

CA MAS-I Chapter 1

1.4.0 Overview

 5m

In this section, we will learn about Poisson processes. To start, let's review a few properties of the Poisson distribution.

Let X be a random variable that follows the Poisson distribution with parameter λ .

- The probability mass function of X is:

$$\Pr(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

- The expected value and variance of X are:

$$\mathbb{E}[X] = \text{Var}[X] = \lambda$$

Note that these properties are provided in the exam table.

1.4.1 Poisson Processes

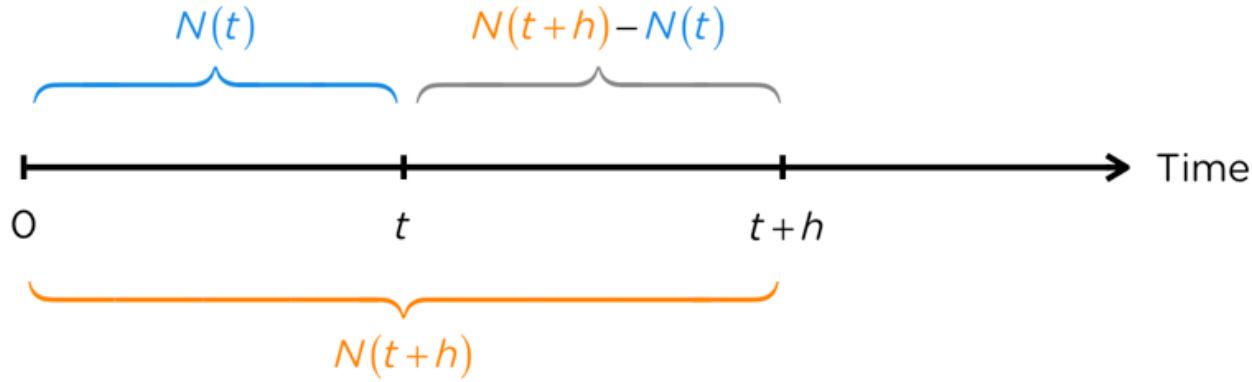
🕒 15m

Before we define a Poisson process, let's define a stochastic process and a counting process.

A stochastic process is defined as a collection of random variables. A *counting process*, N or $\{N(t), t \geq 0\}$, is a stochastic process whose random variables have values that are non-decreasing and non-negative integers. For time $t > 0$, a counting process counts the number of events that occur after time 0 up to (and including) time t . $N(t)$ is defined as the number of events that occur in the time interval $(0, t]$. Therefore,

- $N(t) \geq 0$
- $N(t)$ is an integer
- $N(0) = 0$
- $N(t + h) \geq N(t)$ for $h > 0$

The *increment* $N(t + h) - N(t)$ represents the number of events that occur in the time interval $(t, t + h]$. So, this includes events that occur at time $t + h$ but excludes events at time t .



Then, a *Poisson process* is a counting process where each increment is a Poisson random variable and non-overlapping increments are independent of one another. For a Poisson process N with rate function $\lambda(t)$:

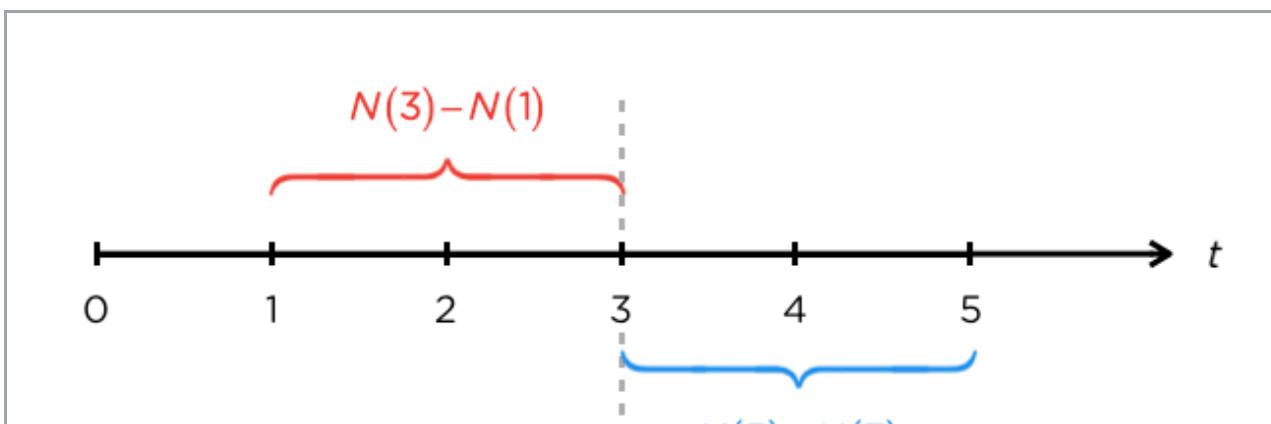
- The increment $N(t + h) - N(t)$ is a Poisson random variable with parameter $\lambda = \int_t^{t+h} \lambda(u) du$.

- If the rate function is constant, i.e. $\lambda(t) = \lambda$, the process N is a **homogeneous Poisson process**. In this case, $N(t + h) - N(t)$ is a Poisson random variable with parameter $h\lambda$.
- If the rate function $\lambda(t)$ varies with t , the process N is a **non-homogeneous Poisson process**.
- Because the increments $N(t)$ and $N(t + h) - N(t)$ are independent, the distribution of the number of events after time t is unaffected by the number of events before or at time t .

$$\begin{aligned}\Pr[N(t + h) - N(t) = x \mid N(t) = n] &= \frac{\Pr[N(t + h) - N(t) = x \cap N(t) = n]}{\Pr[N(t) = n]} \\ &= \frac{\Pr[N(t + h) - N(t) = x] \cdot \Pr[N(t) = n]}{\Pr[N(t) = n]} \\ &= \Pr[N(t + h) - N(t) = x]\end{aligned}$$

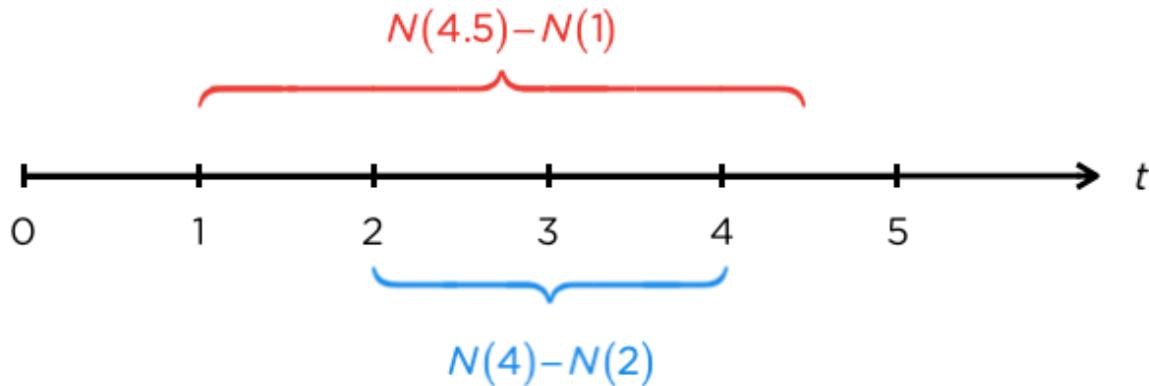
Coach's Remarks

In order for the increments to be independent, they have to be non-overlapping, or disjoint. For a Poisson process N , consider the increments $N(3) - N(1)$ and $N(5) - N(3)$. The increment $N(3) - N(1)$ represents the number of events occurring after time 1 up to (and including) time 3, while the increment $N(5) - N(3)$ represents the number of events occurring after time 3 up to (and including) time 5. These two increments do not overlap; thus, they are independent.



$$N(5) - N(5)$$

Now, consider the increments $N(4.5) - N(1)$ and $N(4) - N(2)$. Because the two increments overlap, the number of events in the time interval $(2, 4]$ and the number of events in the time interval $(1, 4.5]$ are not independent.



The following examples illustrate the properties of Poisson processes.

Example 1.4.1.1

Buses arrive at a bus stop at a Poisson rate of $\lambda = 4$ per hour.

Calculate

1. the probability that two buses arrive in the first twenty minutes.
2. the probability that one bus arrives in the first ten minutes and two buses arrive in the last ten minutes of the same hour.

Solution to (1)

Let N be the number of buses arriving at the bus stop. Because the rate function is a constant, N is a homogeneous Poisson process. Therefore, the number of buses arriving in the interval $(t, t + h]$ when measured in hours is a Poisson random variable with parameter $4h$.

So, $N\left(\frac{1}{3}\right) - N(0)$ is the the number of buses that arrive in the first twenty minutes, or $\frac{1}{3}$ hour. This is a Poisson random variable with parameter $\lambda = 4\left(\frac{1}{3}\right) = \frac{4}{3}$.

Recall that $N(0) = 0$. Therefore, the probability of two buses arriving in the first twenty minutes is

$$\begin{aligned} \Pr\left[N\left(\frac{1}{3}\right) = 2\right] &= \frac{e^{-4/3}\left(\frac{4}{3}\right)^2}{2!} \\ &= \mathbf{0.2343} \end{aligned}$$



Solution to (2)

The number of buses arriving in the first ten minutes, or $\frac{1}{6}$ hour, is a Poisson random variable with parameter $\lambda = 4\left(\frac{1}{6}\right) = \frac{2}{3}$. The number of buses arriving in the last ten minutes of the same hour is also a Poisson random variable with parameter $\lambda = \frac{2}{3}$.

Since the time intervals do not overlap, these two random variables are independent. Therefore, calculate the intersection probability by multiplying together the individual probabilities.

$$\Pr\left[N\left(\frac{1}{6}\right) - N(0) = 1\right] \cdot \Pr\left[N(1) - N\left(\frac{5}{6}\right) = 2\right] = \frac{e^{-2/3} \left(\frac{2}{3}\right)^1}{1!} \cdot \frac{e^{-2/3} \left(\frac{2}{3}\right)^2}{2!} = \mathbf{0.0391}$$

■

Example 1.4.1.2

Taxis arrive at a taxi stand according to a Poisson process with a rate function of $\lambda(t) = 3t$.

Calculate

1. the probability that at least two taxis arrive between $t = 0.2$ and $t = 0.5$.
2. the probability that five taxis arrive by $t = 0.5$ given that two taxis arrived by $t = 0.2$.

Solution to (1)

Let N be the number of taxis arriving at the taxi stand. N is a non-homogeneous Poisson process, so the number of taxis that arrive between $t = 0.2$ and $t = 0.5$ is a Poisson random variable with parameter

$$\lambda = \int_{0.2}^{0.5} 3t \, dt$$

$$\begin{aligned}
 & \int_{0.2}^{0.5} \left[\frac{3t^2}{2} \right]^{0.5} dt \\
 &= 0.315
 \end{aligned}$$

Therefore, the probability of at least two taxis arriving between $t = 0.2$ and $t = 0.5$ is

$$\begin{aligned}
 \Pr[N(0.5) - N(0.2) \geq 2] &= 1 - \Pr[N(0.5) - N(0.2) = 0] - \Pr[N(0.5) - N(0.2) = 1] \\
 &= 1 - e^{-0.315} - e^{-0.315} (0.315) \\
 &= \mathbf{0.0403}
 \end{aligned}$$



Solution to (2)

Use Bayes' Theorem to calculate the probability of observing five taxis by $t = 0.5$ given that two taxis arrived by $t = 0.2$. Note that the probability of observing five taxis by $t = 0.5$ and two taxis by $t = 0.2$ is the probability of observing two taxis in the interval $(0.0, 0.2]$ and three taxis in the interval $(0.2, 0.5]$.

$$\begin{aligned}
 \Pr[N(0.5) = 5 \mid N(0.2) = 2] &= \frac{\Pr[N(0.5) = 5 \cap N(0.2) = 2]}{\Pr[N(0.2) = 2]} \\
 &= \frac{\Pr[N(0.5) - N(0.2) = 3] \cdot \Pr[N(0.2) = 2]}{\Pr[N(0.2) = 2]} \\
 &= \Pr[N(0.5) - N(0.2) = 3]
 \end{aligned}$$

Thus, the probability of five taxis arriving by $t = 0.5$ given that two taxis arrived by

$t = 0.2$ is equal to the probability of observing three taxis between $t = 0.2$ and $t = 0.5$. This is true because non-overlapping increments are independent.

$$\Pr[N(0.5) - N(0.2) = 3] = \frac{e^{-0.315}(0.315)^3}{3!} = \mathbf{0.0038}$$



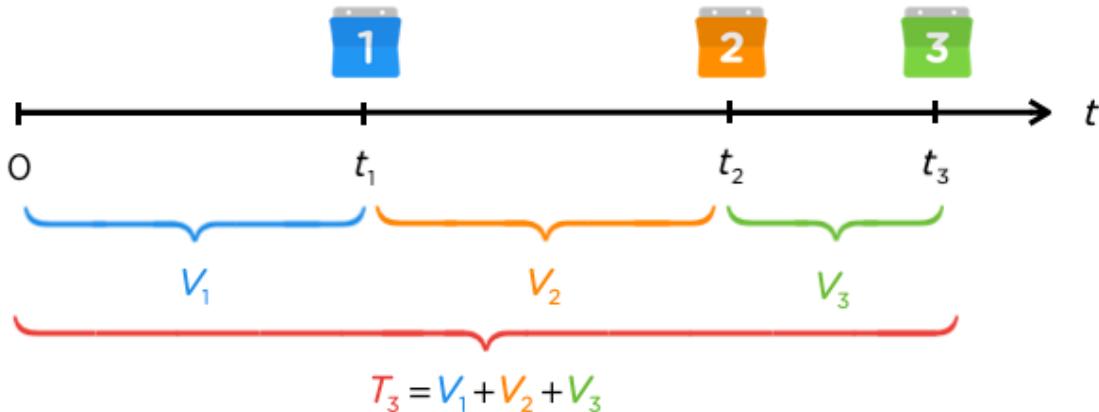
1.4.2 Time Between Events

🕒 25m

Let T_k be the time until the k^{th} event occurs. Let V_k be the time from the $(k-1)^{\text{st}}$ event to the k^{th} event. V_k is also known as the *inter-arrival time*. Then, we have the following relationships:

- $V_k = T_k - T_{k-1}, \quad k = 1, 2, 3, \dots$
- $T_k = V_1 + V_2 + \dots + V_k, \quad k = 1, 2, 3, \dots$
- $T_1 = V_1$
- $T_0 = 0, V_0 = 0$

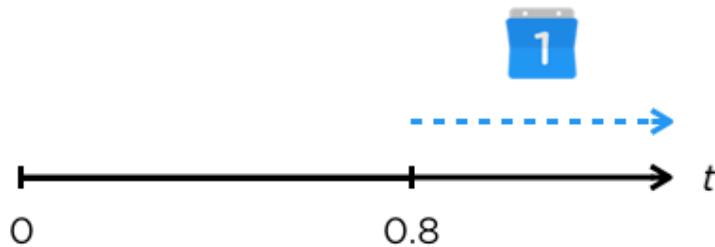
Let t_1 be the time of the first event, t_2 be the time of the second event, and so on. Then, V_1 and T_1 are the time to the first event, V_2 is the time from the first event to the second event, and so on.



