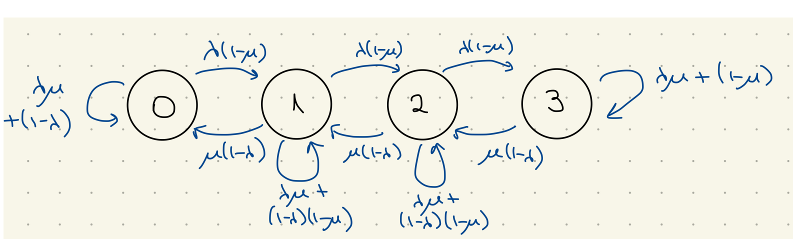
Example 3.15. ***Single server queue with finite waiting room****. Suppose that a patient arrives to a doctor's office with probability λ in each time slot. If the doctor is busy with a patient, the consult finishes with probability μ in each time slot. Due to COVID, the doctor has determined that she can only have 2 patients in the waiting room and that each patient in the waiting room has a cost of \$20. If a patient comes and the waiting room is full, she/he leaves.*

(a) What is the probability that a patient cannot enter the doctor's office?

(b) What is the long-run cost due to patients waiting?

Solution. Let X_n denote the number of patients in the doctor's office after transition n. Then, the probability that a patient cannot enter is $\lambda\pi_3$ and the long-run cost is $\sum_{i=1}^{r} 20i\pi_i$. Then, we compute $\boldsymbol{\pi}$.

We first draw the diagram and build the transition matrix.



Then, the transition probability matrix is

$$P = \begin{bmatrix} \lambda\mu + (1-\lambda) & \lambda(1-\mu) & 0 & 0 \\ \mu(1-\lambda) & \lambda\mu + (1-\lambda)(1-\mu) & \lambda(1-\mu) & 0 \\ 0 & \mu(1-\lambda) & \lambda\mu + (1-\lambda)(1-\mu) & \lambda(1-\mu) \\ 0 & 0 & \mu(1-\lambda) & \lambda\mu + (1-\mu) \end{bmatrix}$$

To compute the stationary distribution $\boldsymbol{\pi}$ we solve $\boldsymbol{\pi} = P^T\boldsymbol{\pi}$. We obtain:

$$\begin{aligned}
\pi_0 &= (\lambda\mu + 1 - \lambda)\,\pi_0 + \mu(1-\lambda)\pi_1 \\
\pi_1 &= \lambda(1-\mu)\pi_0 + (\lambda\mu + (1-\lambda)(1-\mu))\,\pi_1 + \mu(1-\lambda)\pi_2 \\
\pi_2 &= \lambda(1-\mu)\pi_1 + (\lambda\mu + (1-\lambda)(1-\mu))\,\pi_2 + \mu(1-\lambda)\pi_3 \\
\pi_3 &= \lambda(1-\mu)\pi_2 + (\lambda\mu + 1 - \mu)\,\pi_3
\end{aligned}$$

Rearranging terms in the first equation we obtain:

$$\mu(1-\lambda)\pi_1 = \lambda(1-\mu)\pi_0 \qquad \Longrightarrow \qquad \pi_1 = \frac{\lambda(1-\mu)}{\mu(1-\lambda)}\pi_0$$

Then, replacing this result on the second equation, we obtain

$$\pi_2 = \left(\frac{\lambda(1-\mu)}{\mu(1-\lambda)}\right)\pi_1 = \left(\frac{\lambda(1-\mu)}{\mu(1-\lambda)}\right)^2\pi_0$$

Then, using the last equation, we obtain

$$\pi_3 = \left(\frac{\lambda(1-\mu)}{\mu(1-\lambda)}\right)\pi_2 = \left(\frac{\lambda(1-\mu)}{\mu(1-\lambda)}\right)^3\pi_0$$

To solve we use the condition $\sum_{i=0}^{3}\pi_i = 1$. We obtain

$$1 = \sum_{i=0}^{3}\left(\frac{\lambda(1-\mu)}{\mu(1-\lambda)}\right)^i\pi_0 \qquad \Longrightarrow \qquad \pi_0 = \left[\sum_{i=0}^{3}\left(\frac{\lambda(1-\mu)}{\mu(1-\lambda)}\right)^i\right]^{-1}$$

Observe that this chain is positive recurrent for all values of λ and μ because the state space is finite and the chain is irreducible. For a queue with infinite room space we need to make sure that the sum of π's converges (as you'll do in your homework). □

3.6 Time until visiting a state

Before finishing this chapter, we study the time that it takes to visit a particular state for the first time. We do this with an example.

Example 3.16. *Consider the single-server queue described in Example 3.15. If the system starts empty, what is the expected number of transitions until it is full?*

Solution. Let N_i be the number of transitions until the number of customers is i for the first time starting from an empty system, and let $n_i \triangleq \mathbb{E}[N_i]$. Then, we want to compute n_0.

We will condition on the first transition of the DTMC to obtain a recursive equation that will depend on n_1. Then, following the same steps for n_1 and n_2 we will obtain a linear system of equations. The last step is to simply solve the system of equations to obtain our answer.

Using the tower property to condition on the first transition we obtain:

$$\begin{aligned}
n_0 = \mathbb{E}[N_0] &= \mathbb{E}[\mathbb{E}[N_0|X_1]] \\
&= \sum_{i=0}^{3}\mathbb{E}[N_0|X_1 = i]\,p_{0,i} \\
&= (\lambda\mu + 1 - \lambda)\,\mathbb{E}[N_0|X_1 = 0] + \lambda(1-\mu)\mathbb{E}[N_0|X_1 = 1]
\end{aligned}$$

$$= (\lambda\mu + 2 - \lambda)\,(1 + n_0) + \lambda(1-\mu)(1 + n_1),$$

where we used that the first transition adds 1 step to n_0 and $(X_n)_n$ is a Markov chain, so after giving one step the future only depends on the new state. Hence, if we stay at state 0, the expected number of steps until entering state 3 is n_0 and if we go to state 1, the expected number of steps until we get to state 3 is n_1.

Rearranging terms we obtain the following linear equation:

$$n_0 - n_1 = \frac{1}{\lambda(1-\mu)} \tag{3}$$

Now we follow the same steps to obtain an equation starting from n_1. We obtain

$$n_1 = (1 + n_0)\mu(1-\lambda) + (1 + n_1)(\lambda\mu + (1-\lambda)(1-\mu)) + (1 + n_2)\lambda(1-\mu)$$

Reorganizing terms, we obtain

$$1 = \mu(1-\lambda)(n_1 - n_0) + \lambda(1-\mu)(n_1 - n_2)$$

and using the result from Equation (3) we obtain

$$n_1 - n_2 = \frac{1}{\lambda(1-\mu}\left(1 + \frac{\mu(1-\lambda)}{\lambda(1-\mu)}\right) \tag{4}$$

Following similar steps for n_2 we obtain

$$\begin{aligned} n_2 &= (1 + n_1)\mu(1-\lambda) + (1 + n_2)(\lambda\mu + (1-\lambda)(1-\mu)) + (1 + n_3)\lambda(1-\mu) \\ &= (1 + n_1)\mu(1-\lambda) + (1 + n_2)(\lambda\mu + (1-\lambda)(1-\mu)) + \lambda(1-\mu), \end{aligned}$$

where the last equality holds because $n_3 = 0$. Reorganizing terms, we obtain

$$\begin{aligned} n_2 &= \frac{1}{\lambda(1-\mu)}\left(1 + \mu(1-\lambda)(n_1 - n_2)\right) \\ &= \frac{1}{\lambda(1-\mu)}\left(1 + \frac{\mu(1-\lambda)}{\lambda(1-\mu)} + \left(\frac{\mu(1-\lambda)}{\lambda(1-\mu)}\right)^2\right) \end{aligned}$$

Plugging the last result back in Equation (4) to obtain n_1, and then using the result in Equation (3) we obtain the value of n_0.

□

4 Exponential Distribution and Poisson Process

[This section is based on Chapter 5 of the textbook]

After spending some time studying Discrete-Time Markov chains, we move to continuous-time processes. In this chapter we will look at the time between events instead of the number of events in fixed time slots. These processes are fairly complicated, but we can make some mild assumptions to gain tractability without losing practicality. Specifically, we will assume that the time between events follows an exponential distribution. We start the chapter studying the most important properties of this distribution.

4.1 The exponential distribution

We start with the formal definition.

Definition 4.1. *A continuous random variable X is said to have exponential distribution with parameter λ (with $\lambda > 0$) if its probability density function is*

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{, if } x \geq 0 \\ 0 & \text{, otherwise} \end{cases}$$

or, equivalently, if its cumulative distribution function is

$$F_X(x) = \begin{cases} 0 & \text{, if } x < 0 \\ 1 - e^{-\lambda x} & \text{, if } x \geq 0 \end{cases}$$

The mean and variance of an exponential random variable are:

$$\mathbb{E}[X] = \frac{1}{\lambda} \qquad \text{and} \qquad \text{Var}[X] = \frac{1}{\lambda^2}$$

Memoryless property

One of the most important properties of the exponential distribution, and probably the main reason why it is so easy to work with, is the memoryless property. We define it below.

Definition 4.2. *A random variable Y is said to be memoryless if for all $s, t \geq 0$ we have*

$$\mathbb{P}[Y > t + s \,|\, Y > t] = \mathbb{P}[Y > s] \tag{5}$$

In words, the memoryless property says that the random variable Y does not deteriorate with time. For example, if Y represents the lifetime of an instrument, the memoryless property establishes that the current age t of the object does not affect the probability that it lasts for additional s time units. If we look at the equation above, the left-hand side establishes that the lifetime of the instrument is s extra time units considering that it hasn't failed up to t. On the right-hand side, we simply compute the probability of the instrument's lifetime being at least s time units.