Let's relate this to the Poisson process, N . What does the time to the k^{th} event, T_k , tell us about the number of events that have occurred by time t , $N(t)$?

Consider the time to the first event, T_1 . How do we calculate the probability that the time to the first event is greater than $t = 0.8$, i.e. $\Pr(T_1 > 0.8)$?

What must be true about the number of events that occur in the increment $N(0.8) - N(0)$? In order for the first event to occur after time $t = 0.8$, no events can occur before or at time $t = 0.8$.

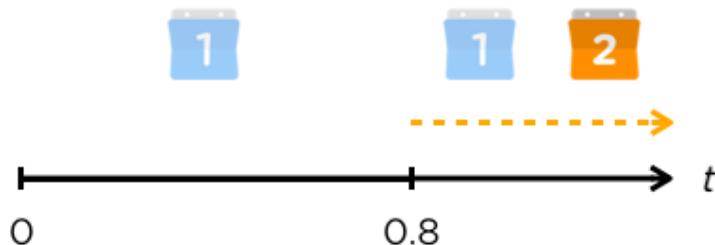


Therefore, the probability that the first event occurs after time $t = 0.8$ is equal to the probability that there are no events before or at time $t = 0.8$, i.e. $\Pr[N(0.8) = 0]$.

$$\Pr(T_1 > 0.8) = \Pr[N(0.8) = 0]$$

What can we say about the time to the second event? Determine the probability that the second event occurs after $t = 0.8$, or $\Pr(T_2 > 0.8)$.

You might think that in order for the second event to occur after time $t = 0.8$, one event must have occurred by time $t = 0.8$. This is not the case. While the first event **can** occur at or before $t = 0.8$, the first event can also occur after $t = 0.8$, as the second event would then also occur after $t = 0.8$. So, if there is either one event or no events by $t = 0.8$, then it is guaranteed that the second event will occur after $t = 0.8$.



Therefore, the probability of the second event occurring after $t = 0.8$ is equal to the probability of observing **at most** one event by time $t = 0.8$.

$$\Pr(T_2 > 0.8) = \Pr[N(0.8) \leq 1]$$

Notice the following:

$$\Pr(T_1 > 0.8) = \Pr[N(0.8) = 0] = \Pr[N(0.8) < 1]$$

$$\Pr(T_2 > 0.8) = \Pr[N(0.8) \leq 1] = \Pr[N(0.8) < 2]$$

From the examples above, we can generalize the relationship between the time to the k^{th} event and the number of events.

$$\Pr(T_k > s) = \Pr[N(s) < k] \quad (1.4.2.1)$$

For a homogeneous Poisson process, the inter-arrival time and the time to the k^{th} event are simplified. If N is a homogeneous Poisson process with rate λ , then

- the inter-arrival time, V_k , is an exponential random variable with mean $\theta = \frac{1}{\lambda}$, and
- the time to the k^{th} event, T_k , is a gamma random variable with parameters $\alpha = k$ and $\theta = \frac{1}{\lambda}$. This is because the sum of independent and identical exponential random variables, $V_1 + V_2 + \dots + V_k$, is a gamma random variable.

The converses are also true, assuming α has a positive integer value.

Vehicles arrive at a restaurant's drive-through window at a Poisson rate of $\lambda = 3$ per hour.

Calculate the probability that on a given day, the first vehicle arrives within 30 minutes.

Vehicles arrive at a Poisson rate of three per hour. This means that the number of vehicles arriving in the first 30 minutes is Poisson with mean $0.5 \cdot 3 = 1.5$. In order for the first vehicle to arrive within 30 minutes, we just need at least one vehicle to arrive during this period. Calculate this probability as

$$\begin{aligned}
 \Pr[N(0.5) \geq 1] &= 1 - \Pr[N(0.5) < 1] \\
 &= 1 - \Pr[N(0.5) = 0] \\
 &= 1 - \frac{e^{-1.5}(1.5)^0}{0!} \\
 &= 1 - e^{-1.5} \\
 &= \mathbf{0.7769}
 \end{aligned}$$

Notice that this can also be solved using the exponential distribution. The time until the first vehicle arrives is an exponential random variable with mean $\theta = \frac{1}{\lambda} = \frac{1}{3}$ hour. Calculate the probability that the first vehicle arrives within 30 minutes (or 0.5 hours) as

$$\begin{aligned}
 \Pr(V_1 \leq 0.5) &= F_{V_1}(0.5) \\
 &= 1 - e^{-0.5/(1/3)} \\
 &= 1 - e^{-1.5} \\
 &= \mathbf{0.7769}
 \end{aligned}$$

Let's work through a few more examples.

Example 1.4.2.1

Buses arrive at a bus stop at a Poisson rate of $\lambda = 4$ per hour.

Calculate the probability that the fourth bus arrives after more than one hour.

Solution

The probability that the fourth bus arrives sometime after one hour, $\Pr(T_4 > 1)$, is equivalent to the probability of observing fewer than four buses (at most three buses) in the first hour.

$$\Pr(T_4 > 1) = \Pr[N(1) < 4]$$

The number of buses arriving in the first hour is a Poisson random variable with $\lambda = 4$

Therefore,

$$\begin{aligned}
 \Pr(T_4 > 1) &= \Pr[N(1) < 4] \\
 &= \Pr[N(1) = 0] + \Pr[N(1) = 1] + \Pr[N(1) = 2] + \Pr[N(1) = 3] \\
 &= e^{-4} + 4e^{-4} + \frac{e^{-4}(4^2)}{2!} + \frac{e^{-4}(4^3)}{3!} \\
 &= \mathbf{0.4335}
 \end{aligned}$$



Coach's Remarks

Since N is a homogeneous Poisson process with rate 4, T_4 is a gamma random variable with parameters $\alpha = 4$ and $\theta = \frac{1}{4}$. Recall that we can use the Poisson distribution as a shortcut for evaluating the gamma CDF when α is a positive integer. This results in the same calculation as above.

Example 1.4.2.2

Taxis arrive at a taxi stand according to a Poisson process with a rate function of $\lambda(t) = 3t$.

Calculate the probability that the third taxi arrives after $t = 0.5$ and by $t = 0.9$.

Solution

The goal is to calculate the probability that T_3 is in the time interval $(0.5, 0.9]$. Notice that

this is equivalent to the probability that T_3 is greater than 0.5 minus the probability that T_3 is greater than 0.9.

$$\Pr(0.5 < T_3 \leq 0.9) = \Pr(T_3 > 0.5) - \Pr(T_3 > 0.9)$$

- If the third taxi arrives after $t = 0.5$, then at most two taxis arrive by time $t = 0.5$.
- If the third taxi arrives after $t = 0.9$, then at most two taxis arrive by time $t = 0.9$.

The increment $N(0.5)$ is a Poisson random variable with mean

$$\begin{aligned}\lambda &= \int_0^{0.5} 3t \, dt \\ &= \left[\frac{3t^2}{2} \right]_0^{0.5} \\ &= 0.375\end{aligned}$$

The increment $N(0.9)$ is a Poisson random variable with mean

$$\begin{aligned}\lambda &= \int_0^{0.9} 3t \, dt \\ &= \left[\frac{3t^2}{2} \right]_0^{0.9} \\ &= 1.215\end{aligned}$$

Therefore,

$$\begin{aligned}\Pr(T_3 > 0.5) &= \Pr[N(0.5) < 3] \\ &= \Pr[N(0.5) = 0] + \Pr[N(0.5) = 1] + \Pr[N(0.5) = 2] \\ &= e^{-0.375} + 0.375e^{-0.375} + \frac{e^{-0.375}(0.375^2)}{2!} \\ &= 0.9933\end{aligned}$$

$$\begin{aligned}\Pr(T_3 > 0.9) &= \Pr[N(0.9) < 3] \\ &= \Pr[N(0.9) = 0] + \Pr[N(0.9) = 1] + \Pr[N(0.9) = 2]\end{aligned}$$

$$\begin{aligned}
 &= e^{-1.215} + 1.215e^{-1.215} + \frac{e^{-1.215} (1.215)^2}{2!} \\
 &= 0.8762
 \end{aligned}$$

Thus, the probability that the third taxi arrives in the interval $(0.5, 0.9]$ is:

$$0.9933 - 0.8762 = \mathbf{0.1171}$$

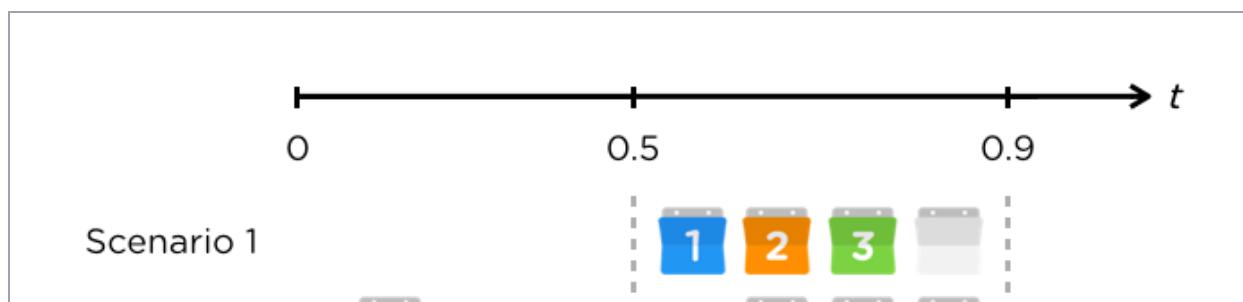


Alternative Solution

We know that non-overlapping increments of a Poisson process are independent. That means the intervals $(0.0, 0.5]$ and $(0.5, 0.9]$ are independent.

There are a few possible scenarios that will result in the third taxi arriving in the interval $(0.5, 0.9]$. There can be at most two taxis arriving in the interval $(0.0, 0.5]$ if the third taxi must be in the interval $(0.5, 0.9]$.

- If no taxis arrive in the interval $(0.0, 0.5]$, there must be at least three taxis that arrive in the interval $(0.5, 0.9]$ in order for the third taxi to be in this interval.
- If one taxi arrives in the interval $(0.0, 0.5]$, there must be at least two taxis that arrive in the interval $(0.5, 0.9]$ in order for the third taxi to be in this interval.
- If two taxis arrive in the interval $(0.0, 0.5]$, there must be at least one taxi that arrives in the interval $(0.5, 0.9]$ in order for the third taxi to be in this interval.





The increment $N(0.5)$ is a Poisson random variable with mean $\lambda = 0.375$, and the increment $N(0.9) - N(0.5)$ is a Poisson random variable with mean

$$\begin{aligned}\lambda &= \int_{0.5}^{0.9} 3t \, dt \\ &= \left[\frac{3t^2}{2} \right]_{0.5}^{0.9} \\ &= 0.84\end{aligned}$$

For Scenario 1,

$$\begin{aligned}\Pr[N(0.5) = 0] \cdot \Pr[N(0.9) - N(0.5) \geq 3] &= (e^{-0.375}) \left(1 - e^{-0.84} - 0.84e^{-0.84} - \frac{0.84^2 e^{-0.84}}{2!} \right) \\ &= 0.0367\end{aligned}$$

For Scenario 2,

$$\begin{aligned}\Pr[N(0.5) = 1] \cdot \Pr[N(0.9) - N(0.5) \geq 2] &= (0.375e^{-0.375}) (1 - e^{-0.84} - 0.84e^{-0.84}) \\ &= 0.0530\end{aligned}$$

For Scenario 3,

$$\begin{aligned}\Pr[N(0.5) = 2] \cdot \Pr[N(0.9) - N(0.5) \geq 1] &= \left(\frac{0.375^2 e^{-0.375}}{2!} \right) (1 - e^{-0.84}) \\ &= 0.0275\end{aligned}$$

Therefore, the probability that the third taxi arrives in the interval $(0.5, 0.9]$ is the sum of the probabilities of the three scenarios.

$$\begin{aligned} \Pr(0.5 < T_3 \leq 0.9) &= 0.0307 + 0.0530 + 0.0275 \\ &= \mathbf{0.1171} \end{aligned}$$

Use this method with caution because it can be easy to miss some of the possible scenarios. In addition, without advance knowledge of the approximate number of possible scenarios, this method can quickly become laborious.



1.4.3 Conditional Distribution of Arrival Times

🕒 15m

Let N be a homogeneous Poisson process. Given that n events happened by time t , what can we say about the distribution of the arrival times of those n events?

To start, we'll need a basic understanding of order statistics (covered in detail in Section 2.7). When we sort a set of data in ascending order, the ordered data points are called **order statistics**. Given n random variables, X_1, X_2, \dots, X_n , the order statistics are $X_{(1)}, X_{(2)}, \dots, X_{(n)}$, where $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$.

$X_{(1)}$ is the minimum of the n values, while $X_{(n)}$ is the maximum. Since $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ are order statistics, the distribution of $X_{(k)}$ is not the same as the distribution of X_k .

The times to events, T_i , have an inherent order. The time to the second event must be greater than the time to the first event, the time to the third event must be greater than the time to the second event, and so on. So, we can consider the time to the k^{th} event, T_k , as the k^{th} order statistic of the arrival times, $X_{(k)}$.

$$\begin{aligned} X_{(1)} &= T_1 \\ X_{(2)} &= T_2 \\ &\vdots && \vdots \\ X_{(k)} &= T_k \\ &\vdots && \vdots \\ X_{(n)} &= T_n \end{aligned}$$

To determine the distribution for T_k , first consider the **unordered** times of the events. Since $X_{(1)}$ through $X_{(n)}$ represent the ordered arrival times of the events, X_1 through X_n will be the unordered arrival times. In other words, X_k does not necessarily represent the time to the k^{th} event.

Then, we are given that n events for a homogeneous Poisson process N happened by time t . So, for a given event, each interval of the same length within $[0, t]$ is equally likely to contain that event.

Therefore, for a homogeneous Poisson process N , the following properties hold:

- Given that n events occur by time t , $N(t) = n$, the **unordered** times of the past events, X_1, X_2, \dots, X_n , follow independent and identical uniform distributions on the interval $[0, t]$.
- Given that the time to the n^{th} event is t , $T_n = t$, the **unordered** times of the past events, X_1, X_2, \dots, X_{n-1} , follow independent and identical uniform distributions on the interval $[0, t]$.