Theorem 4.1. *Exponentially distributed random variables are the only continuous random variables that satisfy the memoryless property.*

Proof. We only prove that exponential random variables are memoryless. We omit the proof of uniqueness. If we use the definition of conditional probability in Equation (5) we obtain:

$$\mathbb{P}[Y > t + s,\, Y > t] = \mathbb{P}[Y > t]\,\mathbb{P}[Y > s]$$

But $Y > t + s$ implies $Y > t$. Then, $\mathbb{P}[Y > t + s,\, Y > t] = \mathbb{P}[Y > t + s]$. Hence, the memoryless property can be equivalently written as

$$\mathbb{P}[Y > t + s] = \mathbb{P}[Y > t]\,\mathbb{P}[Y > s].$$

If $Y \sim \mathit{Expo}(\lambda)$, then $\mathbb{P}[Y > x] = e^{-\lambda x}$. Then, we obtain

$$\begin{aligned} &\mathbb{P}[Y > t + s] = e^{-\lambda(t+s)} \\ \text{and} \quad &\mathbb{P}[Y > t]\,\mathbb{P}[Y > s] = e^{-\lambda t} e^{-\lambda s} = e^{-\lambda(t+s)}. \end{aligned}$$

Hence, the exponential random variable is memoryless. □

Example 4.1 (Example 5.3 from the textbook)**.** *Consider a post office that is run by two clerks. Suppose that when Mr. Smith enters the system he discovers that Mr. Jones is being served by one of the clerks and Mr. Brown by the other. Suppose also that Mr. Smith is told that his service will begin as soon as either Mr. Jones or Mr. Brown leaves. If the amount of time that a clerk spends with a customer is exponentially distributed with mean* $1/\lambda$*, what is the probability that Mr. Smith is the last customer to leave the post office among the three?*

Solution. Mr. Smith only starts being processed when one of the current customers ends. Suppose it's Mr. Jones. Then, Mr. Smith is the last one to leave if his total service time is larger than the remaining service time from Mr. Brown. Since the service times are exponential, by the memoryless property, the remaining service time of Mr. Smith and Mr. Brown have the same distribution. Hence, by symmetry, Mr. Smith is the last one with probability 1/2. □

Example 4.2 (Example 5.4 from the textbook)**.** *The amount of money in damage in an automobile accident is an exponential random variable with mean 1000. Of this, the insurance company only pays the amount exceeding the deductible of 400. Find the expected value and standard deviation of the amount the insurance company pays per accident.*

Solution. Let X be the amount of money in damage in an automobile accident and Y the amount of money paid by the insurance company. Then,$Y = \max\{X - 400, 0\}$. Let I be the indicator function that the insurance company pays something, that is,

$$I = \begin{cases} 1 & \text{, if } X > 400 \\ 0 & \text{, if } X \leq 400 \end{cases}$$

Then, using the memoryless property, we obtain

$$\begin{aligned} &\mathbb{E}[Y \mid I = 1] = 1000 \\ &\mathbb{E}[Y \mid I = 0] = 0 \\ &\mathrm{Var}[Y \mid I = 1] = 1000^2 \\ &\mathrm{Var}[Y \mid I = 0] = 0 \end{aligned}$$

because $Y \mid I = 1$ has the same distribution as X. We can write the expessions above as follows

$$\mathbb{E}[Y \mid I] = 10^3 I \qquad \text{and} \qquad \mathrm{Var}[Y \mid I] = 10^6 I$$

Now, I is a Bernoulli random variable with parameter

$$p = \mathbb{P}[I = 1] = \mathbb{P}[X > 400] = e^{-400/1000} = e^{-0.4}$$

Then, using the tower property we obtain that the mean of what the insurance company pays is

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|I]] = \mathbb{E}\left[10^3 I\right] = 10^3\mathbb{E}[I] = 10^3 e^{-0.4} \approx 670.32$$

Finally, using the law of total variance we have

$$\begin{aligned} \mathrm{Var}[Y] &= \mathbb{E}[\mathrm{Var}[Y|I]] + \mathrm{Var}[\mathbb{E}[X|I]] \\ &= \mathbb{E}\left[10^6 I\right] + \mathrm{Var}\left[10^3 I\right] \\ &= 10^6\mathbb{E}[I] + 10^6\mathrm{Var}[I] \\ &= 10^6 e^{-0.4} + 10^6 e^{-0.4}(1 - e^{-0.4}) \end{aligned}$$

where we used that a Bernoulli random variable with parameter p has mean p and variance $p(1-p)$.

Then, the standard deviation of the amount paid by the insurance company is

$$\sqrt{\mathrm{Var}[Y]} \approx 944.09$$

□

Properties of the exponential distribution

We list some well-known properties of the exponential distribution:

1. If $X_1, X_2, \ldots, X_n$ are iid exponential random variables with parameter λ, then $Y = \sum_{i=1}^{n} X_i$ is Gamma(n, λ), that is,
$$f_Y(y) = \lambda e^{-\lambda y} \frac{(\lambda y)^{n-1}}{(n-1)!}$$
2. If X_1, X_2 are independent random variables with parameter λ_1, λ_2, respectively, then
$$\mathbb{P}[X_1 < X_2] = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$
3. Let $X_1, \ldots, X_n$ be independent exponential random variables with parameter $\lambda_1, \ldots, \lambda_n$, respectively. Then, $Y = \min\{X_1, \ldots, X_n\}$ is exponential with parameter $\lambda_1 + \ldots + \lambda_n$

The last two properties are particularly interesting because they are related to the first event that happens among a list of independent events "competing".

Example 4.3 (Example 5.8 from the textbook)**.** *Suppose that a post office has two clerks. Clerk i takes an exponentially distributed time with rate λ_i to process customers' requirements. You arrive at the post office when both clerks are busy and no one else is waiting in line, that is, whenever the first clerk is free, you start being served. Let T be the amount of time that you spend at the post office. Compute $\mathbb{E}[T]$.*

Solution. The time that you take in the post office depends on which clerk serves you, that is, of the customer that leaves the system first.

Let R_i be the remaining time of the customer at clerk i, for $i = 1, 2$. By the memoryless property, R_i is an exponential random variable with parameter λ_i. Then,

$$\begin{aligned}\mathbb{E}[T] &= \mathbb{E}[T \mid R_1 < R_2]\,\mathbb{P}[R_1 < R_2] + \mathbb{E}[T \mid R_1 > R_2]\,\mathbb{P}[R_1 > R_2] \\ &= \mathbb{E}[T \mid R_1 < R_2]\,\frac{\lambda_1}{\lambda_1 + \lambda_2} + \mathbb{E}[T \mid R_1 > R_2]\,\frac{\lambda_2}{\lambda_1 + \lambda_2}\end{aligned}$$

Finally, we compute $\mathbb{E}[T \mid R_1 < R_2]$. Let S be your service time. Then,

$$\begin{aligned}\mathbb{E}[T \mid R_1 < R_2] &= \mathbb{E}[R_1 + S \mid R_1 < R_2] \\ &= \mathbb{E}[R_1 \mid R_1 < R_2] + \mathbb{E}[S \mid R_1 < R_2] \\ &= \mathbb{E}[\min R_1, R_2] + \mathbb{E}[S \mid R_1 < R_2] \\ &= \frac{1}{\lambda_1 + \lambda_2} + \frac{1}{\lambda_1}\end{aligned}$$

where we used property 3. to compute $\mathbb{E}[\min\{R_1, R_2\}]$ in the last step. Similarly,

$$\mathbb{E}[T \mid R_1 > R_2] = \frac{1}{\lambda_1 + \lambda_2} + \frac{1}{\lambda_2}$$

Putting everything together we obtain

$$\mathbb{E}[T] = \frac{3}{\lambda_1 + \lambda_2}$$

□

4.2 The Poisson Process

The exponential distribution gives us the lifetime of a device, or the time that a customer spends being processed by a server. We now switch to the number of events that happen in a given interval of time. For example, how many devices die in an interval, or how many customers finish service. We start with a basic definition.

Counting processes

Definition 4.3. *A stochastic process $\{N(t) : t \geq 0\}$ is said to be a counting process if $N(t)$ represents the total number of events that happen by time t.*

A counting process satisfies the following properties:

(i) $N(t) \geq 0$ for all $t \geq 0$

(ii) $N(t)$ only takes integer values

(iii) $N(t)$ is nondecreasing in t, that is, if $0 \leq s < t$, then $N(s) \leq N(t)$

(iv) For $0 \leq s < t$, $N(t) - N(s)$ represents the number of events that occur in the interval $(s, t]$.

Examples of counting processes are:

- Number of people who enter a store. Here, an event corresponds to a person entering the store
- Number of people who were born
- Number of goals that a soccer player scores
- Number of holes in the I-64. Here we need to fix the origin, and t represents the distance to the origin in one direction.

Since the counting processes need to be increasing, the number of people in the store of the number of alive people are not counting processes.

Definition 4.4. *A counting process $\{N(t) : t \geq 0\}$ is said to possess:*

(a) Independent increments if the number of events that occur in disjoint time intervals are independent. Mathematically, if $0 \leq s < t < u$, then $N(u) - N(t)$ is independent of $N(s)$. Further, $N(t) - N(s)$ is independent of $N(s)$.

(b) Stationary increments if the distribution of the number of events that occur in any interval of time depends only on the length of the time interval. Mathematically, if $s, t > 0$, then $N(t + s) - N(s)$ has the same distribution than $N(t)$.

The number of people who were born may not have independent nor stationary increments, for the following reason.

- If $N(t)$ is very large, then chances are that a lot of people are alive. Therefore, we expect that more people will be born and, hence, we expect that $N(t') - N(t)$ will be large for $t' > t$. However, if $N(t)$ is small we expect that $N(t') - N(t)$ will also be small.
- If $N(t)$ had stationary increments, then the rate at which population grows would be constant. However, it is well known that this is not true.

Definition of the Poisson Process

We now formally define a very special and useful counting process. But before, we introduce the following notation. We say that a function $f(h)$ is $o(h)$ if

$$\lim_{h \to 0} \frac{f(h)}{h} = 0$$

that is, if $f(h)$ decreases to 0 faster than h.

Definition 4.5 (Definition 5.2 from the textbook)**.** *The counting process $\{N(t) : t \geq 0\}$ is said to be a Poisson process with rate $\lambda > 0$ if:*

(i) $N(0) = 0$

(ii) $\{N(t) : t \geq 0\}$ *has independent increments*

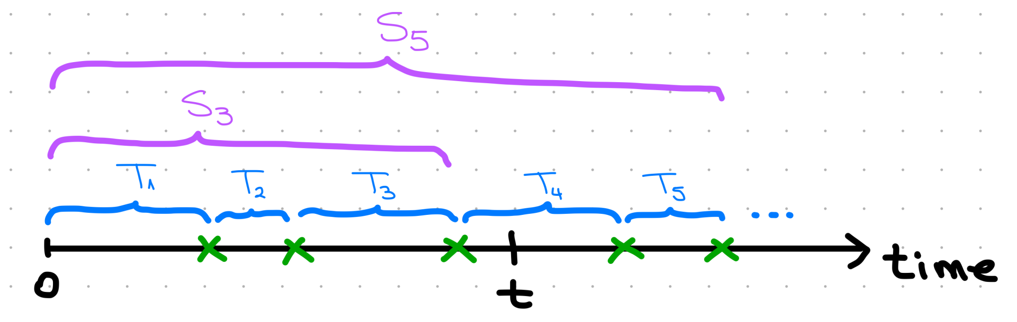
(iii) $\mathbb{P}[N(t+h) - N(t) = 1] = \lambda h + o(h)$*, that is, for small values of* h, $\mathbb{P}[N(t+h) - N(t) = 1] \approx \lambda h$

(iv) $\mathbb{P}[N(t+h) - N(t) \geq 2] = o(h)$*, that is, for small values of* h, $\mathbb{P}[N(t+h) - N(t) \geq 2] \approx 0$

In words, the last two properties say that in small intervals of time, the probability of having 1 event is proportional to the rate λ and the probability of having 2 or more events is nearly 0. That is, there are no simultaneous events.

Definition 4.6. *For integer numbers* i*, let* T_i *be the time between the* i^{th} *and the* $(i-1)^{th}$ *event. Then,* $\{T_i : i = 1, 2, 3 \dots\}$ *is the sequence of interarrival times.*
Additionally, define $S_n \triangleq T_1 + \dots + T_n$ *as the time of the* n^{th} *event.*

Pictorially, the counting process $N(t)$, the interarrival times T_i and the times of events S_n have the following relationship:



Then, $N(t) = 3$ and observe that $S_3 < t$ and $S_4 > t$.

Proposition 4.2. $T_1, T_2, \dots$ *are iid exponential random variables with parameter* λ

We skip the proof, but it is based on a differential equation that we can build using properties (iii) and (iv) from the definition, and by the independent increments property.
As a corollary, we can easily conclude that S_n is a $Gamma(n, \lambda)$ random variable.
The following result gives us the pmf of $N(t)$, and is essential for the rest of the chapter.

Theorem 4.3 (Theorem 5.1 from the textbook)**.** *If* $\{N(t) : t \geq 0\}$ *is a Poisson process with rate* λ*, then* $N(t)$ *is a Poisson random variable with rate* λt*. That is,*

$$\mathbb{P}[N(t) = n] = \frac{e^{-\lambda t}(\lambda t)^n}{n!} \qquad \forall n \in \mathbb{Z}_+$$

Observe that if $\{N(t) : t \geq 0\}$ is a Poisson process, then $\{N_s(t) : t \geq 0\}$ with $N_s(t) = N(t+s) - N(s)$ is also a Poisson process. By the last theorem, we know that

$$\mathbb{P}[N(t) = n] = \mathbb{P}[N_s(t) = n].$$

Therefore, the Poisson process has stationary increments.

Example 4.4. *Suppose that people immigrate into a territory according to a Poisson process with rate 2 per day.*

(a) Find the probability that 10 immigrants arrive in the following week (7 days)

(b) Find the expected time until we have 20 immigrants

Solution.

(a) We use the theorem:

$$\mathbb{P}[N(7) = 10] = \frac{e^{-2\times 7}(2 \times 7)^{10}}{10!}$$

(b) Here we obtain $\mathbb{E}[S_{20}] = \frac{20}{2} = 10$.

$\square$

Example 4.5. *Cars pass a certain street location according to a Poisson process with rate λ. A person who wants to cross the street at that location waits until she/he can see that no cars will come by in the next T time units.*

(a) Find the probability that the person waits 0 time units

(b) Find the expected waiting time of the person.

Solution.

(a) The person does not wait if the time of the first car (event) is greater than T, that is

$$\mathbb{P}[T_1 > T] = \mathbb{P}[N(T) = 0] = e^{-\lambda T}$$

(b) To compute the expected waiting time we condition on the arrival of the first car. Let W be the waiting time. Then,

$$\begin{aligned}
\mathbb{E}[W] &= \mathbb{E}[\mathbb{E}[W|T_1]] \qquad \text{(tower property)} \\
&= \mathbb{E}[W \mid T_1 > T]\,\mathbb{P}[T_1 > T] + \mathbb{E}[W \mid T_1 < T]\,\mathbb{P}[T_1 < T] \\
&= 0 \times \mathbb{P}[T_1 > T] + (\mathbb{E}[W] + \mathbb{E}[T_1 \mid T_1 < T])\,\mathbb{P}[T_1 < T]
\end{aligned}$$

where $\mathbb{P}[T_1 < T] = 1 - e^{-\lambda T}$ and

$$\begin{aligned}
\mathbb{E}[T_1 \mid T_1 < T] &= \int_0^\infty t\, f_{T_1|T_1<T}(t)\, dt \\
&= \int_0^T t\, \frac{f_{T_1}(t)}{\mathbb{P}[T_1 < T]}\, dt \\
&= \frac{1}{\mathbb{P}[T_1 < T]} \int_0^T t\lambda e^{-\lambda t}\, dt \\
&= \frac{1}{\mathbb{P}[T_1 < T]} \left(\frac{1 - e^{-\lambda T}(\lambda T + 1)}{\lambda} \right)
\end{aligned}$$

Hence, we obtain

$$\begin{aligned}
\mathbb{E}[W] &= \left(\frac{1}{1 - e^{-\lambda T}} \right) \left(\frac{1 - e^{-\lambda T}(\lambda T + 1)}{\lambda} \right) \\
&= \left(\frac{1}{1 - e^{-\lambda T}} \right) \left(\frac{1 - e^{\lambda T} - \lambda T e^{-\lambda T}}{\lambda} \right) \\
&= \frac{1}{\lambda} - T \left(\frac{e^{-\lambda T}}{1 - e^{-\lambda T}} \right).
\end{aligned}$$

$\square$

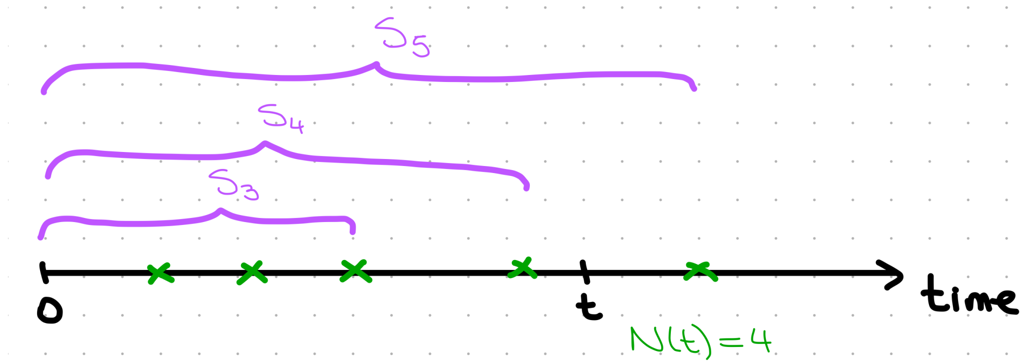
Before digging into some of the most important properties of the Poisson process, let's get a closer look at the relationship between $N(t)$ and S_n. Indeed, the equivalences we write below are true for any counting process such that the interarrival times are independend.

Proposition 4.4. *Let $\{N(t) : t \geq 0\}$ be a counting process, $T_1, T_2, \ldots$ be the interarrival times and $S_n \triangleq \sum_{i=1}^n T_i$, that is, S_n is the time of the n^{th} event. Then,*

(1) $\{N(t) = n\} = \{S_n < t,\, S_{n+1} > t\}$

(2) $\{N(t) < n\} = \{S_n > t\}$. *Observe that both inequalities are strict.*

The proof follows easily after carefully plotting an example.



In the figure, $N(t) = 4$ and observe that saying $S_4 < t$ is not sufficient because we also have $S_3 < t$. To ensure $N(t) = 4$ we need both: $S_4 < t$ and $S_5 > t$. For the second case, observe that $N(t) < n$ is saying that less than n events have occurred by time t. In the picture, we can say $N(t) < 5$. Equitalently, we can say that the time at which the n^{th} event occurs has not come yet, which is equivalent to saying $S_n > t$. In the picture, $S_5 > t$.

Properties of the Poisson process

We start with a property called Poisson splitting.

Proposition 4.5. *Suppose $\{N(t) : t \geq 0\}$ is a Poisson process with rate λ. Each of the events is of type I with probability p and of type II with probability $1 - p$. Let $N_i(t)$ be the number of type-i events occurring in $[0, t]$ for $i = 1, 2$, and observe $N(t) = N_1(t) + N_2(t)$.*

Then, $\{N_1(t) : t \geq 0\}$ and $\{N_2(t) : t \geq 0\}$ are independent Poisson processes with rates λp and $\lambda(1 - p)$, respectively.

The proof is easy, and follows after verifying the properties from the definition of Poisson process. We skip it.

Example 4.6 (Example 5.15 from the textbook)**.** *Suppose that offers to buy an item you are selling arrive according to a Poisson process with rate λ. Each offer is the value of a continuous random variable Y with pdf $f(y)$. Once the offer is presented, you must either accept it, or reject it and wait for the next offer.*

Suppose that you incur in a cost of c per unit time until the item is sold. Your objective is to maximize your expected total return, and suppose you accept the first offer that is greater than some pre-determined value y. What is the expected return in terms of y?

Solution. Let X be the value of an offer, and $F(x)$ be its cdf. Then, the probability that an offer is for a value higher than y is $1 - F(y)$. Using the property above, offers with a value higher than y are a Poisson process with rate $\lambda(1 - F(y))$. Hence, the time until the next offer with value higher than y is $\frac{1}{\lambda(1-F(y))}$.

We compute the expected return in terms of y. Let $R(y)$ be the total return from the policy that accepts an offer larger than y. Then,

$$
\begin{aligned}
\mathbb{E}[R(y)] &= \mathbb{E}[\text{accepted offer}] - c\mathbb{E}[\text{time to accept}] \\
&= \mathbb{E}[X \mid X > y] - \frac{c}{\lambda(1 - F(y))} \\
&= \int_0^\infty x f_{X|X>y}(x)\, dx - \frac{c}{\lambda(1 - F(y))} \\
&= \int_y^\infty x \frac{f_X(x)}{\mathbb{P}[X > y]}\, dx - \frac{c}{\lambda(1 - F(y))} \\
&= \frac{1}{1 - F(y)} \left(\int_y^\infty x f_X(x)\, dx - \frac{c}{\lambda} \right)
\end{aligned}
$$

□

The inverse of the property described in Proposition 4.5 is that the sum of two Poisson processes is also a Poisson process. We formally state it below.

Proposition 4.6. *For $i = 1, 2$, let $\{N_i(t) : t \geq 0\}$ be a Poisson process with rate λ_i, and suppose they are independent. Then, defining $N(t) = N_1(t) + N_2(t)$ we obtain that $\{N(t) : t \geq 0\}$ is a Poisson process with rate $\lambda_1 + \lambda_2$.*

Conditional distribution of the arrival times

Independent and stationary increments of the Poisson process make it very easy to deal with. So far we have been studying the probability of observing a certain number of events in a fixed interval of time. Now we will focus on a fixed number of events and we will study their arrival time.

Suppose we know that exactly one Poisson event occurre in the last t units of time. Can we get an idea of the exact time at which it happened? Since the Poisson process has independent and stationary increments, this question is very easy to answer. Indeed, for any $0 < s < t$ we have

$$\begin{aligned}
\mathbb{P}\left[T_1 < s \,|\, N(t) = 1\right] &= \frac{\mathbb{P}\left[T_1 < s, N(t) = 1\right]}{\mathbb{P}\left[N(t) = 1\right]} && \text{(definition of conditional probability)} \\
&= \frac{\mathbb{P}\left[N(s) = 1,\, N(t) - N(s) = 0\right]}{\mathbb{P}\left[N(t) = 1\right]} \\
&= \frac{\mathbb{P}\left[N(s) = 1\right] \mathbb{P}\left[N(t) - N(s) = 0\right]}{\mathbb{P}\left[N(t) = 1\right]} && \text{(independent increments)} \\
&= \frac{\mathbb{P}\left[N(s) = 1\right] \mathbb{P}\left[N(t - s) = 0\right]}{\mathbb{P}\left[N(t) = 1\right]} && \text{(stationary increments)} \\
&= \frac{\left(e^{-\lambda s}(\lambda s)^1/1!\right)\left(e^{-\lambda(t-s)}\right)}{e^{-\lambda t}(\lambda t)^1/1!} && \text{(Poisson distribution)} \\
&= \frac{s}{t}
\end{aligned}$$

which is the cdf of a uniform random variable in $[0, t]$.

This result can be generalized to conditioning on $N(t) = n$ and computing the joint density of the time of occurrence of events in $[0, t]$. In English, the result says that under the condition $N(t) = n$, the times $S_1, \ldots, S_n$ at which the events occur, considered as unordered random variables, are distributed independently and uniformly in the interval $(0, t)$. We state the formal result below.