Coach's Remarks

There is a subtle difference between the two bullet points above. The difference lies in the time to the n^{th} event.

- In the first case, n events occur by time t . We do not know the exact time of the n^{th} event. All we know is that the event occurred by t . Therefore, the times of all n events are random variables.
- In the second case, we are given the time to the n^{th} event is t . Since we know the exact time of the n^{th} event, there are only $n - 1$ event times that are random variables.

Examples 1.4.3.1 and 1.4.3.2 will illustrate this difference.

Therefore, we can conclude that the time to the k^{th} event is the k^{th} order statistic of the independent and identical uniform distributions. There is a shortcut to calculate the expected time to the k^{th} event, i.e. the expected value of the k^{th} order statistic from a set of n uniform distributions.

Let X_1, X_2, \dots, X_n be independent and identical uniform distributions on the interval $[a, b]$. The expected value of the k^{th} order statistic is:

$$\mathbb{E}[X_{(k)}] = a + \frac{k(b-a)}{n+1} \quad (1.4.3.1)$$

Let's explore these properties in the following examples.

Example 1.4.3.1

Buses arrive at a bus stop at a Poisson rate of $\lambda = 4$ per hour. Buses begin arriving at the bus stop at 8 AM each day.

On a given day, it is known that three buses arrived between 8 AM and 9 AM.

Calculate

1. the expected time of arrival of the fifth bus.
2. the expected time of arrival of the second bus.
3. the probability that exactly one bus arrives by 8:30 AM.

Solution to (1)

This is a homogeneous Poisson process because the rate function is constant. Therefore, the time to the k^{th} event follows a gamma distribution with $\alpha = k$ and $\theta = \frac{1}{4} = 0.25$.

From Section 1.4.1, we know that the inter-arrival times are independent. This means that the events that happen after 9 AM are not affected by the events that happened before 9 AM. The distribution of the time of arrival of the fifth bus given that three buses arrived before 9 AM is the same as the distribution of the time of arrival of the second bus after 9 AM.

We know that the time to the second event follows a gamma distribution with $\alpha = 2$ and $\theta = \frac{1}{4}$. Thus, the fifth bus is expected to arrive $\alpha\theta = 2 \cdot \frac{1}{4} = \frac{1}{2}$ hour after the first hour. In other words, the expected arrival time is **9:30 AM**.

Solution to (2)

Given that three buses arrived in the first hour, $N(1) = 3$, the unordered arrival times of the first three buses are independent, identical uniform distributions on the interval $[0, 1]$.

The time to the second bus is the second order statistic, $X_{(2)} = T_2$, of three independent and identical uniform distributions. Using Equation 1.4.3.1, the expected value of the second order statistic is:

$$\mathbb{E}[T_2] = 0 + \frac{2(1-0)}{3+1} = 0.5$$

Therefore, the expected time of arrival of the second bus is **8:30 AM.**

Solution to (3)

Since the unordered arrival times of the first three buses are independent, identical uniform distributions on the interval $[0, 1]$, the probability of a given

bus arriving by 8:30 AM is 0.5.

The probability that exactly 1 of the 3 buses arrives by 8:30 AM is

$$\binom{3}{1} 0.5^1 (1 - 0.5)^2 = \mathbf{0.375}$$



Example 1.4.3.2

Buses arrive at a bus stop at a Poisson rate of $\lambda = 4$ per hour. Buses begin arriving at the bus stop at 8 AM each day.

On a given day, it is known that the third bus arrived at 9 AM.

Calculate the expected time of arrival of the second bus.

Solution

This is similar to the previous question, but note that instead of three buses arriving in the first hour (between 8 AM and 9 AM), this question states that the third bus arrives at 9 AM. So, since we know the exact arrival time of the third bus, only the arrival times of the first two buses are uniform on $[0, 1]$. The expected value of the second order statistic is:

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

$$\mathbb{E}[T_2] = 0 + \frac{\cdot}{2+1} = \frac{1}{3}$$

Thus, the expected time of arrival of the second bus is **8:40 AM.**



1.4.4 Properties of Poisson Processes

🕒 30m

Thinning

If a Poisson process can be broken down into two or more disjoint subprocesses, then these disjoint subprocesses are also Poisson processes with proportional rate functions. This process is called *thinning*.

Let N be a Poisson process with rate function $\lambda(t)$, and N_1, N_2, \dots, N_n be disjoint subprocesses of N with proportions $\pi_1, \pi_2, \dots, \pi_n$, respectively. Note that $\pi_1 + \pi_2 + \dots + \pi_n = 1$. Then, N_1, N_2, \dots, N_n are independent Poisson processes with rate functions $\pi_1\lambda(t), \pi_2\lambda(t), \dots, \pi_n\lambda(t)$, respectively.

Assume N is a homogeneous Poisson process. If the proportions $\pi_1, \pi_2, \dots, \pi_n$ are constant, the disjoint subprocesses N_1, N_2, \dots, N_n will also be homogeneous Poisson processes. However, if the proportions are not constant, the subprocesses will be non-homogeneous Poisson processes. This is one way that non-homogeneous Poisson processes can come about.

Example 1.4.4.1

Taxis arrive at a taxi stand according to a Poisson process with a rate function of $\lambda(t) = 3t$.

Taxis in the district are registered under their own counties. 30% of taxis are from county A, 45% of taxis are from county B, and 25% of taxis are from county C.

Calculate

1. the probability of at least two taxis from county B arriving at the taxi stand by $t = 2$.
2. the probability of exactly three taxis from county A and exactly two taxis from county C arriving at the taxi stand by $t = 2$.

Solution to (1)

The number of taxis that arrive at a taxi stand by $t = 2$ is a Poisson random variable with rate

$$\begin{aligned}\lambda &= \int_0^2 3t \, dt \\ &= \left[\frac{3t^2}{2} \right]_0^2 \\ &= 6\end{aligned}$$

We can use this to determine the respective rates for counties A, B, and C since the number of taxis from counties A, B, and C are disjoint subprocesses of the Poisson process.

- The number of county A taxis that arrive by $t = 2$ is a Poisson random variable with $\lambda = 0.30(6) = 1.8$.
- The number of county B taxis that arrive by $t = 2$ is a Poisson random variable with $\lambda = 0.45(6) = 2.7$.
- The number of county C taxis that arrive by $t = 2$ is a Poisson random variable with $\lambda = 0.25(6) = 1.5$.

Therefore, the probability that at least two county B taxis arrive at the taxi stand by $t = 2$ is:

$$\begin{aligned}\Pr[N_B(2) \geq 2] &= 1 - \Pr[N_B(2) = 0] - \Pr[N_B(2) = 1] \\ &= 1 - e^{-2.7} - 2.7e^{-2.7} \\ &= \mathbf{0.7513}\end{aligned}$$



Solution to (2)

Note that the number of taxis from county A and the number of taxis from county C are independent. Thus, the probability of exactly three taxis from county A and exactly two taxis from county C arriving at the taxi stand by $t = 2$ is:

$$\begin{aligned}\Pr[N_A(2) = 3] \cdot \Pr[N_C(2) = 2] &= \frac{1.8^3 e^{-1.8}}{3!} \cdot \frac{1.5^2 e^{-1.5}}{2!} \\ &= 0.0403\end{aligned}$$



Example 1.4.4.2

Customers enter a store at a Poisson rate of 100 per shift. Each shift is from 8 AM to 8 PM. The amount of money spent by each customer follows a Pareto distribution with $\alpha = 3$ and $\theta = 300$.

A customer who spends more than 400 in a visit is considered a high-spending customer.

What is the probability that a store will have exactly two high-spending customers

within the first two hours of the shift, i.e. from 8 AM to 10 AM?

Solution

The number of high-spending customers can be considered a disjoint subprocess of the Poisson process for the number of customers. To determine the rate of this subprocess, we need to calculate the probability that a customer is high-spending. This is equal to the survival function of the Pareto distribution evaluated at 400. Refer to the exam table for the Pareto survival function.

$$S(400) = \left(\frac{300}{400 + 300} \right)^3 = 0.0787$$

Therefore, the number of high-spending customers between 8 AM and 10 AM is a Poisson random variable with mean:

$$\lambda = 100 \left(\frac{2}{12} \right) (0.0787) = 1.3120$$

Note that we multiply by $\frac{2}{12}$ because a shift is 12 hours long and there are 2 hours from 8 AM to 10 AM. In other words, 8 AM to 10 AM is represented by the interval $(0, \frac{2}{12}]$. Then, the probability that a store will have two high-spending customers within the first two hours is:

$$\begin{aligned} \Pr \left[N_H \left(\frac{2}{12} \right) = 2 \right] &= \frac{1.3120^2 e^{-1.3120}}{2!} \\ &= \mathbf{0.2318} \end{aligned}$$



Superposition

The opposite of thinning a Poisson process is *superposition*. The superposition of Poisson processes is the sum of two or more Poisson processes. The sum of two or more **independent** Poisson processes is also a Poisson process.

Let N_1, N_2, \dots, N_n be independent Poisson processes with respective rate functions, $\lambda_1(t), \lambda_2(t), \dots, \lambda_n(t)$. Then, $N_1 + N_2 + \dots + N_n$ is a Poisson process with rate function $\lambda = \lambda_1(t) + \lambda_2(t) + \dots + \lambda_n(t)$.

Example 1.4.4.3

Taxis arrive at a taxi stand according to a Poisson process with a rate function of $\lambda(t) = 3t$.

Taxis in the district are registered under their own counties. 30% of taxis are from county A, 45% of taxis are from county B, and 25% of taxis are from county C.

Calculate the probability that a total of three taxis between counties A and C, and any number of taxis from county B, arrive at the taxi stand by $t = 2$.

Solution

From Example 1.4.4.1, we know that the number of county A taxis that arrive by $t = 2$ is a Poisson random variable with a mean of 1.8 and that the number of county C taxis that arrive by $t = 2$ is a Poisson random variable with a mean of 1.5.

Therefore, the number of taxis from county A or county C that arrive by $t = 2$ is a Poisson random variable with a mean of $1.8 + 1.5 = 3.3$.

The probability that three taxis from either of these two counties arrive at the

taxi stand by $t = 2$ is:

$$\Pr[N_{A,C}(2) = 3] = \frac{3.3^3 e^{-3.3}}{3!} \\ = 0.2209$$



Joint Probabilities

Assume we have two independent Poisson processes, N_1 and N_2 , with respective rates λ_1 and λ_2 . We can determine the probability of observing one event from N_1 before observing one event from N_2 . For this, denote the times to the first event for each process as $T_{1,1}$ and $T_{2,1}$, respectively. Since $T_{1,1}$ and $T_{2,1}$ are exponential, apply Equation 1.1.6.2.

$$\Pr(T_{1,1} < T_{2,1}) = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

We can generalize this to determine the probability of observing n events from the first process before observing m events from the second process. For this, it does not matter when we actually observe the m^{th} event from the second process, as long as it occurs after the n^{th} event from the first process. So, we can use the negative binomial distribution to calculate the probability of observing at most $m - 1$ events from the second process before observing the n^{th} event from the first process. Each event from the first process has probability $\frac{\lambda_1}{\lambda_1 + \lambda_2}$, which means each event from the second process has probability $\frac{\lambda_2}{\lambda_1 + \lambda_2}$.

$$\Pr(T_{1,n} < T_{2,m}) = \sum_{i=0}^{m-1} \binom{n+i-1}{n-1} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^n \left(\frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^i$$

$\text{t=0} \quad \backslash \quad / \quad \times \quad + \quad - \quad \times \quad \times \quad - \quad ,$

We will see this formula in action in Example 1.4.4.5 below. Note that the parameterization is different from what is used in the exam table for negative binomial.

Coach's Remarks

The probability above can also be calculated using the binomial distribution. Out of the first $n + m - 1$ events, there need to be at least n events from the first process and at most $m - 1$ events from the second process. This guarantees that the n^{th} event from the first process occurs before the m^{th} event from the second process. So, using the binomial distribution, the probability is

$$\Pr(T_{1,n} < T_{2,m}) = \sum_{i=n}^{n+m-1} \binom{n+m-1}{i} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^i \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{n+m-1-i}$$

Example 1.4.4.4

The time until the next bus arrives at a bus stop follows an exponential distribution with mean $\theta = 0.25$ hours. 25% of the buses that arrive at that bus stop are express buses while the rest are local buses.

Josh and Rachel ride the bus to work. Josh always takes the first bus that arrives, while Rachel always takes the first express bus. An express bus takes ten minutes to get to the office, while a local bus takes twenty minutes.

Josh has been waiting at the bus stop for four minutes before Rachel arrives.

Calculate

1. Rachel's expected total time to get to the office.
2. the probability that Josh arrives at the office at least five minutes before Rachel does.

Solution to (1)

A mean wait time of 0.25 hours for the next bus means there is a rate of $\frac{1}{0.25} = 4$ buses every hour. Since 25% of buses are express buses, the rate for express buses is $4 \cdot 25\% = 1$ every hour. Equivalently, the mean wait time is $\frac{1}{1} = 1$ hour for the next express bus.

Since Rachel takes the first express bus, she will spend 10 minutes on the bus. So, her expected total time to the office is $60 + 10 = 70$ minutes.



Solution to (2)

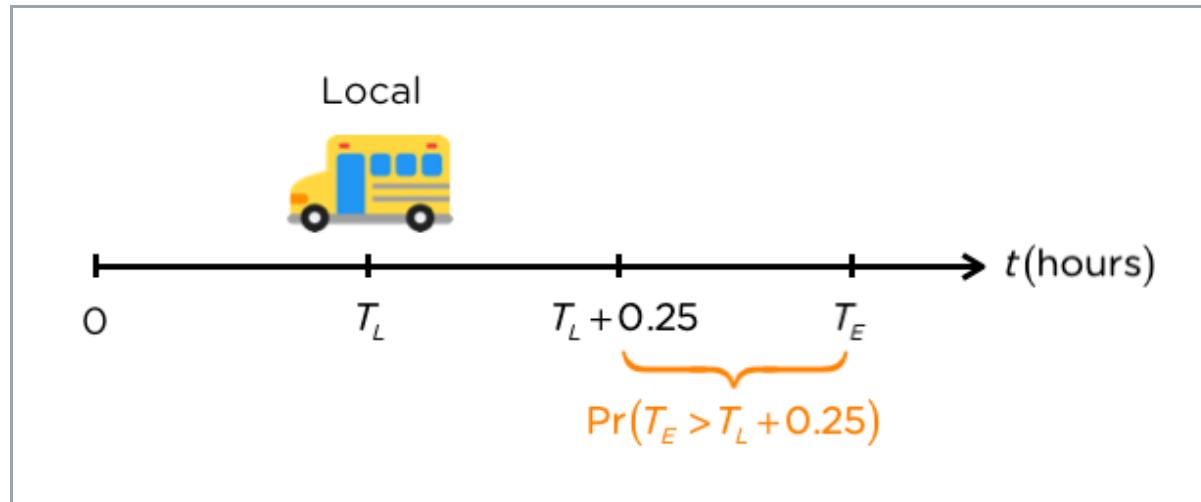
The possible scenarios are:

1. If the first bus is an express bus, Josh and Rachel will arrive at the office at the same time.
2. If the first bus is a local bus and an express bus arrives within 10 minutes of the local bus, Rachel will arrive at the office before Josh (because the local bus takes 10 more minutes to get to the office).
3. If the first bus is a local bus and an express bus arrives between 10 and

15 minutes after the local bus, Josh will arrive at the office less than five minutes before Rachel.

4. If the first bus is a local bus and an express bus arrives at least 15 minutes after the local bus, then Josh will arrive at the office at least five minutes before Rachel.

The goal is to calculate the probability of the 4th scenario. Let T_L be the time to the first local bus and T_E be the time to the first express bus.



- T_L is an exponential random variable with rate $\lambda_L = 4 \cdot 75\% = 3$.
- T_E is an exponential random variable with rate $\lambda_E = 4 \cdot 25\% = 1$.

Then, using the memoryless property, the probability is:

$$\begin{aligned}
 \Pr(T_E > T_L + 0.25) &= \Pr(T_E > T_L) \cdot \Pr(T_E > T_L + 0.25 \mid T_E > T_L) \\
 &= \Pr(T_E > T_L) \cdot \Pr(T_E > 0.25) \\
 &= \frac{3}{3+1} \cdot e^{-0.25} \\
 &= \mathbf{0.5841}
 \end{aligned}$$

Example 1.4.4.5

Buses arrive at a bus stop at a Poisson rate of $\lambda = 4$ per hour. 25% of the buses that arrive at that bus stop are express buses while the rest are local buses.

Calculate the probability of four local buses arriving before two express buses arrive.

Solution

We determined in Example 1.4.4.4 that the time to a local bus, T_L , is an exponential random variable with rate $\lambda_L = 3$. And the time to an express bus, T_E , is an exponential random variable with rate $\lambda_E = 1$. So, the probability of one local bus arriving before one express bus is

$$\Pr(T_L < T_E) = \frac{\lambda_L}{\lambda_L + \lambda_E} = \frac{3}{3 + 1} = 0.75$$

In order for 4 local buses to arrive before 2 express buses, we need at most 1 express bus to arrive before the 4th local bus.

$$\begin{aligned}\Pr(T_{L,4} < T_{E,2}) &= \sum_{i=0}^{2-1} \binom{4+i-1}{4-1} (0.75)^4 (0.25)^i \\ &= \binom{3}{3} (0.75)^4 (0.25)^0 + \binom{4}{3} (0.75)^4 (0.25)^1 \\ &= \mathbf{0.6328}\end{aligned}$$

The probability can also be calculated using Excel function = NEGBINOM.DIST(1, 4, 0.75, 1).

Alternative Solution

For 4 local buses to arrive before 2 express buses, at most 1 express bus can arrive among the first 5 buses.

$$\begin{aligned}\Pr(T_{L,4} < T_{E,2}) &= \sum_{i=4}^{4+2-1} \binom{4+2-1}{i} (0.75)^i (0.25)^{2+4-1-i} \\ &= \binom{5}{4} (0.75)^4 (0.25)^1 + \binom{5}{5} (0.75)^5 (0.25)^0 \\ &= \mathbf{0.6328}\end{aligned}$$

The probability can also be calculated using Excel function = `BINOM.DIST(1, 5, 0.25, 1)`.

1.4.5 Mixture of Poisson Processes

🕒 15m

Thus far, the rate function $\lambda(t)$ has either been a constant or a function of time. What if the rate function is a random variable?

In this case, we would have a mixture of Poisson processes. This creates a new process, but note that a mixture of Poisson processes is **not** a Poisson process. To define a mixture of Poisson processes, we will use the concepts covered in Section 1.1.9.

Let $N \mid \Lambda$ be a Poisson random variable with rate Λ , where Λ is a random variable. Then, the unconditional distribution of N has the following properties:

1. If Λ is a discrete random variable with probability mass function $p(\lambda)$, then the probability mass function of N is:

$$\begin{aligned}\Pr(N = n) &= \mathbb{E}_\Lambda[\Pr(N = n \mid \Lambda = \lambda)] \\ &= \mathbb{E}_\Lambda\left[\frac{e^{-\lambda} \lambda^n}{n!}\right] \\ &= \sum_{\lambda=0}^{\infty} \frac{e^{-\lambda} \lambda^n}{n!} \cdot p_\Lambda(\lambda)\end{aligned}$$

2. If Λ is a continuous random variable with probability density function $f(\lambda)$, then the probability mass function of N is:

$$\begin{aligned}\Pr(N = n) &= \mathbb{E}_\Lambda[\Pr(N = n \mid \Lambda = \lambda)] \\ &= \mathbb{E}_\Lambda\left[\frac{e^{-\lambda} \lambda^n}{n!}\right] \\ &= \int_0^\infty \frac{e^{-\lambda} \lambda^n}{n!} \cdot f_\Lambda(\lambda) d\lambda\end{aligned}$$

3. The mean is:

$$\begin{aligned}\mathbb{E}[N] &= \mathbb{E}_\Lambda[\mathbb{E}_N[N \mid \Lambda]] \\ &= \mathbb{E}_\Lambda[\Lambda]\end{aligned}$$

4. The variance is:

$$\begin{aligned}\text{Var}[N] &= \mathbb{E}_\Lambda[\text{Var}_N[N | \Lambda]] + \text{Var}_\Lambda[\mathbb{E}_N[N | \Lambda]] \\ &= \mathbb{E}_\Lambda[\Lambda] + \text{Var}_\Lambda[\Lambda]\end{aligned}$$

These concepts can be seen in the following examples.

Example 1.4.5.1

The annual number of claims is a Poisson random variable with mean Λ . Λ is a discrete random variable with the following possible values and probabilities:

$$p(\lambda) = \begin{cases} 0.8, & \text{if } \lambda = 2 \\ 0.2, & \text{if } \lambda = 6 \end{cases}$$

Calculate

1. the probability that there are five claims in any given year.
2. the variance of the number of claims.

Solution to (1)

Let N be the annual claim count. To calculate the probability that there are five claims in any given year, calculate the conditional probabilities of five claims for each value of λ . Then, compute the weighted average of the conditional probabilities, using the probabilities of λ as weights.

$$\Pr(N = 5) = \Pr(N = 5 | \Lambda = 2) \cdot \Pr(\Lambda = 2) + \Pr(N = 5 | \Lambda = 6) \cdot \Pr(\Lambda = 6)$$

$$2^5 e^{-2} \quad . \quad 6^5 e^{-6} \quad .$$

$$\begin{aligned} &= \frac{1}{5!}(0.8) + \frac{4}{5!}(0.2) \\ &= \mathbf{0.0610} \end{aligned}$$



Solution to (2)

The variance of the number of claims is the sum of the expectation and variance of Λ . The expectation of Λ is

$$\begin{aligned} E_{\Lambda}[\Lambda] &= 0.8(2) + 0.2(6) \\ &= 2.8 \end{aligned}$$

The variance of Λ is

$$\begin{aligned} \text{Var}_{\Lambda}[\Lambda] &= E_{\Lambda}[\Lambda^2] - (E_{\Lambda}[\Lambda])^2 \\ &= 0.8(2^2) + 0.2(6^2) - 2.8^2 \\ &= 2.56 \end{aligned}$$

Thus, the variance of the number of claims is

$$\begin{aligned} \text{Var}[N] &= 2.8 + 2.56 \\ &= \mathbf{5.36} \end{aligned}$$



Example 1.4.5.2

The annual number of claims is a Poisson random variable with mean Λ . Λ is a random variable with the following density function:

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

Calculate the probability that there are no claims in any given year.

Solution

Let N be the annual number of claims. The conditional annual number of claims, $N | \Lambda$, is a Poisson random variable. The unconditional annual number of claims has the following probability mass function:

$$\begin{aligned} \Pr(N = n) &= \int_{-\infty}^{\infty} \frac{e^{-\lambda} \lambda^n}{n!} \cdot f(\lambda) d\lambda \\ &= \int_0^{\infty} \frac{e^{-\lambda} \lambda^n}{n!} \cdot \lambda e^{-\lambda} d\lambda \end{aligned}$$

To calculate the probability that there are no claims, substitute $n = 0$ into the integral above.

$$\Pr(N = 0) = \int_0^{\infty} e^{-\lambda} \cdot \lambda e^{-\lambda} d\lambda$$

$$\begin{aligned}
 &= \int_0^\infty e^{-2\lambda} \lambda d\lambda \\
 &= \left[-\frac{\lambda e^{-2\lambda}}{2} - \frac{e^{-2\lambda}}{4} \right]_0^\infty \\
 &= \frac{1}{4}
 \end{aligned}$$



Coach's Remarks

Do not worry about questions where the rate function follows a distribution with a complex density function. It is highly unlikely that the exam would require a complex integration.

Poisson and Gamma

There is a special case where the rate function of a Poisson process, Λ , follows a gamma distribution.

Let $N | \Lambda$ be a Poisson random variable with parameter Λ , where Λ is a gamma random variable with parameters α and θ . Then, recall that for a Poisson-gamma mixture, N is a negative binomial random variable with parameters $r = \alpha$ and $\beta = \theta$.

Coach's Remarks

As you may have noticed, the distribution of Λ in Example 1.4.5.2 follows a gamma distribution with $\alpha = 2$ and $\theta = 1$. Thus, the annual number of claims is a negative binomial random variable with $r = 2$ and $\beta = 1$.

This relationship provides a much quicker approach to compute unconditional probabilities. Let's recalculate the probability of zero claims.

$$\Pr(N = 0) = (1 + \beta)^{-r} = (1 + 1)^{-2} = \frac{1}{4}$$

Example 1.4.5.3

The annual number of claims is a Poisson random variable with mean Λ . Λ is a gamma random variable with parameters $\alpha = 2$ and $\theta = 50$.

Calculate

1. the probability that there are five claims in any given year.
2. the variance of the number of claims.

Solution to (1)

The unconditional distribution of N is negative binomial with parameters $r = 2$ and $\beta = 50$. Refer to the exam table for the probability mass function of a negative binomial distribution.

The probability that there are five claims in any given year is

$$\begin{aligned}\Pr(N = 5) &= \frac{r(r+1)\dots(r+5-1)\beta^5}{5!(1+\beta)^{r+5}} \\ &= \frac{2(3)(4)(5)(6)(50^5)}{5!(1+50)^{2+5}} \\ &= \mathbf{0.0021}\end{aligned}$$



Solution to (2)

The variance of N is

$$\text{Var}[N] = r\beta(1 + \beta) = \mathbf{5,100}$$



1.4.6 Compound Poisson Processes

🕒 30m

Let's consider a scenario involving the claims filed by an insurance company. Insurance companies are often interested in both the number of claims and the amount of each claim, which together make up the aggregate claim amount. If claims occur at a Poisson rate and all claim amounts are independent and identically distributed, then the aggregate claim amount is what we call a *compound Poisson process*.

Let N be a Poisson process and $X_1, X_2, \dots, X_{N(t)}$ be independent and identically distributed random variables. Then,

$$S(t) = \sum_{i=1}^{N(t)} X_i$$

is a compound Poisson process. For a compound Poisson process,

- $S(0) = 0$
- If $N(t) = 0$, then $S(t) = 0$
- The expected value of $S(t)$ is

$$\mathbb{E}[S(t)] = \mathbb{E}[N(t)] \cdot \mathbb{E}[X]$$

- The variance of $S(t)$ is

$$\text{Var}[S(t)] = \mathbb{E}[N(t)] \cdot \mathbb{E}[X^2]$$

Refer to the appendix at the end of this section for a derivation of the expectation and variance of $S(t)$.

The distribution of $S(t)$ is difficult to determine. However, because $S(t)$ is the sum of independent and identically distributed random variables, we can estimate the probability of events in $S(t)$ by applying the Central Limit Theorem. Then, $S(t)$ is *approximately normal* with mean $\mathbb{E}[S(t)]$ and variance $\text{Var}[S(t)]$. Therefore, using normal approximation, the probability that the aggregate claim amount is less than s is:

$$\Pr[S(t) < s] = \Phi \left[\frac{s - \mathbb{E}[S(t)]}{\sqrt{\text{Var}[S(t)]}} \right]$$

Remember that **continuity correction** is needed when a continuous (e.g. normal) distribution is used to approximate discrete distributions.

Continuity Correction

If $S(t)$ is a discrete distribution, approximating $S(t)$ with a continuous distribution, such as the normal distribution, will require continuity correction.

Continuity correction is needed because a discrete distribution can only take on discrete values, while a continuous distribution includes all real numbers. Thus, we include one-half of the distance between discrete values to provide a better approximation (see below).

How is continuity correction performed? It is a simple two-step process:

1. Convert desired probability

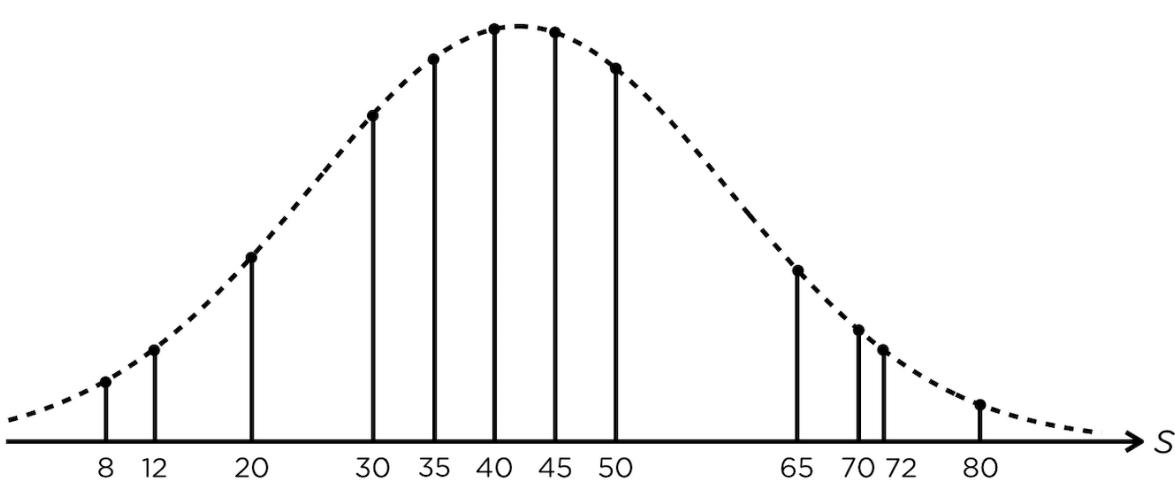
In this first step, we convert the desired probability to be written in terms of $\Pr(S < s)$. In particular, the equivalent probability should satisfy two conditions. First, the numerical value s must be a possible value of $S(t)$. Second, the probability is expressed using the less-than sign, $<$.