Theorem 4.7 (Theorem 5.2 from the textbook)**.** *Given that $N(t) = n$, the n arrival times $S_1, \ldots, S_n$ have the same distribution as the order statistics corresponding to n random variables uniformly distributed on the interval $(0, t)$, that is*

$$f_{S_1,\ldots,S_n \,|\, N(t)}(s_1, \ldots, s_n | n) = \frac{n!}{t^n}.$$

The expression above means that each of the n events has density $1/t$ and they are independent. Hence, we get $1/t^n$. Since there are $n!$ ways to order n events, the density must be multiplied by this factor.

Example 4.7 (Exercise 5-64 from the textbook)**.** *Suppose that people arrive to a bus stop in accordance with a Poisson process at rate λ. The bus departs at time t (fixed). Let X be the total waiting time of all those who get on the bus at time t, that is, if there are two people at the bus stop by time t, the first person arrived at $t/3$ and the second one at $t/2$, then $X = 2t/3 + t/2$. Determine $Var[X]$.*

Solution. The first observation is that X is the sum of the waiting time of all the passengers. Then, if W_i denotes the waiting time of the i^{th} passenger,

$$X = \sum_{i=1}^{N(t)} W_i.$$

Since X depends on the counting process $N(t)$, we solve this problem using the law of total variance as follows:

$$\text{Var}\left[X\right] = \text{Var}\left[\mathbb{E}\left[X|N(t)\right]\right] + \mathbb{E}\left[\text{Var}\left[X|N(t)\right]\right]$$

We compute each of the terms. Since we are conditioning on the number of evens in the interval $(0, t]$, the arrival time of each of the passengers is uniform and the arrival time of different passengers are independent. That is, the arrival times of the $N(t)$ passengers are i.i.d. uniform random variable in $(0,t)$. If a passenger arrived at time s, then his/her waiting time is $t-s$. Hence, the waiting times are also i.i.d. random variables with Uniform distribution in $(0,t)$. We obtain

$$\begin{aligned}
\mathbb{E}\left[X|N(t)\right] &= \mathbb{E}\left[\left.\sum_{i=1}^{N(t)} W_i\right| N(t)\right] \\
&= N(t)\mathbb{E}\left[W_1 \mid N(t)\right] \qquad \text{(since } W_i|N(t) \text{ are i.i.d. random variables)} \\
&= N(t)\mathbb{E}\left[U(0,t)\right] \\
&= N(t)\frac{t}{2}
\end{aligned}$$

Then, we obtain

$$\mathrm{Var}\left[\mathbb{E}\left[X|N(t)\right]\right] = \mathrm{Var}\left[N(t)\frac{t}{2}\right] = \frac{t^2}{4}\mathrm{Var}\left[N(t)\right] = \frac{\lambda t^3}{4}$$

Now we compute the second term. Under similar reasoning, we obtain

$$\begin{aligned}
\mathrm{Var}\left[X|N(t)\right] &= \mathrm{Var}\left[\left.\sum_{i=1}^{N(t)} W_i\right| N(t)\right] \\
&= \sum_{i=1}^{N(t)} \mathrm{Var}\left[W_i \mid N(t)\right] \qquad \text{(since the } W_i\text{'s are independent)} \\
&= N(t)\mathrm{Var}\left[W_1 \mid N(t)\right] \qquad \text{(since the } W_i\text{'s are identically distributed)} \\
&= N(t)\frac{t^2}{12} \qquad \text{(since } \mathrm{Var}\left[U(0,t)\right] = t^2/12\text{)}
\end{aligned}$$

Therefore,

$$\mathbb{E}\left[\mathrm{Var}\left[X|N(t)\right]\right] = \frac{t^2}{12}\mathbb{E}\left[N(t)\right] = \frac{\lambda t^3}{12}$$

Putting everything together, we obtain

$$\mathrm{Var}\left[X\right] = \frac{\lambda t^3}{4} + \frac{\lambda t^3}{12} = \frac{t^3}{3}$$

□

In the next example we study a queueing model with infinitely many servers. That is, the customers do not wait in line because there is always a server available. One application of this model is self-service systems.

Example 4.8 (Example 5.18 from the textbook)**.** *Suppose that customers arrive at a self-service station according to a Poisson process with rate λ. The time each person takes to complete their request has cdf G, and each person's service time is independent from everybody else's and from the number of people in the system. Let $X(t)$ denote the number of customers that have completed service by time t and $Y(t)$ the number of customers in the system at time t. Compute the distribution of $X(t)$ and $Y(t)$.*

Solution. An equivalent way to define $X(t)$ is as the number of customers that arrived to the system at some time $s<t$ and their service time is below $t-s$. Similarly, $Y(t)$ is the number of customers that arrived to the system at some time $s<t$ and their service time is larger than $t-s$.

Hence, a convenient way to solve this problem is by conditioning on the number of customers that have arrived to the system by t, denoted by $N(t)$. Then, the arrival time of each customer is uniform and we can solve.

Let's start with $X(t)$. First observe that, since the arrival times given $N(t)$ are uniform in $(0,t)$, we have

$$\begin{aligned}
\mathbb{P}\left[X(t)=1 \mid N(t)=1\right] &= \int_0^t \mathbb{P}\left[\text{service time} \le t-s \,|N(t)=1\right] f_{U(0,t)}(s)\, ds \\
&= \int_0^t G(t-s)\, \frac{1}{t}\, ds
\end{aligned}$$

and let's denote

$$I_t \triangleq \int_0^t G(t-s)ds$$

Given the number of events $N(t)$, the arrival time of different customers are independent. Therefore, for $n \geq k$ we have

$$\mathbb{P}\left[X(t) = k \mid N(t) = n\right] = \binom{n}{k}\left(\frac{I_t}{t}\right)^k \left(1 - \frac{I_t}{t}\right)^{n-k}$$

that is, the customers that leave distribute as a binomial random variable where success is defined as "finishing service by t". Therefore, we can compute the distribution of the number of customers that have completed service by t as follows:

$$\begin{aligned}
\mathbb{P}\left[X(t) = k\right] &= \sum_{n=k}^{\infty} \mathbb{P}\left[X(t) = k \mid N(t) = n\right]\mathbb{P}\left[N(t) = n\right] \\
&= \sum_{n=k}^{\infty} \binom{n}{k}\left(\frac{I_t}{t}\right)^k \left(1 - \frac{I_t}{t}\right)^{n-k} \frac{e^{-\lambda t}(\lambda t)^n}{n!} \\
&= \frac{e^{-\lambda t}(\lambda I_t)^k}{k!} \sum_{n=k}^{\infty} \frac{\left[\lambda(t - I_t)\right]^{n-k}}{(n-k)!} \qquad \text{(reorganizing terms)} \\
&= \frac{e^{-\lambda t}(\lambda I_t)^k}{k!} \sum_{i=0}^{\infty} \frac{\left[\lambda(t - I_t)\right]^{i}}{i!} \\
&= \frac{e^{-\lambda t}(\lambda I_t)^k}{k!} e^{(t-I_t)\lambda} \\
&= \frac{e^{-I_t t}(\lambda I_t)^k}{k!}
\end{aligned}$$

That is, $X(t)$ is Poisson with rate $I_t\lambda$.

To compute the distribution of $Y(t)$, observe

$$\begin{aligned}
\mathbb{P}\left[Y(t) = 1 \mid N(t) = 1\right] &= \mathbb{P}\left[X(t) = 0 \mid N(t) = 1\right] \\
&= 1 - \mathbb{P}\left[X(t) = 1 \mid N(t) = 1\right] \\
&= 1 - \frac{1}{t} I_t \\
&= \frac{1}{t}\int_0^t \overline{G}(s)\, ds
\end{aligned}$$

where $\overline{G}(s) = 1 - G(s)$ is the complementary cdf of the service time. Defining

$$\overline{I}_t \triangleq \int_0^t \overline{G}(s)\, ds$$

and following similar steps as for $X(t)$ we can easily show that

$$\mathbb{P}\left[Y(t) = j\right] = \frac{e^{-\overline{I}_t t}(\overline{I}_t t)^j}{j!}$$

In other words, we can define the customers that are still in the system by time t as "type I" and the customers who leave as "type II". Then, these two types of customers represent a splitting of the Poisson process. □

Generalizations of the Poisson Process

Before finishing this chapter we will quickly review some of the most important extensions of the Poisson process.

The first extension is the **nonhomogeneous Poisson process**. In this case, we relax the stationary increments assumption and we allow the rate to be a function of t. That is, a nonhomogeneous Poisson process $\{N^{nh}(t) : t \geq 0\}$ with intensity function $\lambda(t)$ satisfies:

1. $N^{nh}(0) = 0$
2. $\{N^{nh}(t) : t \geq 0\}$ has independent increments
3. $\mathbb{P}\left[N^{nh}(t+h) - N(t) = 1\right] = \lambda(t)h + o(h)$ and $\mathbb{P}\left[N^{nh}(t+h) - N(h) \geq 2\right] = o(h)$.
4. Define the mean value function

$$m(t) \triangleq \int_0^t \lambda(\tau)\, d\tau$$

   Then, the probability mass function of $N^{nh}(t)$ is Poisson with mean $m(t)$, that is,

$$\mathbb{P}\left[N^{nh}(t) = n\right] = \frac{e^{-m(t)}[m(t)]^n}{n!}$$

5. For every $s, t \geq 0$, $N(t+s) - N(s)$ has Poisson distribution with mean $m(t+s) - m(s) = \displaystyle\int_s^{s+t} \lambda(\tau)\, d\tau$

Nonhomogeneous Poisson processes are very useful to model arrivals because many applications see independent increments, but they have rush hours. Then, it suffices to define $\lambda(t)$ as a piecewise function and the analysis becomes pretty simple.

The second (and last) extension we will cover is called compound Poisson process, and it is defined as follows.

Definition 4.7. *A stochastic process $\{X(t) : t \geq 0\}$ is said to be a compound Poisson process if it can be represented as*

$$X(t) = \sum_{i=1}^{N(t)} Y_i,$$

where $\{N(t) : t \geq 0\}$ is a Poisson process and $\{Y_i : i \in \mathbb{Z}_+\}$ is a family of independent and identically distributed random variables that is also independent of $\{N(t) : t \geq 0\}$.

The compound Poisson process is very useful when we need to assign a cost/reward to each event of a Poisson process. The following properties are particularly useful. If $\{N(t) : t \geq 0$ is a Poisson process with rate λ, then

$$\mathbb{E}\left[X(t)\right] = \lambda t \mathbb{E}\left[Y_1\right] \qquad \text{and} \qquad \mathrm{Var}\left[X(t)\right] = \lambda t \mathbb{E}\left[Y_1^2\right]$$

5 Continuous-Time Markov Chains (CTMC)

[This section is based on Chapter 6 of the textbook]

Continuous-Time Markov Chains (CTMC) are a class of models that have a very wide variety of real-life applications. They are similar to the DTMC's we studied in Section 3, but now we do not require time windows to observe what happened. Instead, we wait until the next event and observe one event at a time. However, in both cases the information about past events can be "summarized" in the present state, that is, they both satisfy the Markovian property. We will see that the time until the next event will be exponential. Therefore, this chapter can be considered as a mix of the last two chapters.

One of the most celebrated examples of CTMC's are queueing models, which we will study deeply in the next chapter. We start defining the model.

5.1 Definition of CTMC

We start with a formal definition.