2. Apply continuity correction

Then, in the second step, we apply continuity correction by replacing s with the midpoint between s and the closest possible value of $S(t)$ that is less than s .

The example below will demonstrate this two-step approach.

Let S be a discrete distribution with the possible values shown in the diagram below. The distribution of S is approximated with a continuous distribution.



Approximate the probability that S is less than or equal to 48, i.e. $\Pr(S \leq 48)$.

Note that 48 is not a possible value of S , and the less-than sign is not used. So, the first step requires us to modify the desired probability so that it uses a possible value of S and is in terms of the $<$ sign.

The possible values of S that are closest to 48 are 45 and 50. But $\Pr(S \leq 48)$ includes the value $S = 45$. So,

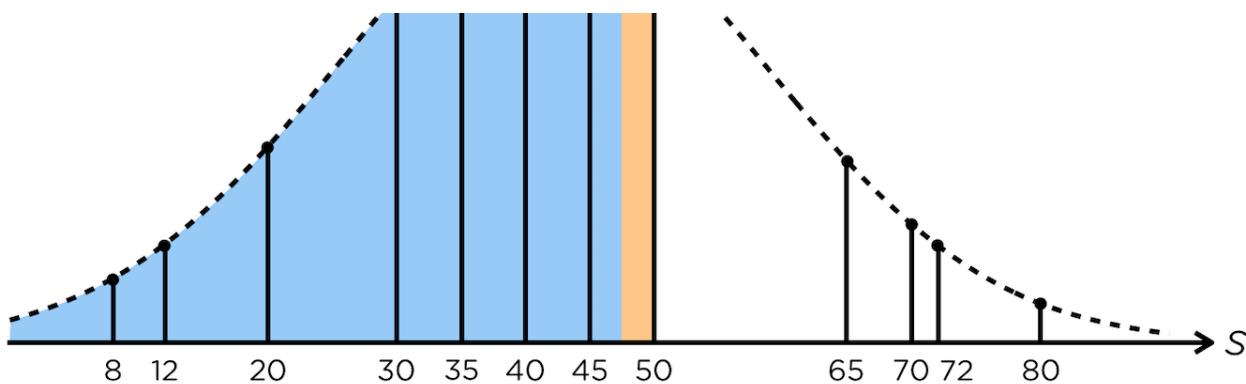
$$\Pr(S \leq 48) = \Pr(S < 50)$$

However, if we compute this probability as

$$\Pr(S < 50) = \Phi\left(\frac{50 - \text{E}[S]}{\sqrt{\text{Var}[S]}}\right)$$

we would be including the unwanted probability of the shaded region (see below).





This is where the second step comes in. To exclude this probability, we make a correction on 50. Since the next possible value of S below 50 is 45, we **subtract 2.5** to reach the **midpoint** of 45 and 50, which is 47.5.

Thus, with continuity correction:

$$\Pr(S < 50) \Rightarrow \Pr(S < 47.5) = \Phi\left(\frac{47.5 - E[S]}{\sqrt{\text{Var}[S]}}\right)$$

What if we want to calculate $\Pr(S > 50)$? The first step requires us to convert the probability. Since 50 is a possible value of S , we only need to make changes to the inequality sign.

$$\begin{aligned}\Pr(S > 50) &= 1 - \Pr(S \leq 50) \\ &= 1 - \Pr(S < 65)\end{aligned}$$

Next, apply continuity correction. The midpoint of 50 and 65 is 57.5. Thus, with continuity correction:

$$\begin{aligned}\Pr(S > 50) &= 1 - \Pr(S \leq 65) \\ &\Rightarrow 1 - \Pr(S < 57.5) \\ &= 1 - \Phi\left(\frac{57.5 - E[S]}{\sqrt{\text{Var}[S]}}\right)\end{aligned}$$

Coach's Remarks

Always use continuity correction when approximating a discrete distribution with a continuous distribution, unless explicitly instructed not to do so in the problem.

Example 1.4.6.1

The number of claims received by an insurance company follows a Poisson process with a rate of 1,000 per year. Each claim amount is normally distributed with a mean of 500 and a variance of 2,000. The number of claims and the claim amounts are mutually independent.

Calculate the probability that the aggregate claim amount in a year is greater than 520,000.

Solution

Let N be the annual number of claims received, X be the individual claim amount, and S be the aggregate claim amount.

The aggregate claim amount, S , is a compound Poisson process with mean and variance:

$$\begin{aligned} E[S] &= E[N] \cdot E[X] \\ &= 1,000 (500) \\ &= 500,000 \end{aligned}$$

$$\begin{aligned} \text{Var}[S] &= E[N] \cdot E[X^2] \\ &= 1,000 (2,000 + 500^2) \end{aligned}$$

$$\begin{aligned}
 &= 1,000(2,000 + 500) \\
 &= 252,000,000
 \end{aligned}$$

By the Central Limit Theorem, S is approximately normal with mean 500,000 and variance 252,000,000. Continuity correction is not needed because claim amounts follow a continuous distribution (normal). So, the probability that the aggregate claim amount is greater than 520,000 is:

$$\begin{aligned}
 \Pr(S > 520,000) &= 1 - \Pr(S \leq 520,000) \\
 &= 1 - \Phi\left(\frac{520,000 - 500,000}{\sqrt{252,000,000}}\right) \\
 &= 1 - \Phi(1.26) \\
 &= 1 - 0.8962 \\
 &= \mathbf{0.1038}
 \end{aligned}$$



Example 1.4.6.2

The number of the claims received by an insurance company follows a Poisson process with a rate of $4t^2$. Each claim amount is randomly distributed as follows:

Claim Amount	Probability
250	0.3
500	0.6
1,000	0.1

Calculate the probability that the aggregate claim amount from $t = 2$ to $t = 6$ is

Calculate the probability that the aggregate claim amount from $t = 3$ to $t = 6$ is less than 100,000.

Solution

Let N be the number of claims received, X be the individual claim amount, and S be the aggregate claim amount. The number of claims from $t = 3$ to $t = 6$ is a Poisson random variable with parameter

$$\begin{aligned}\lambda &= \int_3^6 4t^2 dt \\ &= \left[\frac{4t^3}{3} \right]_3^6 \\ &= 252\end{aligned}$$

The expected value and the second moment of the claim amount are:

$$\begin{aligned}\mathbb{E}[X] &= 0.3(250) + 0.6(500) + 0.1(1,000) \\ &= 475\end{aligned}$$

$$\begin{aligned}\mathbb{E}[X^2] &= 0.3(250^2) + 0.6(500^2) + 0.1(1,000^2) \\ &= 268,750\end{aligned}$$

Then, the aggregate claim amount has an expected value and variance of:

$$\begin{aligned}\mathbb{E}[S] &= \mathbb{E}[N] \cdot \mathbb{E}[X] \\ &= 252(475) \\ &= 119,700\end{aligned}$$

$$\begin{aligned}\text{Var}[S] &= \mathbb{E}[N] \cdot \mathbb{E}[X^2] \\ &= 252(268,750) \\ &= 67,725,000\end{aligned}$$

Because claim amounts are discrete, continuity correction is needed. The smallest claim amount is 250 and the other possible claim amounts are multiples of 250, so the aggregate claim amount random variable will be discrete in multiples of 250.

Since the desired probability satisfies the two conditions in the first step, we can move on to applying continuity correction. The closest possible value of S that is less than 100,000 is 99,750. Thus, the midpoint is $(100,000 + 99,750) / 2 = 99,875$.

$$\begin{aligned}\Pr(S < 100,000) &\Rightarrow \Pr(S < 99,875) \\&= \Phi\left(\frac{99,875 - 119,700}{\sqrt{67,725,000}}\right) \\&= \Phi(-2.41) \\&= 1 - 0.9920 \\&= \mathbf{0.0080}\end{aligned}$$



Example 1.4.6.3

The number of customers entering a store follows a Poisson process with a rate of 100 per shift. The amount of money spent by each customer follows a gamma distribution with $\alpha = 20$ and θ .

It is known that the probability of aggregate sales exceeding 295,000 in any given shift is 0.0951.

Calculate θ .

Solution

Let's calculate the mean and variance of the aggregate sales, S . Note that the mean of a gamma distribution is $\alpha\theta$ and the variance is $\alpha\theta^2$.

$$\begin{aligned} E[S] &= E[N] \cdot E[X] \\ &= 100 (\alpha\theta) \\ &= 2,000\theta \end{aligned}$$

$$\begin{aligned} \text{Var}[S] &= E[N] \cdot E[X^2] \\ &= 100 (20\theta^2 + 20^2\theta^2) \\ &= 42,000\theta^2 \end{aligned}$$

It is given that $\Pr(S > 295,000) = 0.0951$. Use this information to calculate θ . Do not apply continuity correction because claim amounts follow a continuous distribution (gamma).

$$\begin{aligned} \Pr(S > 295,000) &= 0.0951 \\ \Phi\left(\frac{295,000 - 2,000\theta}{\sqrt{42,000\theta^2}}\right) &= 0.9049 \\ \frac{295,000 - 2,000\theta}{\sqrt{42,000\theta^2}} &= 1.31 \\ 295,000 - 2,000\theta &= 268.4701\theta \\ \theta &= \mathbf{130.0436} \end{aligned}$$



Sums of Compound Poisson Processes

Let N_1 be a Poisson process with rate λ_1 and $X_{1,1}, X_{1,2}, \dots, X_{1,N_1(t)}$ be independent and identically distributed random variables, such that

$$S_1(t) = \sum_{i=1}^{N_1(t)} X_{1,i}$$

is a compound Poisson process. In addition, let N_2 be a Poisson process with rate λ_2 and $X_{2,1}, X_{2,2}, \dots, X_{2,N_2(t)}$ be independent and identically distributed random variables, such that

$$S_2(t) = \sum_{j=1}^{N_2(t)} X_{2,j}$$

is a compound Poisson process. Assume N_1, N_2 , each $X_{1,i}$, and each $X_{2,j}$ are independent. Then, $S(t) = S_1(t) + S_2(t)$ is also a compound Poisson process with rate

$$\lambda = \lambda_1 + \lambda_2$$

and with a discrete mixture severity. Specifically, the severity probability function has the form of Equation 1.1.9.1, i.e. a weighted average of the probability functions of X_1 and X_2 , with respective weights $\frac{\lambda_1}{\lambda_1 + \lambda_2}$ and $\frac{\lambda_2}{\lambda_1 + \lambda_2}$.

This can be easily extended to the sum of more than two compound Poisson processes.

Example 1.4.6.4

The number of auto insurance claims received by an insurance company follows a Poisson process with a rate of 400 per year. Each claim amount is exponentially distributed with mean 1,000.

The number of homeowners insurance claims received by the same insurance company follows a Poisson process with a rate of 150 per year. Each claim amount is normally distributed with a mean of 2,000 and a variance of 500.

Calculate the expected aggregate claim amount in a year for the insurance company.

Solution

Let N_1 be the annual number of auto insurance claims received, X_1 be the individual auto insurance claim amounts, and S_1 be the aggregate auto insurance claim amount.

Similarly, let N_2 be the annual number of homeowners insurance claims received, X_2 be the individual homeowners insurance claim amounts, and S_2 be the aggregate homeowners insurance claim amount.

Then, $S = S_1 + S_2$ is a compound Poisson process with rate

$$\lambda = 400 + 150 = 550$$

and with severity that is from X_1 with probability $\frac{400}{550}$ and from X_2 with probability $\frac{150}{550}$.

Therefore,

$$\begin{aligned} E[S] &= E[N] \cdot E[X] \\ &= 550 \left(\frac{400}{550} \cdot E[X_1] + \frac{150}{550} \cdot E[X_2] \right) \\ &= 550 \left(\frac{400}{550} \cdot 1,000 + \frac{150}{550} \cdot 2,000 \right) \\ &= 700,000 \end{aligned}$$



1.4 Summary

🕒 10m

- A Poisson process is a counting process where non-overlapping increments are independent.
- The increment $N(t + h) - N(t)$ of a Poisson process is a Poisson random variable with parameter $\lambda = \int_t^{t+h} \lambda(u) du$.
- If the rate function is constant, i.e. $\lambda(t) = \lambda$, the process N is a homogeneous Poisson process, and $N(t + h) - N(t)$ is a Poisson random variable with parameter $h\lambda$.
- If the rate function $\lambda(t)$ varies with t , the process N is a non-homogeneous Poisson process.

Time between Events

- Let T_k be the time until the k^{th} event occurs. Let V_k be the time from the $(k-1)^{\text{th}}$ event to the k^{th} event.
 - $V_k = T_k - T_{k-1}$
 - $T_k = V_1 + V_2 + \dots + V_k$
 - $T_1 = V_1$
 - $T_0 = 0, V_0 = 0$
- The probability that the time to the k^{th} event, T_k , is greater than any arbitrary value, s , is the probability of observing at most $k-1$ events by time s .

$$\Pr(T_k > s) = \Pr[N(s) < k]$$

- If N is a homogeneous Poisson process with rate λ ,
 - V_k is an exponential random variable with mean $\theta = \frac{1}{\lambda}$.

- T_k is a gamma random variable with parameters $\alpha = k$ and $\theta = \frac{1}{\lambda}$.