Definition 5.1. *A stochastic process $\{X(t) : t \geq 0\}$ taking on values in $\mathbb{Z}_+$ is a Continuous-Time Markov Chain if for all $s, t \geq 0$ and $i, j \in \mathbb{Z}_+$, $x(u) \in \mathbb{Z}_+$ for all $0 \leq u < s$, we have*

$$\mathbb{P}\left[X(t+s) = j \mid X(s) = i,\, X(u) = x(u),\, 0 \leq u < s\right] = \mathbb{P}\left[X(t+s) = j \mid X(s) = i\right]$$

In other words, Definition 5.1 says that all the information about the states that the chain visited in the interval $[0, s]$ can be condensed in $X(s)$, that is,

The future depends on the past only through the present state.

Another desirable property is that the transition probabilities only depend on the length of the interval, and not on the "position". We define below.

Definition 5.2. *Let $\{X(t) : t \geq 0\}$ be a CTMC. If, in addition, $\mathbb{P}\left[X(t+s) = j \mid X(s) = i\right]$ is independent of s, we say that the CTMC has stationary (or homogeneous) transition probabilities.*

In this class, all the CTMC's will be assumed to have stationary transition probabilities.

Proposition 5.1. *Suppose $\{X(t) : t \geq 0\}$ is a CTMC with stationary transition probabilities. Then, the time that the chain stays in each state is exponentially distributed.*

We provide a proof sketch for this statement.

Proof. Suppose that the chain enters state i at time 0, and denote T_i the time it stays there. Now let's try to write the stationary transition probabilities property in terms of T_i. Since we want to conclude something about the time the chain stays at state i, let's write the stationary property for $i = j$. We obtain

$$\mathbb{P}\left[X(t+s) = i \mid X(t) = i\right] = \mathbb{P}\left[T_i > t + s \mid, T_i > t\right].$$

Hence, the stationary transition probabilities property is equivalent to saying that the time the chain stays in state i is memoryless. Since we know that the only memoryless distribution is the exponential, we conclude that T_i is exponential for every state i. □

The properties we just reviewed give us an equivalent way to define CTMC's.

Definition 5.3 (equivalent definition of CTMC)**.** *A stochastic process $\{X(t) : t \geq 0\}$ having the properties that each time it enters state i:*

(i) The amount of time it spends in i before making a transition into a different state is exponentially distributed with mean $1/\nu_i$

(ii) When the process leaves state i, it next enters state j with some probability p_{ij} that satisfies

$$p_{i,i} = 0 \qquad \forall i$$
$$\sum_j p_{i,j} = 1$$

In words, this definition says that a CTMC is a DTMC that spends an exponential time in each state before it makes a transition to a different state.

Additionally the time spent in each state is independent of the time spent in the previous and future states.

Let's go through a simple example.

Example 5.1 (Example 6.1 from the textbook)**.** *Consider a shoe shine establishment consisting of two chairs. Upon arrival, a customer goes initially to chair 1, where his shoes are cleaned and polish is applied. After this is done, the customer moves on to chair 2, where the polish is buffed.*

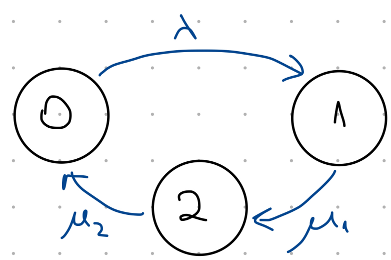
The service time at each chair is exponentially distributed with rate μ_i, $i = 1, 2$. Suppose that potential customers arrive caccording to a Poisson process with rate λ, and a potential customer will enter the system only if both chairs are empty. Can we model this situation as a CTMC? If yes, provide the rates ν_i and the transition probabilities.

Solution. Let $X(t)$ be the chair at which the current customer is at, and let $X(t) = 0$ if there are no customers. Then,

$$\nu_0 = \lambda, \quad \nu_1 = \mu_1, \quad \nu_2 = \mu_2$$

and $p_{0,1} = p_{1,2} = p_{2,0} = 1$.

We may additionally draw the transition diagram for this situation. In this case, however, we write the rate at which the chain goes from one state to the other instead of the probabilities. We obtain



□

5.2 Birth-Death Processes

A very important class of CTMC's are birth-death processes, which we properly define below.

Definition 5.4. *A birth-death process is a CTMC with state space $\mathbb{Z}_+ = \{0, 1, 2, \ldots\}$ for which the transitions from state n can only be to state $n-1$ and $n+1$.*

For each state $n \in \mathbb{Z}_+$, we use λ_n to denote the rate at which the chain goes from state n to $n+1$ and μ_n to the rate at which it goes from n to $n-1$. The parameters $(\lambda_n)_{n=0}^{\infty}$ are called birth or arrival rates and $(\mu_n)_{n=1}^{\infty}$ are called death or departure rates.

Observe that in a birth-death process, we have an arrival/birth if the minimum time until a birth and a death is the time until a birth. Similarly for departures/deaths. Therefore, we can compute the transition probabilities using the property of exponential distributed random variables that says: If $X_i \sim Expo(\alpha_i)$ for $i = 1, 2$, then

$$\mathbb{P}[X_1 < X_2] = \frac{\alpha_1}{\alpha_1 + \alpha_2}$$

Indeed,

$$\begin{aligned} p_{0,1} &= 1 \\ p_{n,n+1} &= \frac{\lambda_n}{\lambda_n + \mu_n} \qquad \forall n \geq 1 \\ p_{n,n-1} &= \frac{\mu_n}{\lambda_n + \mu_n} \qquad \forall n \geq 1 \end{aligned}$$

Let's see some examples.

Example 5.2 (Example 6.2 from the textbook)**.** *The Poisson process is a birth-death process with* $\lambda_n = \lambda$ *and* $\mu_n = 0$ *for all* n*.*

Since the death rate is 0 for all the states in the above example, it is called a pure birth process.

Example 5.3 (Example 6.5 from the textbook)**.** ***The*** $M/M/1$ ***queue:*** *Customers arrive to a single-server station according to a Poisson process with rate* λ*. Upon arrival, each customer goes directly into service if the server is free. If not, then the customer joins the queue. When the server finishes serving a customer, the customer leaves the system adn the next customer in line (if any) enters the service.*

The successive service times are assumes to be independent exponential random variables with rate μ*.*

Model this system as a birth-death process.

Solution. We let $X(t)$ be the number of customers in the system (considering the one in service, if any). Then, $\{X(t) : t \geq 0\}$ is a birth-death process with $\lambda_n = \lambda$ for all $n \geq 0$ and $\mu_n = \mu$ for all $n \geq 1$. □

The notation $M/M/1$ above means that the arrivals are Markovian, the service times are Markovian and there is one server.

Example 5.4 (Example 6.6 from the textbook)**.** ***The M/M/s system:*** *Consider now a multiple-server station. Customers arrive according to a Poisson process with rate* λ*. If one of the* s *servers is available, then the customer immediately starts service with one of them. If not, the customer joins a centralized queue, that is, there is a single waiting line for all the customers.*

All the servers have the same service rate μ*, and the service times are exponentially distributed.*

Model this system as a birth-death process.

Solution. We again define the state $X(t)$ as the number of customers in the system. However, now the rates are different. We have

$$\lambda_n = \lambda \qquad \forall n \geq 0$$
$$\mu_n = \begin{cases} n\mu & \text{if } n \leq s \\ s\mu & \text{if } n > s \end{cases}$$

□

We end this section with the following proposition.

Proposition 5.2. *Consider a birth-death process with birth rates* $(\lambda_n)_{n\geq 0}$ *and death rates* $(\mu_n)_{n\geq 1}$*. Let* T_i *be the time, starting from state* i*, it takes for the process to enter state* $i+1$*. Then,*

$$\mathbb{E}[T_0] = \frac{1}{\lambda_0} \qquad \textit{and} \qquad \mathbb{E}[T_i] = \frac{1}{\lambda_i} + \frac{\mu_i}{\lambda_i}\mathbb{E}[T_{i-1}] \quad \forall i \geq 1$$

Additionally, for $k < j$*, the expected time to reach state* j *starting from state* k *is*

$$\mathbb{E}[\textit{time to go from } j \textit{ to } k] = \sum_{i=k}^{j-1} \mathbb{E}[T_i]$$

Proof. First observe that $\mathbb{E}[\text{time to go from } j \text{ to } k] = \sum_{i=k}^{j-1} \mathbb{E}[T_i]$ holds by definition of T_i because, in a birth-death process, going from state k to j implies that we need to pass through all the states in between.

Now we show the recursive equation. For state 0 there is nothing to prove because the chain can only go to state 1 if it's currently at state 0, and the expected time spent in state 0 is $1/\lambda_0$.

To prove the result for $i \geq 1$ we condition on the first transition after leaving state i. Define

$$I_i = \begin{cases} 1 & \text{if the first transition from } i \text{ is to } i+1 \\ 0 & \text{if the first transition from } i \text{ is to } i-1 \end{cases}$$

Then, since the expected time spent at state i is always $1/(\lambda_i + \mu_i)$, we obtain

$$\mathbb{E}[T_i \mid I_i = 1] = \frac{1}{\lambda_i + \mu_i}$$

$$\mathbb{E}[T_i \mid I_i = 0] = \frac{1}{\lambda_i + \mu_i} + \mathbb{E}[T_{i-1}] + \mathbb{E}[T_i]$$

where the last equality holds because if $I_i = 0$, then the chain goes from state i to $i-1$. Then, the time to go to state $i+1$ will be the time until it goes back to state i (represented by $\mathbb{E}[T_{i-1}]$) and the time until it goes to $i+1$ (represented by $\mathbb{E}[T_i]$).

Then, using the tower property of the expectation, we obtain

$$\begin{aligned}\mathbb{E}[T_i] &= \mathbb{E}[T_i \mid I_i = 1]\,\mathbb{P}[I_i = 1] + \mathbb{E}[T_i \mid I_i = 0]\,\mathbb{P}[I_i = 0] \\ &= \left(\frac{1}{\lambda_i + \mu_i}\right)\left(\frac{\lambda_i}{\lambda_i + \mu_i}\right) + \left(\frac{1}{\lambda_i + \mu_i} + \mathbb{E}[T_{i-1}] + \mathbb{E}[T_i]\right)\left(\frac{\mu_i}{\lambda_i + \mu_i}\right)\end{aligned}$$

Reorganizing terms we obtain the result. □

Example 5.5. *Consider an $M/M/1$ queue. Compute the expected time until the queue grows by one customer, for every state i.*

Solution. In this case, $\lambda_i = \lambda$ for all $i \geq 0$ and $\mu_i = \mu$ for all $i \geq 1$. Then,

$$\begin{aligned}\mathbb{E}[T_0] &= \frac{1}{\lambda} \\ \mathbb{E}[T_1] &= \frac{1}{\lambda}\left(1 + \frac{\mu}{\lambda}\right) \\ \mathbb{E}[T_2] &= \frac{1}{\lambda}\left(1 + \frac{\mu}{\lambda} + \frac{\mu^2}{\lambda^2}\right) \\ &\vdots \\ \mathbb{E}[T_i] &= \frac{1}{\lambda}\left(1 + \frac{\mu}{\lambda} + \frac{\mu^2}{\lambda^2} + \cdots + \frac{\mu^i}{\lambda^i}\right) = \frac{1 - (\mu/\lambda)^{i+1}}{\mu - \lambda} \qquad \forall i \geq 1\end{aligned}$$

□

The variance can be computed similarly, but we will skip it in this class. We will now start studying properties of any CTMC, and we will keep using the birth-death process as an application.