Conditional Distribution of Arrival Times

- Given that n events occur by time t , $N(t) = n$, the unordered times of the past events, X_1, X_2, \dots, X_n , follow independent and identical uniform distributions on the interval $[0, t]$.
- Given that the time to the n^{th} event is t , $T_n = t$, the unordered times of the past events, X_1, X_2, \dots, X_{n-1} , follow independent and identical uniform distributions on the interval $[0, t]$.
- Let X_1, X_2, \dots, X_n be n independent and identical uniform distributions on the interval $[a, b]$. The expected value of the k^{th} order statistic is:

$$\mathbb{E}[X_{(k)}] = a + \frac{k(b-a)}{n+1}$$

Properties of Poisson Processes

- Thinning of a Poisson process is the classifying of events of a Poisson process into disjoint types.
 - Subprocesses of a Poisson process N with rate function $\lambda(t)$ are Poisson processes with proportional rate functions.
- The superposition of two or more independent Poisson processes is the sum of two or more independent Poisson processes.
 - The sum of two or more independent Poisson processes is a Poisson process whose rate function is the sum of the respective rate functions.
- The probability of observing n events from a Poisson process N_1 before observing m events from another Poisson process N_2 (where N_1 and N_2 are independent) is found using the binomial distribution.

$$\Pr(T_{1,n} < T_{2,m}) = \sum_{i=n}^{n+m-1} \binom{n+m-1}{i} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^i \left(\frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{n+m-1-i}$$

Mixture of Poisson Processes

- If Λ is a discrete random variable with probability mass function $p(\lambda)$, then the probability mass function of N is:

$$\Pr(N = n) = \sum_{\lambda=0}^{\infty} \frac{e^{-\lambda} \lambda^n}{n!} \cdot p(\Lambda = \lambda)$$

- If Λ is a continuous random variable with probability density function $f(\lambda)$, then the probability mass function of N is:

$$\Pr(N = n) = \int_0^{\infty} \frac{e^{-\lambda} \lambda^n}{n!} \cdot f(\lambda) d\lambda$$

- Let $N | \Lambda$ be a Poisson random variable with parameter Λ , where Λ is a gamma random variable with parameters α and θ . Consequently, N is a negative binomial random variable with parameters $r = \alpha$ and $\beta = \theta$.

Compound Poisson Processes

- A compound Poisson process, $S(t)$, has the following mean and variance:

$$\mathbb{E}[S(t)] = \mathbb{E}[N(t)] \cdot \mathbb{E}[X]$$

$$\text{Var}[S(t)] = \mathbb{E}[N(t)] \cdot \mathbb{E}[X^2]$$

- The probability of an event in $S(t)$ can be estimated using normal approximation:

$$\Pr[S(t) < s] = \Phi \left[\frac{s - \text{E}[S(t)]}{\sqrt{\text{Var}[S(t)]}} \right]$$

- Continuity correction is needed if $S(t)$ is discrete.
- The sum of independent compound Poisson processes is also a compound Poisson process with a rate equal to the sum of the individual rate functions and with a mixed severity distribution.

Appendix

⌚ 5m

Mean and Variance for Compound Poisson Process

Let N be a Poisson process and $X_1, X_2, \dots, X_{N(t)}$ be independent and identically distributed random variables. Then,

$$S(t) = \sum_{i=1}^{N(t)} X_i$$

is a compound Poisson process. The mean and variance of $S(t)$ can be calculated using the Law of Total Expectation and Law of Total Variance.

$$\begin{aligned} E[S(t)] &= E\left[\sum_{i=1}^{N(t)} X_i\right] \\ &= E_{N(t)}\left[E\left[\sum_{i=1}^{N(t)} X_i \mid N(t)\right]\right] \\ &= E_{N(t)}[N(t) \cdot E[X]] \\ &= E[N(t)]E[X] \end{aligned}$$

$$\begin{aligned} \text{Var}[S(t)] &= \text{Var}\left[\sum_{i=1}^{N(t)} X_i\right] \\ &= E_{N(t)}\left[\text{Var}\left[\sum_{i=1}^{N(t)} X_i \mid N(t)\right]\right] + \text{Var}_{N(t)}\left[E\left[\sum_{i=1}^{N(t)} X_i \mid N(t)\right]\right] \\ &= E_{N(t)}[N(t) \cdot \text{Var}[X]] + \text{Var}_{N(t)}[N(t) \cdot E[X]] \\ &= E[N(t)]\text{Var}[X] + \text{Var}[N(t)]E[X]^2 \end{aligned}$$

$$\begin{aligned} &= \mathbb{E}[N(t)]\text{Var}[X] + \mathbb{E}[N(t)]\mathbb{E}[X]^2 \\ &= \mathbb{E}[N(t)] \left[\text{Var}[X] + \mathbb{E}[X]^2 \right] \\ &= \mathbb{E}[N(t)]\mathbb{E}[X^2] \end{aligned}$$

1.5.0 Overview

 5m

In basic terms, reliability theory is about the probability that a system with multiple components is functioning. The functioning of components in a system determines the functioning of the entire system.

In this section, we will start by introducing the basic concepts of reliability and different types of systems. We will then investigate the reliability and lifetime of systems.

1.5.1 Introduction to Systems

🕒 10m

A system consists of multiple components. Each component has its own lifetime, a random variable underlying its lifetime, and its state. The state of a component can be either functioning or not at a point in time, so it can be expressed as a binary variable. Let x_i be the state of component i .

$$x_i = \begin{cases} 1, & \text{if component } i \text{ is functioning} \\ 0, & \text{if component } i \text{ has failed} \end{cases}$$

The states of all components in a system can be written as a vector. Let $\mathbf{x} = (x_1, x_2, \dots, x_n)$ be the **state vector**. The state vector indicates which components are functioning and which have failed. A system with n components has 2^n distinct state vectors because each of the n elements within a vector can be one of two values.

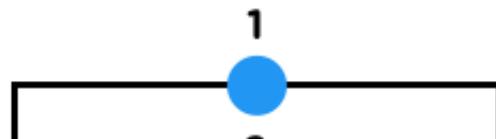
The state of a system depends on the state of its components. Like the components, the state of a system can be expressed as a binary variable. We represent the state of a system as a function, $\phi(\mathbf{x})$, where:

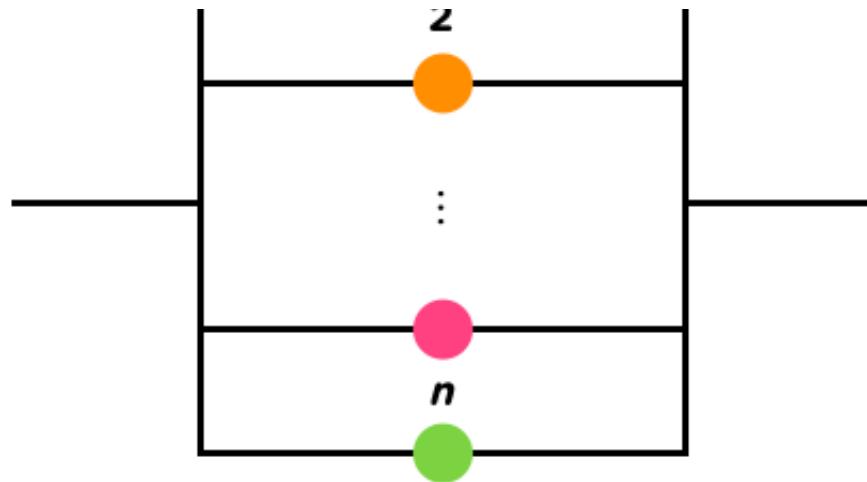
$$\phi(\mathbf{x}) = \begin{cases} 1, & \text{if the system is functioning} \\ 0, & \text{if the system has failed} \end{cases}$$

$\phi(\mathbf{x})$ is the **structure function** of a system. We will discuss structure functions in Section 1.5.3. For now, let's learn about a few common systems.

Parallel System

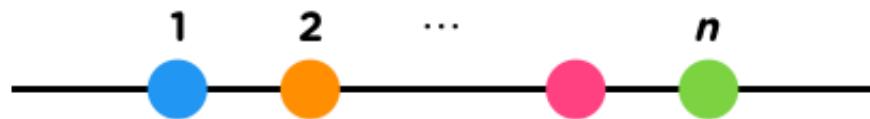
A parallel system is a system that will function as long as one of the components is functioning. The functioning of at least one component guarantees the functioning of the system.





Series System

A series system is a system that will function only when all components are functioning. If one of the components fails, then the whole system fails.



k-out-of-n System

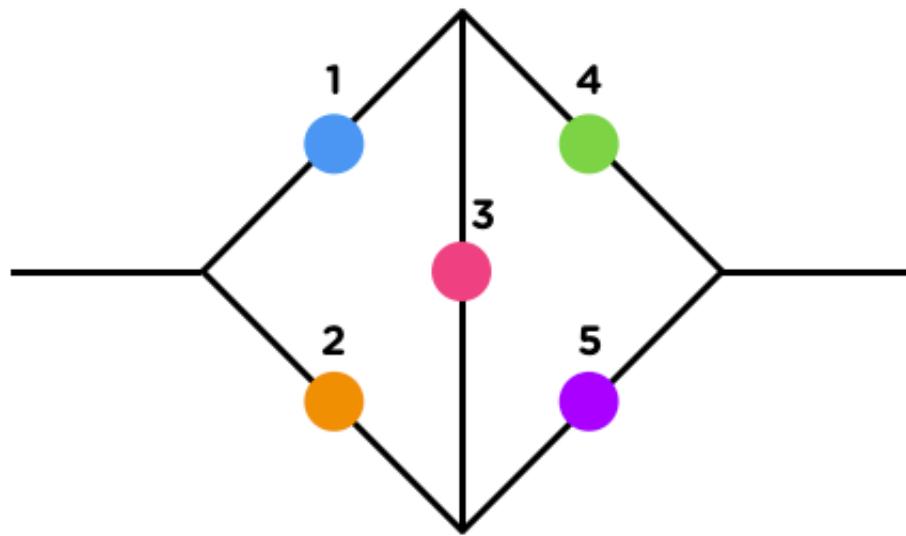
A *k-out-of-n system* will function if and only if at least k out of n components are functioning. For example, a two-out-of-three system will function if at least two of the three components are functioning.

Note that a parallel system is a one-out-of- n system, while a series system is an n -out-of- n system.

Bridae Svstem

-----, -----

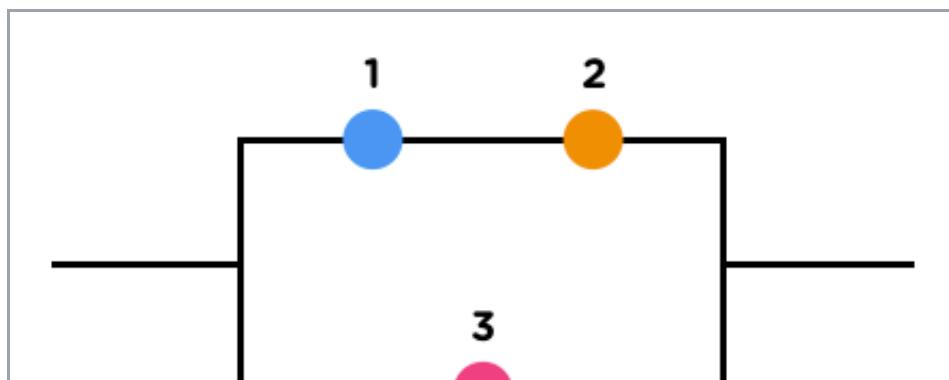
A bridge system is a system as pictured in the diagram below. It has two branches that are "bridged" by a third branch somewhere in the middle.



Other Systems

In reality, there are an unlimited number of possible systems. Systems can be constructed using a combination of any of the four systems mentioned above. For example, there could be a system with five components that will only function if the first two components are functioning and if two of the remaining three components are also functioning.

Consider the following system with three independent components.





Determine the possible combinations of states of the components and the resulting state of the system.

With three components, we have $2^3 = 8$ possible combinations of states. The table below lists all 8 possibilities and the state of the system for each possibility.

Component 1, x_1	Component 2, x_2	Component 3, x_3	System, $\phi(\mathbf{x})$
0	0	0	0
0	0	1	1
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1

As practice, let's work through a couple examples.

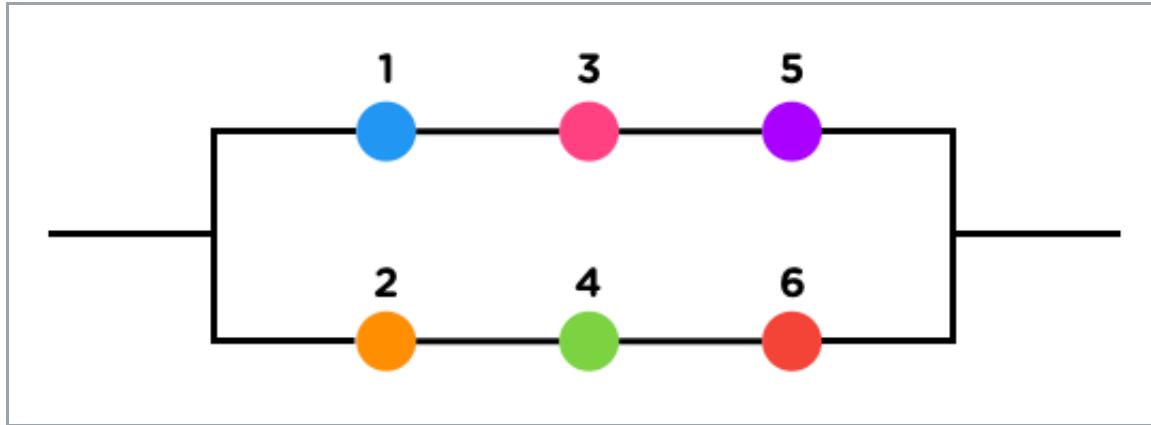
Example 1.5.1.1

A backup generator has six parts labeled 1 through 6. The generator will work as long as all even numbered parts are working or all odd numbered parts are working.

Draw the system.

Solution

This system works as long as at least parts 1, 3, and 5 are all working or parts 2, 4, and 6 are all working. Thus, we can consider this as two series systems – one with components 1, 3, and 5, and the other with components 2, 4, and 6 – placed in a parallel system together. So, the system can be drawn as:



Example 1.5.1.2

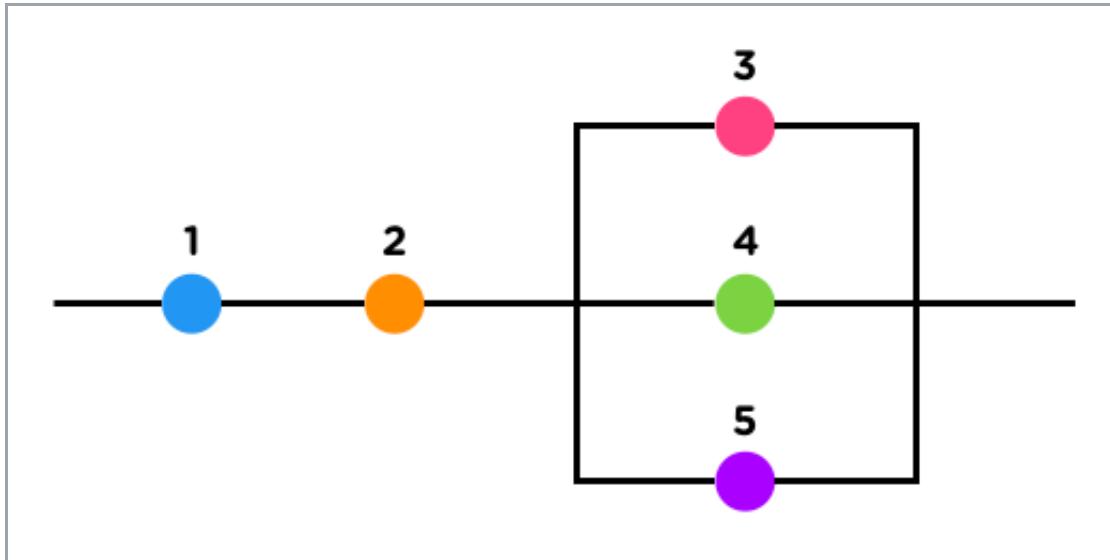
A 2-out-of-2 system is placed in a series with a 1-out-of-3 system.