5.3 The transition probability function $p_{i,j}(t)$

When we studied DTMC's, all the information about the next state probability mass function was condensed in the transition matrix P, which had elements $p_{i,j}$ representing the probability of going to state j if we are currently at state i. Indeed, when we defined CTMC's we also used this definition of $p_{i,j}$. However, we are also interested in the time until the next event, we need a more sophisticated definition of the transition probabilities.

Definition 5.5. *The transition probabilities of a CTMC, denoted by $p_{i,j}(t)$ represent the probability that a process presently in state i will be in state j a time t later.*

If t is very small, then $p_{i,j}(t)$ will be positive only if the chain can visit state j immediately after state i. However, for larger values of t the function $p_{i,j}(t)$ is more complicated. Indeed, for larger values of t we need to consider the time the chain remains in state i, all the possible transitions from state i, the time the chain spends in the new state, etc. Even for a pure-birth process, finding an explicit expression for $p_{i,j}(t)$ is complicated.

In the rest of this section we focus on building a set of differential equations that the transition probabilities $p_{i,j}(t)$ must satisfy. We start with some definitions and lemmas.

Definition 5.6. *Consider a CTMC with ν_i representing the mean time spent at state i and $p_{i,j}$ the probability of visiting state j immediately after state i. The instantaneous transition rate at which the process makes a transition from state i to j $(i \neq j)$ is*

$$q_{i,j} \triangleq \nu_i p_{i,j}$$

Observe that

$$\nu_i = \sum_j q_{i,j} \qquad \text{and} \qquad p_{i,j} = \frac{q_{i,j}}{\nu_i}$$

Hence, $q_{i,j}$ for all i, j completely defines the CTMC.

The following lemmas relate $p_{i,j}(t)$ to $q_{i,j}$ and will help us to work with $p_{i,j}(t)$.

Lemma 5.3 (Lemma 6.2 from the textbook)**.** *Consider a CTMC with instantaneous transition rates $q_{i,j}$, mean time spent at state i ν_i, and transition probabilities $p_{i,j}(t)$. Then,*

$$\lim_{h\to 0} \frac{1 - p_{i,i}(h)}{h} = \nu_i \qquad \text{and} \qquad \lim_{h\to 0} \frac{p_{i,j}(h)}{h} = q_{i,j}$$

Proof. We will skip the proof of the lemma, but let us analyze what it says intuitively. First, notice that $1 - p_{i,i}(h)$ represents the probability that the chain leaves state i in h units of time to any other state $j \neq i$. In other words, $1 - p_{i,i}(h)$ is the probability that a transition occurs before h. Since the time until the next transition is $Expo(\nu_i)$, the probability that there is at least one event in h time units is $\nu_i h + o(h)$ (relating to the Poisson process order property).

Similarly, $p_{i,j}(h)$ represents the probability of leaving state i and being in state j after h time units. As h gets small, the probability of making more than one transition rapidly decreases to zero. Hence, $p_{i,j}(h)$ approaches the rate at which the process leaves the current state and goes to state j, that is, $q_{i,j}$. ☐

Lemma 5.4 (Lemma 6.3 from the textbook)**.** ***Chapman-Kolmogorov equations:*** *Consider a CTMC, and let $s, t \geq 0$. Then,*

$$p_{i,j}(t+s) = \sum_k p_{i,k}(t) p_{k,j}(s)$$

Similarly to DTMC's, this property is true due to the Markovian property. Hence, we can easily prove it using the law of total probability and conditioning on $X(t)$.

Now that we know how to deal with transitions in small intervals of time, and we know that Chapman-Kolmogorov equations hold, we are ready for our system of differential equations.

Theorem 5.5 (Theorem 6.1 from the textbook)**.** ***Kolmogorov's backward equations:*** *For all states i, j and times $t \geq 0$, we have*

$$p'_{i,j}(t) = \sum_{k\neq i} q_{i,k} p_{k,j}(t) - \nu_i p_{i,j}(t)$$

The proof comes immediately from the definition of derivative and the previous two lemmas.

Proof. By definition of derivative, we have

$$p'_{i,j}(t) = \lim_{h\to 0} \frac{p_{i,j}(t+h) - p_{i,j}(t)}{h}$$

The numerator can be reorganized as follows using Chapman-Kolmogorov equations:

$$\begin{aligned} p_{i,j}(t+h) - p_{i,j}(t) &= \sum_k p_{i,k}(h) p_{k,j}(t) - p_{i,j}(t) \\ &= \sum_{k\neq i} p_{i,k}(h) p_{k,j}(t) + p_{i,i}(h) p_{i,j}(t) - p_{i,j}(t) \\ &= \sum_{k\neq i} p_{i,k}(h) p_{k,j}(t) - p_{i,j}(t) \left(1 - p_{i,i}(h)\right) \end{aligned}$$

The rest of the proof holds using Lemma 5.3 in the limit. ☐

Example 5.6 (Example 6.10 from the textbook)**.** *Compute the backward equations for:*

(a) A pure-birth process

(b) A birth-death process

Solution.

(a) In a pure-birth process, $p_{i,j} = 1$ if $j = i + 1$ and 0 otherwise. Similarly, $q_{i,i+1} = \lambda_i = \nu_i$. Then, we obtain

$$p'_{i,j}(t) = \lambda_i p_{i+1,j}(t) - \lambda_i p_{i,j}(t)$$

(b) In a birth-death process, we have

$$\nu_i = \lambda_i + \mu_i, \quad q_{i,i+1} = \lambda_i, \quad q_{i,i-1} = \mu_i, \quad p_{i,i+1} = \frac{\lambda_i}{\lambda_i + \mu_i}, \quad \text{and} \quad p_{i,i-1} = \frac{\mu_i}{\lambda_i + \mu_i}.$$

for all $i \geq 0$, where $\mu_0 = 0$. Then, the backward equations are:

$$\begin{aligned} p'_{0,j}(t) &= \lambda_0 p_{1,j}(t) - \lambda_0 p_{0,j}(t) \\ p'_{i,j}(t) &= \lambda_i p_{i+1,j}(t) + \mu_i p_{i-1,j}(t) - (\lambda_i + \mu_i) p_{i,j}(t) \end{aligned}$$

□

The Kolmogorov backward equations completely determine the behavior of the CTMCs in short periods of time. Indeed, for many systems one can easily solve the equations and get information about the state of the chain at, say $t = 10$. However, in this class we are more interested in the long-run behavior of the CTMCs, so we will not focus on solving the transient.

When we derived Kolmogorov's backward equations, we wrote

$$p_{i,j}(t + h) = \sum_k p_{i,k}(h) p_{k,j}(t)$$

However, we can equivalently write

$$p_{i,j}(t + h) = \sum_k p_{i,k}(t) p_{k,j}(h)$$

This change in the use of Chapman-Kolmogorov equations leads to the following theorem.

Theorem 5.6 (Theorem 6.2 from the textbook)**.** *Under suitable regularity conditions,*

$$p'_{i,j}(t) = \sum_{k \neq j} q_{k,j} p_{i,k}(t) - \nu_j p_{i,j}(t)$$

Example 5.7 (Example 6.10 from the textbook)**.** *Compute the forward equations for:*

(a) A pure-birth process

(b) A birth-death process

Solution. For the pure-birth process we obtain

$$p'_{i,j}(t) = \lambda_{j-1} p_{i,j-1}(t) - \lambda_j p_{i,j}(t)$$

Since in a pure-birth process we can only go from states i to states $j > i$, we obtain

$$p'_{i,i}(t) = -\lambda_i p_{i,i}(t) \qquad \text{and} \qquad p'_{i,j}(t) = \lambda_{j-1} p_{i,j-1}(t) - -\lambda_j p_{i,j}(t) \quad j \geq +1$$

For a birth-death process we obtain

$$\begin{aligned} p'_{i,0}(t) &= \mu_1 p_{i,1}(t) - \lambda_0 p_{i,0}(t) \\ p'_{i,j}(t) &= \lambda_{j-1} p_{i,j-1}(t) + \mu_{j+1} p_{i,j+1}(t) - (\lambda_j + \mu_j) p_{i,j}(t) \end{aligned}$$

□

5.4 Limiting probabilities

Analogously to DTMCs, we want to compute the probability of being in state j in the long run, independently of where the process started. In other words, we want to compute the value of

$$\pi_j = \lim_{t\to\infty} p_{i,j}(t) \qquad \forall j$$

We expect that in the long run, the probabilities $p_{i,j}(t)$ do not change. Then, $p'_{i,j}(t) = 0$ for large t. Then, if we take the limit in the forward equations we obtain:

$$\begin{aligned}
& \lim_{t\to\infty} p'_{i,j}(t) = \sum_{k\neq j} q_{k,j} \lim_{t\to\infty} p_{i,k}(t) - \nu_j \lim_{t\to\infty} p_{i,j}(t) \\
\implies \quad & 0 = \sum_{k\neq j} q_{k,j}\pi_k - \nu_j\pi_j \\
\implies \quad & \nu_j\pi_j = \sum_{k\neq j} q_{k,j} \qquad (6)
\end{aligned}$$

Let's analyze the set of equations (6). The left-hand side represents the rate at which we leave state j, and the right-hand side the rate at which we enter state j. Hence, the equations just say that in the long run, the rate at which we enter a state equals the rate at which we leave the state. Indeed, the equations (6) are called balance equations.

Observe that in the argument above, we interchanged the limit and the summation, and this step cannot always be performed. We won't spend a lot of time on when we can do it. It suffices to know that if the CTMC is irreducible (all states communicate) and is positive recurrent, then we can interchange the limit and the summation. Hence, for irreducible and positive recurrent chains, the balance equations give us the limiting probability.

Theorem 5.7. *In an irreducible and positive recurrent CTMC, solving the balance equations and $\sum_j \pi_j = 1$ give the limiting probabilities π_j that the chain is in a state j in the long run.*

Observe that in DTMC's we also required aperiodicity. In this case, we do not need that because we are also studying the time of permanence in each state (which is exponential).