Draw the system and determine all combinations of states which result in a functioning system.

Solution

Label the components in the 2-out-of-2 system as 1 and 2, and label the

components in the 1-out-of-3 system as 3, 4, and 5. Notice that the 2-out-of-2 system is a series system, while the 1-out-of-3 system is a parallel system. Since these systems are placed in a series together, this system works as long as at least components 1 and 2 and one of components 3, 4, and 5 are working. So, we can draw the system as:



The system functions as long as component 1 functions, component 2 functions, and at least one of components 3, 4, and 5 function. Thus, the possible combinations of states resulting in a functioning system are:

Component 1, x_1	Component 2, x_2	Component 3, x_3	Component 4, x_4	Component 5, x_5
1	1	1	0	0
1	1	0	1	0
1	1	0	0	1
1	1	1	1	0
1	1	1	0	1
1	1	0	1	1
1	1	1	1	1

1.5.2 Minimal Path Sets and Minimal Cut Sets

🕒 20m

Minimal Path Sets

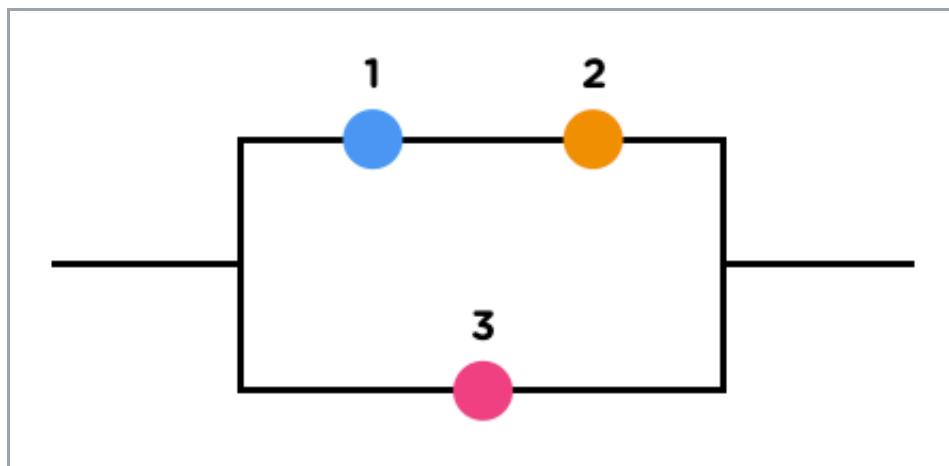
A state vector \mathbf{x} is a *path vector* if the system is functioning, i.e. $\phi(\mathbf{x}) = 1$. A path vector \mathbf{x} is a *minimal path vector* if the system would fail if any of the functioning components in \mathbf{x} were to fail, i.e. there are no "backup" components.

So, \mathbf{x} is a minimal path vector if $\phi(\mathbf{y}) = 0$ for every $\mathbf{y} < \mathbf{x}$. The notation $\mathbf{y} < \mathbf{x}$ means that each y_i must be less than or equal to x_i , with at least one y_i that is strictly less than x_i . In other words, $\mathbf{y} < \mathbf{x}$ if $y_i \leq x_i$ for $i = 1, 2, \dots, n$, with $y_i < x_i$ for some i .

Each minimal path vector can be represented as a *minimal path set*, or a minimal set of components whose functioning guarantees the functioning of the system. Let A be a minimal path set. Then, given a minimal path vector \mathbf{x} , the minimal path set A is $\{i : x_i = 1\}$.

This means that a system will function if and only if all components of at least one minimal path set function.

Consider the earlier system with three independent components.



Determine the minimal path sets for this system.

Recall the possible combinations of states of the components and the resulting state of the system.

Component 1, x_1	Component 2, x_2	Component 3, x_3	System, $\phi(\mathbf{x})$
0	0	0	0
0	0	1	1
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1

So, we have five path vectors for this system, i.e. five state vectors where the system functions. To determine which of these are minimal path vectors, we first list, for each path vector \mathbf{x} , all vectors \mathbf{y} that satisfy $\mathbf{y} < \mathbf{x}$. We do this by comparing each y_i to x_i for $i = 1, \dots, n$. For example, for the path vector $\mathbf{x} = (0, 0, 1)$:

- $(0, 0, 0)$ satisfies $\mathbf{y} < \mathbf{x}$ because $y_1 \leq x_1$, $y_2 \leq x_2$, and $y_3 \leq x_3$ with y_3 being strictly less than x_3 .
- $(0, 0, 1)$ does not satisfy $\mathbf{y} < \mathbf{x}$ even though $y_1 \leq x_1$, $y_2 \leq x_2$, and $y_3 \leq x_3$. This is because none of the y_i 's are strictly less than x_i .

The table below shows all of the vectors \mathbf{y} for each path vector \mathbf{x} .

Path Vector, \mathbf{x}	All \mathbf{y}
$(0, 0, 1)$	$(0, 0, 0)$
$(0, 1, 1)$	$(0, 0, 0), (0, 0, 1), (0, 1, 0)$
$(1, 0, 1)$	$(0, 0, 0), (0, 0, 1), (1, 0, 0)$
$(1, 1, 0)$	$(0, 0, 0), (0, 1, 0), (1, 0, 0)$
$(1, 1, 1)$	$(0, 0, 0), (0, 0, 1), (0, 1, 0), (0, 1, 1), (1, 0, 0), (1, 0, 1), (1, 1, 0)$

After listing all vectors \mathbf{y} for each path vector, we check if $\phi(\mathbf{y}) = 0$. Only three vectors have their structure functions equal to zero: $(0, 0, 0)$, $(0, 1, 0)$, and $(1, 0, 0)$.

A path vector can only be a minimal path vector if **all** of its corresponding vectors \mathbf{y} are from the set of vectors where $\phi(\mathbf{y}) = 0$.

Only path vectors $(0, 0, 1)$ and $(1, 1, 0)$ have all vectors \mathbf{y} such that $\phi(\mathbf{y}) = 0$. Therefore, $(0, 0, 1)$ and $(1, 1, 0)$ are minimal path vectors. This means $\{3\}$ and $\{1, 2\}$ are the minimal path sets for this system. As long as component 3 or components 1 and 2 are functioning, the system will function.

Coach's Remarks

We do not have to go through the steps shown above to figure out the minimal path sets. We can instead use logic and the definition of a minimal path set. We know that a minimal path set is a **minimal** set of components whose functioning guarantees the functioning of the system. We can then look for sets of non-overlapping components for which no component can be removed without causing the system to fail.

Looking at the system, we know that as long as component 3 or components 1 and 2 function, the system functions. Thus, $\{3\}$ and $\{1, 2\}$ are the minimal path sets.

Minimal Cut Sets

In contrast to a path vector, a state vector \mathbf{x} is a **cut vector** if the system is not functioning, i.e. $\phi(\mathbf{x}) = 0$. A cut vector \mathbf{x} is a **minimal cut vector** if repairing any of the non-functioning components in \mathbf{x} would result in the whole system functioning.

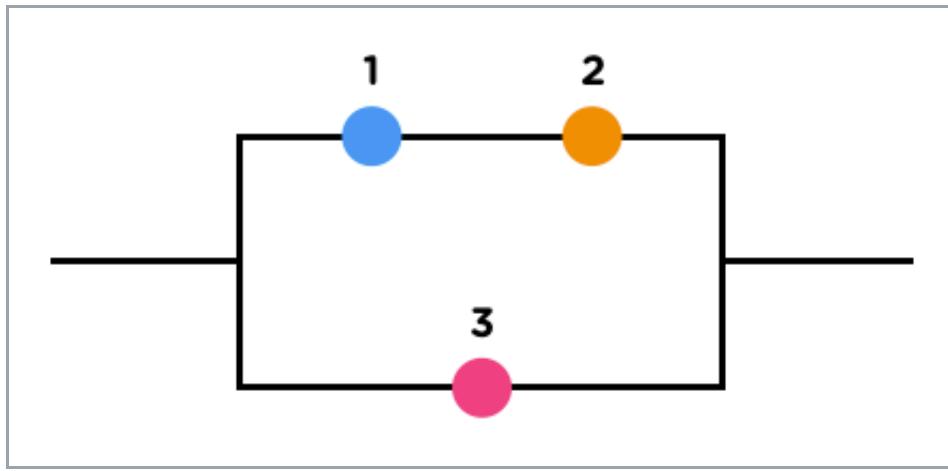
So, \mathbf{x} is a minimal cut vector if $\phi(\mathbf{y}) = 1$ for every $\mathbf{y} > \mathbf{x}$. The notation $\mathbf{y} > \mathbf{x}$ means that each y_i must be greater than or equal to x_i , with at least one y_i that is strictly greater than x_i . In other words, $\mathbf{y} > \mathbf{x}$ if $y_i \geq x_i$ for $i = 1, 2, \dots, n$, with $y_i > x_i$ for some i .

Each minimal cut vector can be represented as a **minimal cut set**, or a minimal set of

components whose failure guarantees the failure of the system. Let C be a minimal cut set. Then, given a minimal cut vector \mathbf{x} , the minimal cut set C is $\{i : x_i = 0\}$.

This means that a system will fail if and only if all components of at least one minimal cut set fail.

Determine the minimal cut sets for this system with three independent components.



From the vectors we listed previously for this system, three are cut vectors, i.e. three are state vectors where the system fails. Similar to our approach earlier, we begin by listing, for each cut vector \mathbf{x} , all vectors \mathbf{y} that satisfy $\mathbf{y} > \mathbf{x}$.

Cut Vector, \mathbf{x}	All \mathbf{y}
(0, 0, 0)	(0, 0, 1), (0, 1, 0), (0, 1, 1), (1, 0, 0), (1, 0, 1), (1, 1, 0), (1, 1, 1)
(0, 1, 0)	(0, 1, 1), (1, 1, 0), (1, 1, 1)
(1, 0, 0)	(1, 0, 1), (1, 1, 0), (1, 1, 1)

Then, we check each \mathbf{y} to see if $\phi(\mathbf{y}) = 1$. Only five vectors have their structure functions equal to one: (0, 0, 1), (0, 1, 1), (1, 0, 1), (1, 1, 0), (1, 1, 1).

A cut vector can only be a minimal cut vector if **all** of its corresponding vectors \mathbf{y} are from the set of vectors where $\phi(\mathbf{y}) = 1$.

Only cut vectors $(0, 1, 0)$ and $(1, 0, 0)$ have all vectors \mathbf{y} such that $\phi(\mathbf{y}) = 1$. Therefore, $(0, 1, 0)$ and $(1, 0, 0)$ are minimal cut vectors. This means that $\{1, 3\}$ and $\{2, 3\}$ are the minimal cut sets for this system. If components 1 and 3 or components 2 and 3 stop functioning, the system is guaranteed to fail.

Coach's Remarks

Using similar logic, we can find the minimal cut sets without going through the steps shown above.

The system will fail if all three components fail. However, it is not necessary for all three to fail to guarantee system failure. Component 3 must fail, but only one of the other two components needs to fail. Thus, $\{1, 3\}$ and $\{2, 3\}$ are the minimal cut sets.

Note that all minimal path sets must have at least one component from each minimal cut set, and all minimal cut sets must have at least one component from each minimal path set. For example, for the system of three independent components above, the minimal path sets are $\{1, 2\}$ and $\{3\}$, and the minimal cut sets are $\{1, 3\}$ and $\{2, 3\}$. So, the minimal path set $\{1, 2\}$ shares component 1 with the first minimal cut set and component 2 with the second minimal cut set. And the minimal path set $\{3\}$ shares component 3 with both minimal cut sets.

We can calculate the number of minimal path sets and minimal cut sets for the common systems introduced in the previous subsection as shown below:

System	Number of Minimal Path Sets	Number of Minimal Cut Sets
Parallel	n	1
Series	1	n
k -out-of- n	$\binom{n}{k}$	$\binom{n}{n-k+1}$
Bridge	4	4

In addition, assume we have multiple systems, where we know the number of minimal

path sets and minimal cut sets for each system. If the systems are placed in parallel, the total number of minimal path sets is the sum of the individual amounts, while the number of minimal cut sets is the product of the individual amounts. On the other hand, if the systems are placed in series, the total number of minimal path sets is the product of the individual amounts, while the number of minimal cut sets is the sum of the individual amounts. We will see how this works in the example below.

Example 1.5.2.1

Calculate the number of minimal path sets if

1. a 3-out-of-4 system is placed in series with a 2-out-of-6 system.
2. a 3-out-of-4 system is placed in parallel with a 2-out-of-6 system.