Example 5.8. *Compute the limiting probabilities in a birth-death process.*

Solution. The balance equations are:

State	Rate at which leave	=	Rate at which enter
0	$\lambda_0\pi_0$	=	$\mu_1\pi_1$
1	$(\lambda_1+\mu_1)\pi_1$	=	$\mu_2\pi_2 + \lambda_0\pi_0$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$n \geq 1$	$(\lambda_n+\mu_n)\pi_n$	=	$\mu_{n+1}\pi_{n+1} + \lambda_{n-1}\pi_{n-1}$

Reorganizing terms, we obtain

$$\lambda_n\pi_n = \mu_{n+1}\pi_{n+1} \qquad \forall n \geq 0$$

Hence, using that $\sum_j \pi_j = 1$ we obtain

$$\begin{aligned}
\pi_0 &= \left(1 + \sum_{n=1}^{\infty}\prod_{i=1}^{n}\frac{\lambda_{i-1}}{\mu_i}\right)^{-1} \\
\pi_n &= \left(\prod_{i=1}^{n}\frac{\lambda_{i-1}}{\mu_i}\right)\pi_0
\end{aligned}$$

□

Example 5.9 (Example 6.14 from the textbook)**.** *Compute the limiting probabilities of an $M/M/1$ queue.*

Solution. We may use the result from the birth-death process with $\lambda_n = \lambda$ for all $n \geq 0$ and $\mu_n = \mu$ for all $n \geq 1$. However, let's solve the problem from scratch to practice using the balance equations. Observe that all the states are similar, except for state 0. Then, we have:

State	Rate at which leave	=	Rate at which enter
0	$\lambda\pi_0$	=	$\mu\pi_1$
1	$(\lambda+\mu)\pi_1$	=	$\lambda\pi_0 + \mu\pi_2$
2	$(\lambda+\mu)\pi_2$	=	$\lambda\pi_1 + \mu\pi_3$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$i \geq 1$	$(\lambda+\mu)\pi_i$	=	$\lambda\pi_{i-1} + \mu\pi_{i+1}$

Now we solve the system of equations, including the equation

$$\sum_{i=0}^{\infty} \pi_i = 1$$

From the balance equation for state 0, we obtain

$$\pi_1 = \frac{\lambda}{\mu}\pi_0$$

Now, if we add the equations for state 0 and 1, we obtain

$$\begin{aligned} & \lambda\pi_0 + \lambda\pi_1 + \mu\pi_1 = \mu\pi_1 + \lambda\pi_0 + \mu\pi_2 \\ \implies \quad & \pi_2 = \frac{\lambda}{\mu}\pi_1 = \left(\frac{\lambda}{\mu}\right)^2 \pi_0 \end{aligned}$$

where we used that $\pi_1 = (\lambda/\mu)\pi_0$ in the last equality. Now if we add the equations for states 1 and 2, we obtain

$$\begin{aligned} & \lambda\pi_1 + \mu\pi_1 + \lambda\pi_2 + \mu\pi_2 = \lambda\pi_0 + \mu\pi_2 + \lambda\pi_1 + \mu\pi_3 \\ \implies \quad & \mu\pi_3 = \mu\pi_1 + \lambda\pi_2 - \lambda\pi_0 \\ \implies \quad & \pi_3 = \left(\frac{\lambda}{\mu}\right)^3 \pi_0 \end{aligned}$$

where we used that $\pi_1 = (\lambda/\mu)\pi_0$ and $\pi_2 = (\lambda/\mu)^2\pi_0$ to obtain the last equation.

If we repeat these steps, we can easily conclude that

$$\pi_i = \left(\frac{\lambda}{\mu}\right)^i \pi_0 \qquad \forall i \geq 1$$

To obtain the value of π_0 we use that $\sum_i \pi_i = 1$. We obtain

$$\begin{aligned} & 1 = \sum_{i=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^i \pi_0 = \frac{1}{1-\frac{\lambda}{\mu}} \qquad \text{if } \lambda < \mu \text{, solving the geometric sum} \\ \implies \quad & \pi_0 = 1 - \frac{\lambda}{\mu} \end{aligned}$$

Hence, we obtain

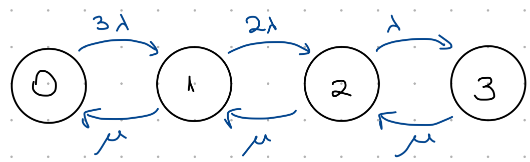
$$\pi_i = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^i \qquad \forall i \geq 0$$

□

Let's do one more example.

Example 5.10 (Example 6.13 from the book)**.** *Consider a job shop that consists of $M = 3$ machines and one serviceman. Suppose that the amount of time each machine runs before breaking down is exponentially distributed with mean $1/\lambda$ and suppose that the amount of time that it takes to the service man to fix a machine is exponentially distributed with mean $1/\mu$. What is the average number of machines not in use?*

Solution. Let's first specify the model. Define $X(t)$ as the number of machines that are damaged (not in use) at time t. Then, the transition diagram with the corresponding rates at which we go from one state to another is:



The repair rate is always μ because there is only one serviceman. However, the time until the next machine fails depends on how many machines are working. If 3 machines are working (0 not in use), the time until the next failure is the minimum of three $Expo(\lambda)$ random variables and, hence, it is an $Expo(3\lambda)$ random variable. Similarly, when 2 machines are working, the time until the next failure is $Expo(2\lambda)$.

Now we compute the balance equations. We obtain

State	Rate at which leave	=	Rate at which enter
0	$3\lambda\pi_0$	=	$\mu\pi_1$
1	$(2\lambda+\mu)\pi_1$	=	$3\lambda\pi_0+\mu\pi_2$
2	$(\lambda+\mu)\pi_2$	=	$2\lambda\pi_1+\mu\pi_3$
3	$\mu\pi_3$	=	$\lambda\pi_2$

We may solve the balance equations, or we may use the formula we obtained for the birth-death process. Let's do the last one. We obtain

$$\pi_0 = \left(1+\frac{3\lambda}{\mu}+\frac{(3\lambda)(2\lambda)}{\mu^2}+\frac{(3\lambda)(2\lambda)\lambda}{\mu^3}\right)^{-1} = \left(1+3\frac{\lambda}{\mu}+6\frac{\lambda^2}{\mu^2}+6\frac{\lambda^3}{\mu^3}\right)^{-1}$$

$$\pi_1 = \frac{3\lambda}{\mu}\pi_0$$

$$\pi_2 = \frac{6\lambda^2}{\mu^2}\pi_0$$

$$\pi_3 = \frac{6\lambda^3}{\mu^3}\pi_0$$

Therefore, the average of machines not in use is:

$$\mathbb{E}[X] = \sum_{i=0}^{3} i\pi_i = \pi_0\left(\frac{3\lambda}{\mu}+\frac{12\lambda^2}{\mu^2}+\frac{18\lambda^3}{\mu^3}\right)$$

□

5.5 Uniformization

When we defined the transition probabilities $p_{i,j}(t)$ we showed that they satisfy a differential system of equations that we should solve. However, solving these equations might be considerably hard. In this short section we will learn an alternative to compute these probabilities. We start with a simplified version of the computation.

Consider a CTMC in which the mean time spent at each state is constant, that is, $\nu_i = \nu$ for all i. Then, the amount of time spent at each of the states is an exponential random variable with rate ν. Further, the number of transitions in the interval $(0, t)$ represents a Poisson process with rate ν. Let's use this information to compute $p_{i,j}(t)$. We condition on the number of transitions $N(t)$, and obtain

$$\begin{aligned} p_{i,j}(t) &= \mathbb{P}[X(t)=j \mid X(0)=i] \\ &= \sum_{n=0}^{\infty} \mathbb{P}[X(t)=j \mid X(0)=i, N(t)=n]\,\mathbb{P}[N(t)=n] \end{aligned}$$

We know exactly what is $\mathbb{P}[N(t)=n]$, since $\{N(t) : t \geq 0\}$ is a Poisson Process. To compute $\mathbb{P}[X(t)=j \mid X(0)=i, N(t)=n]$ observe that the expected time spent in each of the n states the chain visits in $(0, t)$ is the same. Then, it does not matter in which order or how much time we actually spent after each transition and we can think of these n transitions as the transitions of a DTMC with transition probabilities $p_{i,j}$. Therefore,

$$p_{i,j}(t) = \sum_{n=0}^{\infty} p_{i,j}^{(n)} \left(\frac{e^{-\nu t}(\nu t)^n}{n!} \right)$$

Now let's generalize to CTMC's where ν_i might be different for each state and let ν be any number satisfying

$$\nu_i \leq \nu \qquad \forall i$$

Inspired by the splitting of a Poisson process properties, when the chain is at state i, we can pretend that the time until the next potential transition is an exponential random variable with rate ν, but only a fraction ν_i/ν of the transitions occur. Then, the time until the next transition would be an exponential random variable with rate $\nu \times \nu_i/\nu = \nu_i$. Hence, the CTMC can be thought of as a process that spends an $Expo(\nu)$ time in each state i and makes transitions to state j with probability

$$\overline{p}_{i,j} = \begin{cases} 1 - \frac{\nu_i}{\nu} & \text{, if } i = j \\ \frac{\nu_i}{\nu} p_{i,j} & \text{, if } i \neq j \end{cases}$$

Using our previous computation, we have that the transition probabilities of any CTMC can be computed as

$$p_{i,j}(t) = \sum_{n=0}^{\infty} \overline{p}_{i,j}^{(n)} \frac{e^{-\nu t}(\nu t)^n}{n!}$$

The DTMC defined by the transition probabilities $\overline{p}_{i,j}$ is often called the <u>embedded DTMC</u>, and the technique of *uniformizing* the mean time spent in each state is called <u>uniformization</u>.

Observe that uniformization gives a natural method to simulate CTMC's.

Example 5.11 (Example 6.23 from the textbook)**.** *Consider a two-state CTMC with $p_{0,1} = p_{1,0} = 1$, and rates $\nu_0 = \lambda$, $\nu_1 = \mu$. Compute the transition probabilities using uniformization.*

Solution. It suffices to use $\nu = \lambda + \mu$. Then,

$$\overline{p}_{0,0} = 1 - \frac{\lambda}{\lambda + \mu} = \frac{\mu}{\lambda + \mu}$$
$$\overline{p}_{0,1} = \frac{\lambda}{\lambda + \mu}$$
$$\overline{p}_{1,0} = \frac{\mu}{\lambda + \mu}$$
$$\overline{p}_{1,1} = 1 - \frac{\mu}{\lambda + \mu} = \frac{\lambda}{\lambda + \mu}$$

Observe that $\overline{p}_{0,0} = \overline{p}_{1,0} = \mu/(\lambda+\mu)$. Then, the probability of transitioning to state 0 is $\mu/(\lambda+\mu)$ no matter what is the current state. Similarly, the probability of transitioning to state 1 is $\lambda/(\lambda + \mu)$. Hence, for all $n \geq 1$ and $i = 0, 1$ we have

$$\overline{p}_{i,0}^{(n)} = \frac{\mu}{\lambda + \mu} \qquad \text{and} \qquad \overline{p}_{i,1}^{(n)} = \frac{\lambda}{\lambda + \mu}$$

Therefore, applying the uniformization formula we obtain

$$p_{0,0}(t) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)t}$$
$$p_{1,1}(t) = \frac{\lambda}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} e^{-(\lambda+\mu)t}$$
$$p_{0,1}(t) = 1 - p_{0,0}(t) = \frac{\lambda}{\lambda + \mu} \left(1 - e^{-(\lambda+\mu)t}\right)$$
$$p_{1,0}(t) = 1 - p_{1,1}(t) = \frac{\mu}{\lambda + \mu} \left(1 - e^{-(\lambda+\mu)t}\right)$$

□

6 Simulation

[This chapter is based on the lecture note "Introduction to Discrete-Event Simulation and SimPy Language" by Norm Marloff, Chapters 1 and 2. Available here.]

So far in the semester we have been studying theoretical analysis of some stochastic processes, such as Discrete and Continuous-time Markov chains, and the Poisson process. We have seen that we can model a very wide variety of real-life problems using these stochastic processes. Further, we can provide explicit mathematical expressions to understand their behavior. However, the analysis becomes difficult (or even intractable) quickly as the number of possible states increases.

In this chapter we will quickly review another approach to analyze systems: Simulation. We will still use the models we learned before to better understand the nature of our real-life systems, and to decide which variables are important. However, we will not construct expressions anymore. Now we will determine what is important to model, and build some code that will represent the main characteristics of our system.

We will center on Discrete-Event Simulation. The main idea is that we have a system that evolves in time, and we need to detect what changes are relevant. These changes, are called events. To make it easier, let's use an example throughout the chapter.

Example 6.1. *We will consider an $M/M/1$ queue. That is, a single-server queue where the inter-arrival times are exponential with rate λ, and the processing time of each customer is exponential with rate μ.*

Our goal in this chapter is to construct a simulation that will tell us the mean waiting time of the customers that enter the system in an interval $[0, T]$.

6.1 Components

There are four main components of a discrete-event simulation, which we list and define below.

(i) **State:** Variables that capture the main characteristics of the system. In other words, the state answers the question: what are we observing from the system?

(ii) **Clock:** We need a variable that keeps track of time, and can alert us of when there are changes in the state(s).

(iii) **Events:** Depending on the problem, there might be different things that can happen and affect the state of the system. The most common events in stochastic networks are arrivals and departures.

(iv) **Ending condition:** Unfortunately, the simulation cannot run forever. Then, we need to establish when the simulation ends. Some common ending conditions are a simulation horizon, or a number of events to happen.

When we build the simulation, we must make sure that we are also able to generate random numbers and do some output analysis (statistics). Most languages have packages that do that, so we will not focus on those.

In the example of the single-server queue, we have the following components:

(i) State: Number of customers in the system

(ii) Clock: When we build the simulation, we just need to keep track of time

(iii) Events set: Arrivals and departures

(iv) Ending condition: Until the clock hits T.

6.2 Simulation paradigms

There are different ways to carry out discrete-event simulations. In this section we discuss three paradigms, all sharing the four components described above.

6.2.1 Fixed-increment time (or activity oriented) paradigm

As the name says, under this paradigm we discretize time. Then, we keep track of what happened in each time slot. Under this approach, we model our system as a DTMC.

Below we provide a pseudo-code associated to the single-server queue example. Suppose that the set of time slots to revise is $\mathcal{T}$, and recall that we want to study the average waiting time. Then, we need to keep track of the waiting time of each customer and the number of customers we process. One way to keep track of the waiting times is to keep an array that represents the queue, and in each entry, save the arrival time of each customer. Of course, we need to update the queue whenever a customer leaves and we should estimate how long the queue will be to make sure we have enough slots in the array.

0. **Initialization:**
   Set time slots $\mathcal{T}$
   Set clock $t \leftarrow 0$, state $q \leftarrow 0$
   Generate next arrival: $a \leftarrow Expo(\lambda)$
   Generate next departure: $s \leftarrow \infty$
   Initialize queue vector: $q_{vec} \leftarrow (0, 0, 0, 0, 0...)$
   Initialize cumulative waiting time $w \leftarrow 0$ and total number of processed jobs $n \leftarrow 0$

1. **Simulation loop:**
   For every $t \in \mathcal{T}$:
   - If $t = a$ (that is, if there is an arrival),
     Add new customer to the queue: $q \leftarrow q + 1$, $q_{vec}(q) \leftarrow t$
     Generate new arrival time: $a \leftarrow t + Expo(\lambda)$
     If the server was empty, start a new server: If $q = 1$, generate $s \leftarrow Expo(\mu)$
   - If $t = s$ (that is, if there is a departure),
     Update statistics:
     - Cumulative waiting time: $w \leftarrow w + q_{vec}(1)$
     - Number of customers: $n \leftarrow n + 1$

     Update queue: for every $i \in \{1, \dots, q\}$, do $q_{vec}(i) \leftarrow q_{vec}(i + 1)$
     Update number in system: $q \leftarrow q - 1$
     Generate next departure: If $q > 0$, generate $s \leftarrow Expo(\mu)$. Otherwise, $s \leftarrow \infty$.

2. **End:** Return w/n as mean waiting time.

Observe that we are looping through all the elements of $\mathcal{T}$. Then, if our discretization is too fine, nothing will happen in most of them. Also, the discretization should be so fine that at most one event happens in each slot. We can also simulate in discrete time, as in a DTMC. However, we lose too much information because multiple events can happen in the same slot.

We can avoid all this trouble if, instead, we follow the next-event time (a.k.a. event-oriented) simulation paradigm.

6.3 Next-event time (a.k.a. event-oriented) paradigm

Under this paradigm we do not discretize the simulation horizon. Instead, we immediately advance the clock until the next event. That is, we check the minimum time of occurrence among the events in the event set, and we advance the clock. Then, we update the system according to what is the event (arrival or departure in the example), and we generate the next event.

The following flowchart shows a pesudocode.

Advantages of the next-event time paradigm are its simplicity and flexibility. These models are very easy to code and modify for debugging or to evaluate different scenarios. However, they can only model discrete events and state spaces. If we wanted to observe the evolution of the system while a customer is being served, for example, we cannot use the next-event time paradigm. Instead, we have to use the continuous simulation paradigm.

6.4 Continuous simulation (a.k.a. process-oriented) paradigm

Here the model is very similar to the next-event time paradigm. The main difference is that now we build the simulation in modules, and we need an additional "main" module that manages the simulation.

The main advantage of this paradigm is the flexibility to model discrete and continuous events. However, it is not as easy and flexible as the next-event time paradigm. Most of the commercial simulation softwares (such as Arena or SIMIO) use this paradigm, but we won't use this paradigm in this class.

7 Queueing Theory

[In this chapter we take some elements of Chapter 8 of the book]

In this chapter we cover a quick overview of nonmarkovian queueing systems. We have been using queues with Poisson arrivals and exponential service times as examples of CTMC's. However, these models might be idealistic in some scenarios. In this chapter we will cover a quick overview of what can be done in cases of a general distribution of interarrival or service times (or both). We start with some preliminary overviews.

7.1 Preliminaries

One of the most beautiful equations in queueing theory is Little's law. Its beauty is due to its simplicity and usefulness. Let's define some notation and we will then present the equation.

Define:

L	$=$	Average number of customers in the system
L_Q	$=$	Average number of customers waiting in line
W	$=$	Average time a customer spends in the system
W_Q	$=$	Average time a customer spends waiting in the queue
λ	$=$	Average arrival rate of entering customers

Then, Little's law establishes:

$$L = \lambda W \qquad \text{and} \qquad L_Q = \lambda W_Q$$

These very simple equations are valid for most queueing systems, regardless of the arrival process, service time distribution, number of servers and queue discipline.

The next property is another reason why the Poisson process is so important.

Proposition 7.1 (PASTA: Poisson Arrivals See Time Averages)**.** *Let π_n be the limiting probability of having n customers in the system, and a_n be the proportion of customers that encounter n in the system upon arrival. Then, if the arrivals occur according to a Poisson process, we have*

$$\pi_n = a_n \qquad \forall n$$

The proof relies on the independent increments property. The main idea is that having an arrival at time t gives no information about the past. Then, the conditional distribution of what an arrival sees is equal to the unconditional distribution of the state of the system.

PASTA seems like a reasonable property for most distributions. However, not all distributions satisfy it. Let's see a short counter-example.

Example 7.1 (Example 8.1 from the textbook)**.** *Consider a queueing model in which all the customers have service time equal to 1, and the inter-arrival times are uniformly distributed in $(1, 2)$.*

The interarrival times are always greater than the service times. Hence,

$$a_0 = 1$$

However,

$$\pi_0 < 1$$

because the system is not always empty.

Here, PASTA fails because if there is an arrival at t, we immediately know that there was no arrival in the interval $(t-1, t)$. That is, having an arrival at t gives us information about the past.

Let's see another example. This time, we use PASTA to simplify computations.

Example 7.2 (Example 8.2 from the textbook)**.** *People arrive at a bus stop according to a Poisson process with rate λ. Buses arrive at the stop according to a Poisson process with rate μ, and each arriving bus picks up all the currently waiting people.*

(a) What is the mean number of customers waiting at the bus stop, at any time?

(b) What is the mean number of customers that each bus picks up?

Solution.

(a) We use Little's law. Using the notation introduced before, we are trying to compute L_Q.

The time between buses is $Expo(\mu)$ and, since the exponential distribution is memoryless, the average time between buses equals the average time each customer waits. Then,

$$W_Q = \frac{1}{\mu}$$

Therefore, the average number of customers waiting in the bus stop is

$$L_Q = \lambda W_Q = \frac{\lambda}{\mu}$$

(b) Now let

$$\begin{aligned} X_i &= \text{Number of people picked up by the } i^{\text{th}} \text{ bus} \\ T_i &= \text{Time between } (i-1)^{\text{th}} \text{ and } i^{\text{th}} \text{ buses} \end{aligned}$$

Then, we know

$$\begin{aligned} \mathbb{E}[X_i] &= \mathbb{E}[\mathbb{E}[X_i \mid T_i]] \qquad \text{(tower property of expectation)} \\ &= \mathbb{E}[\lambda T_i] \qquad \text{(mean number of events of a Poisson process)} \\ &= \frac{\lambda}{\mu} \end{aligned}$$

Observe that both numbers are equal, even though the random variables are different. In the first case, we compute the average number of people in a bus stop at any time. In the second case, we only look at the number of customers exactly when the bus arrives. Since the bus arrivals are Poisson, both quantities are exactly the same.

□

When we drop the exponential assumption for inter-arrival or service times, there is not much we can do. In other words, the problems rapidly become intractable. In the next section we illustrate the $M/G/1$ queue.

7.2 The $M/G/1$ queue

The $M/G/1$ queue is a single-server queue where the customers arriva according to a Poisson process with rate λ, and the processing time of each server is a random variable with a general distribution with cdf G.

A generalization of Little's law is as follows. Imagine that entering customers are forced to pay money to the system. Then,

$$\mathbb{E}[\text{money that enters the system}] = \lambda \mathbb{E}[\text{money an arriving customer pays}]$$

Little's law is a consequence of establishing that every customer pays \$1 per unit time spent in the system.

If we now assume that each customer pays at a rate of y per unit time when his/her remaining time in the system is y, then we obtain

$$\mathbb{E}[\text{money that enters the system}] = \mathbb{E}[\text{remaining work in the system}] = W_Q$$

where the last equality holds by PASTA, because W_Q is the work in the system seen by an arrival.

Now, to compute the amount of money paid by a customer, let's define S as the customer's total service time, and w_Q the waiting time (as a random variable, not average). Then, since the remaining service time is S whenever the customer is in the queue, and $S-x$ when the customer has been at the server for x units of time, we obtain

$$\mathbb{E}[\text{money an arriving customer pays}] = \mathbb{E}\left[S w_Q + \int_0^S (S-x)\,dx\right]$$

Since the service times of the customers are independent and identically distributed, S is independent of the waiting time. Hence,

$$\mathbb{E}\left[S w_Q\right] = \mathbb{E}[S]\,\mathbb{E}\left[w_Q\right] = \mathbb{E}[S]\, w_Q$$

On the other hand,

$$\int_0^S (S-x)\,dx = \int_0^S x\,dx = \frac{S^2}{2}$$

Then,

$$\mathbb{E}[\text{money an arriving customer pays}] = \mathbb{E}[S]\, W_Q + \frac{S^2}{2}$$

Putting everything together,

$$\begin{aligned} W_Q &= \lambda\left(\mathbb{E}[S]\, W_Q + \frac{\mathbb{E}\left[S^2\right]}{2}\right) \\ \implies \quad W_Q &= \frac{\lambda \mathbb{E}\left[S^2\right]}{2(1-\lambda \mathbb{E}[S]} \end{aligned}$$