Solution to (1)

A k -out-of- n system has $\binom{n}{k}$ minimal path sets. So, the 3-out-of-4 system has

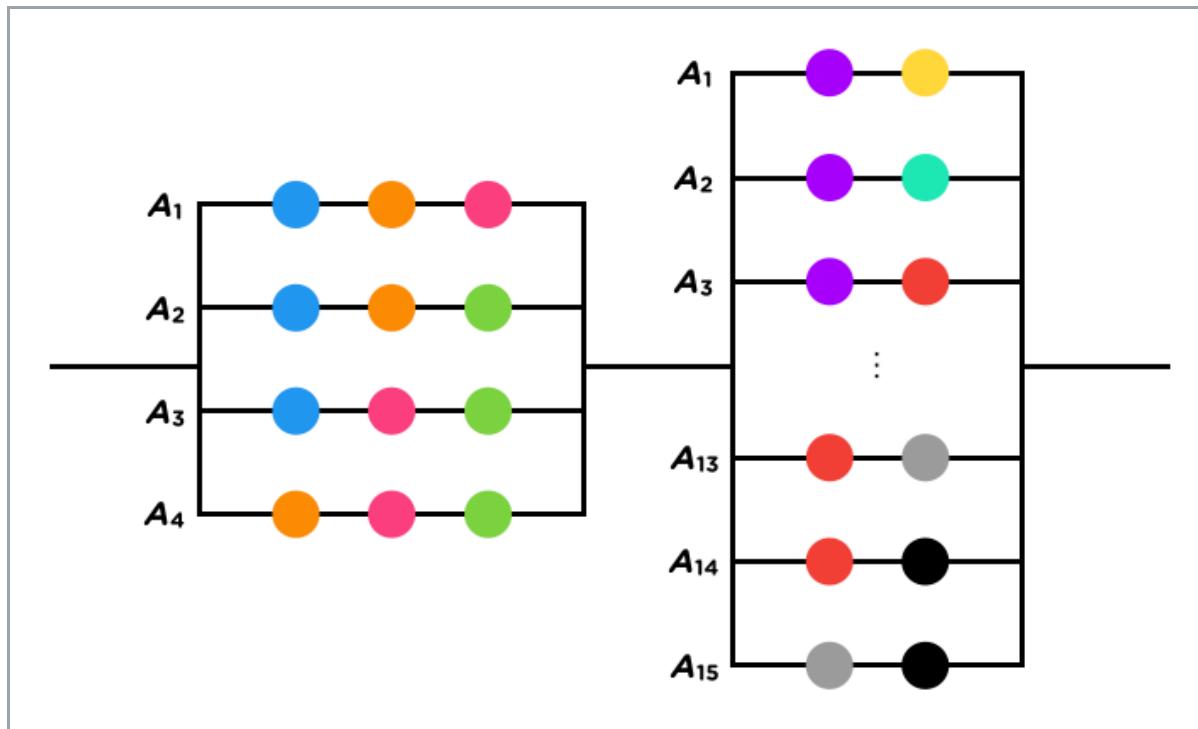
$$\binom{4}{3} = 4$$

minimal path sets. In addition, the 2-out-of-6 system has

$$\binom{6}{2} = 15$$

minimal path sets. If the two systems are placed in a series together, we need 1 out of the 4 minimal path sets from the first system **and** 1 out of the 15 minimal

path sets from the second system to be working in order for the series system to work.



So, the total number of minimal path sets is the product:

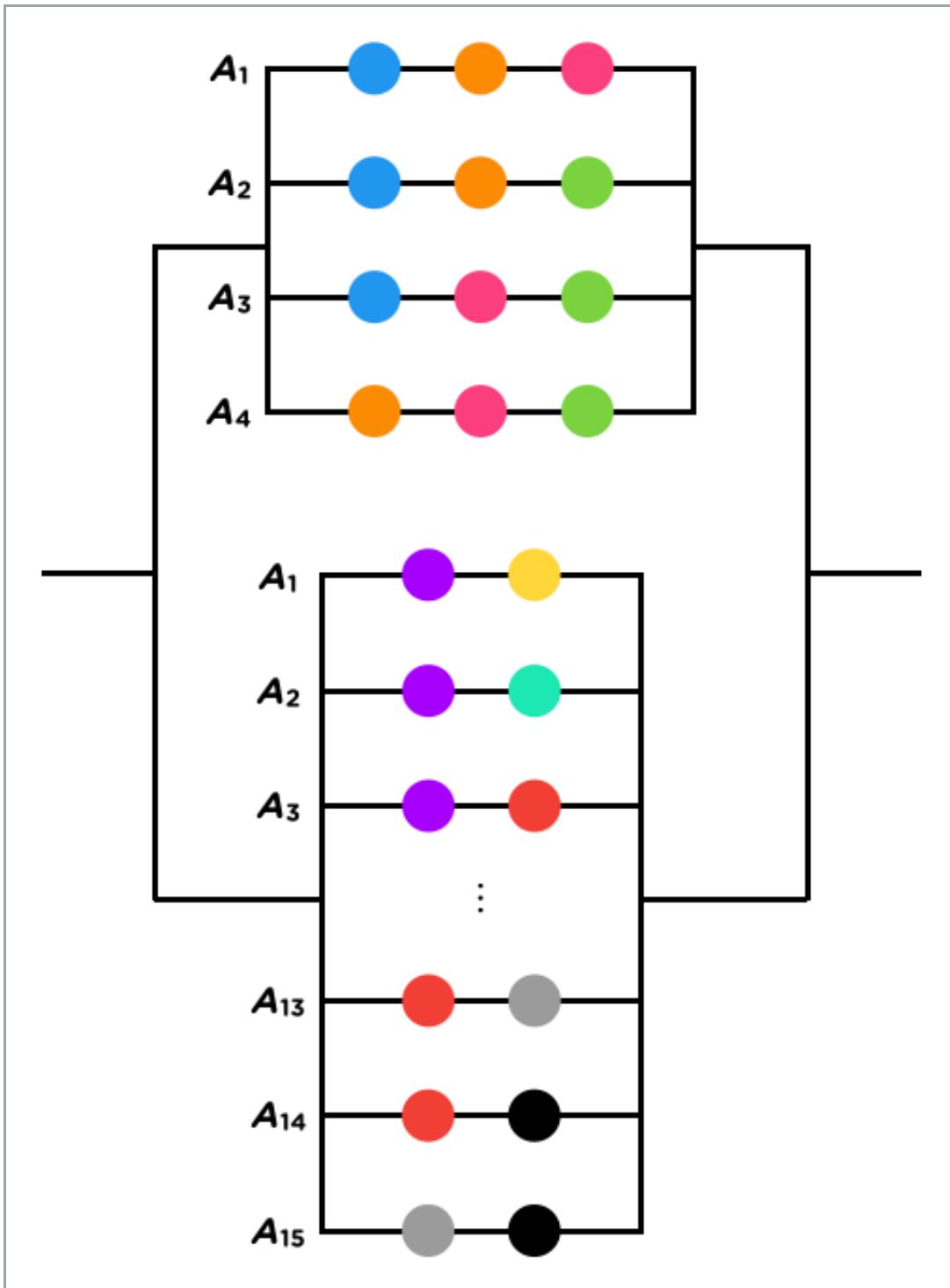
$$4 \cdot 15 = 60$$



Solution to (2)

The 3-out-of-4 system has 4 minimal path sets, and the 2-out-of-6 system has 15 minimal path sets. If the two systems are placed in a parallel system together,

we just need 1 out of the 4 minimal path sets from the first system **or** 1 out of the 15 minimal path sets from the second system to be working in order for the parallel system to work.



So, the total number of minimal path sets is the sum:

$$4 + 15 = 19$$



1.5.3 Structure Functions

(20m)

Structure functions, $\phi(\mathbf{x})$, are functions that represent the state of a system. They are functions of the state of each component in a system. We can derive the structure function of a system using either minimal path sets or minimal cut sets.

Before we proceed, let's examine the structure functions for a parallel system and for a series system.

A parallel system functions as long as at least one of the components functions. So, if at least one component has a state of 1, the system's structure function will equal 1. Thus, a parallel system with n components has the following structure function:

$$\phi(\mathbf{x}) = \max(x_1, \dots, x_n)$$

This should not be surprising because as long as any $x_i = 1$, $\phi(\mathbf{x}) = 1$.

A series system functions if and only if all of its components function. So, if at least one component has a state of 0, the system's structure function will equal 0. Then, a series system with n components has the structure function:

$$\phi(\mathbf{x}) = \min(x_1, \dots, x_n)$$

Again, this should not come as a surprise because if any $x_i = 0$, $\phi(\mathbf{x}) = 0$.

Since x_i is a binary variable, the following relationships are true:

$$\max(x_1, \dots, x_n) = 1 - \prod_{i=1}^n (1 - x_i) \quad (1.5.3.1)$$

$$\min(x_1, \dots, x_n) = \prod_{i=1}^n x_i \quad (1.5.3.2)$$

$$x_i^k = x_i, \quad k > 0 \quad (1.5.3.3)$$

Note that because x_i can only be 0 or 1, raising x_i to the k^{th} power (where $k > 0$) will not change its value.

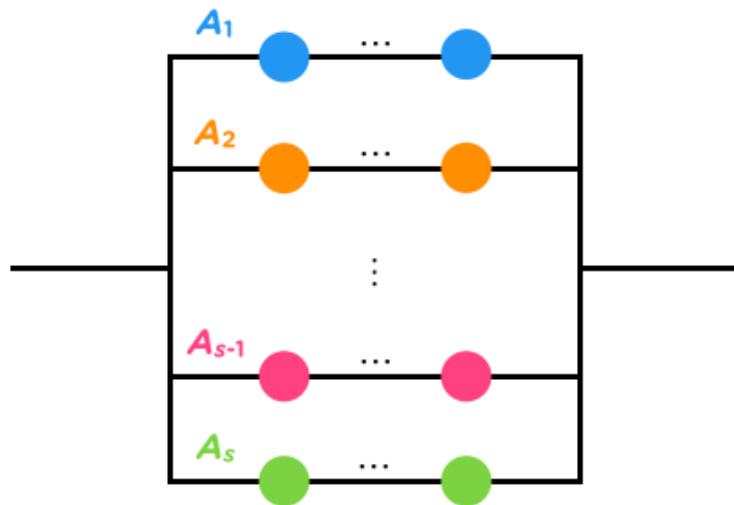
Minimal Path Sets Approach

Consider these two statements:

- A system will function if and only if all components of at least one minimal path set function.
- A parallel system will function if and only if at least one component functions.

So, we can treat any system as a parallel system of its minimal path sets. Therefore, given the minimal path sets of a system, we can express the structure function of the system the same way we express the structure function of a parallel system. For this, just treat the minimal path sets as the individual components of a parallel system.

Let A_1, \dots, A_s be the minimal path sets of a system. We can illustrate the system as such:

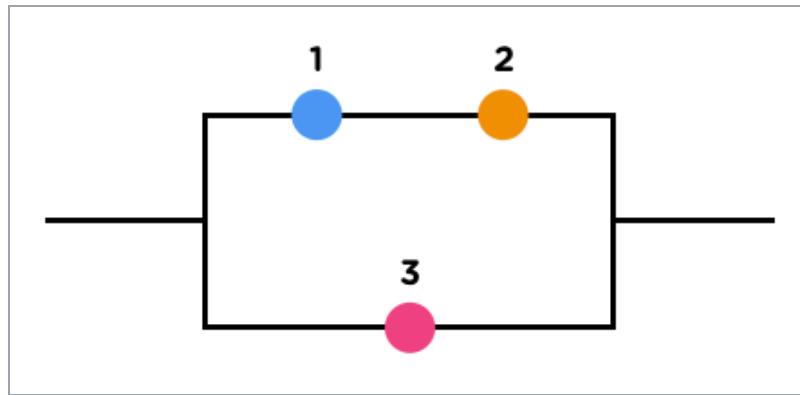


The structure function of the system can be defined as:

$$\begin{aligned}\phi(\mathbf{x}) &= \max \left(\min_{i \in A_1} x_i, \min_{i \in A_2} x_i, \dots, \min_{i \in A_s} x_i \right) \\ &= \max_j \prod_{i \in A_j} x_i\end{aligned}\tag{1.5.3.4}$$

Note that the symbol \in means "is an element of". So, for example, $i \in A_j$ means that we are considering all components i that are in the minimal path set A_j .

Consider the same system with three independent components.



Determine the structure function of the system using the minimal path sets approach.

We have determined that the minimal path sets are $A_1 = \{1, 2\}$ and $A_2 = \{3\}$. Thus, by using Equation 1.5.3.4, we have the structure function of this system:

$$\phi(\mathbf{x}) = \max(x_1x_2, x_3)$$

Then, use Equation 1.5.3.1 to simplify the structure function:

$$\begin{aligned}\phi(\mathbf{x}) &= \max(x_1x_2, x_3) \\ &= 1 - (1 - x_1x_2)(1 - x_3) \\ &= 1 - (1 - x_1x_2 - x_3 + x_1x_2x_3) \\ &= x_1x_2 + x_3 - x_1x_2x_3\end{aligned}$$

Minimal Cut Sets Approach

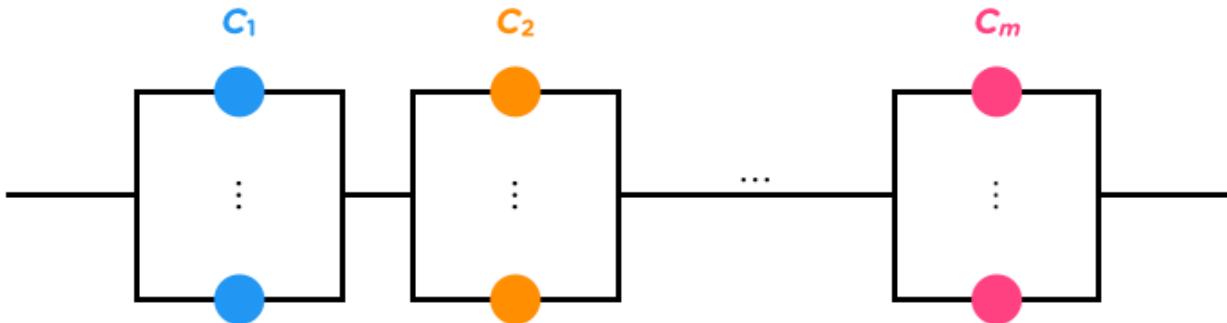
Similar to the relationship between minimal path sets and parallel systems, there is a relationship between minimal cut sets and series systems. Consider these two statements:

- A system will fail if and only if all components of at least one minimal cut set fail.
- A series system will fail if and only if at least one component fails.

So, any system can be thought of as a series system of its minimal cut sets. Therefore, given the

minimal cut sets of a system, we can express the structure function of the system the same way we express the structure function of a series system. For this, just treat the minimal cut sets as the individual components of a series system.

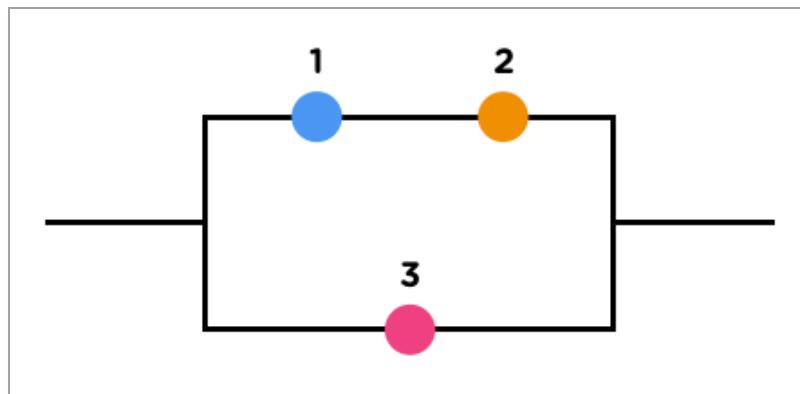
Let C_1, \dots, C_m be the minimal cut sets of a system. We can illustrate the system as such:



The structure function of the system can be defined as:

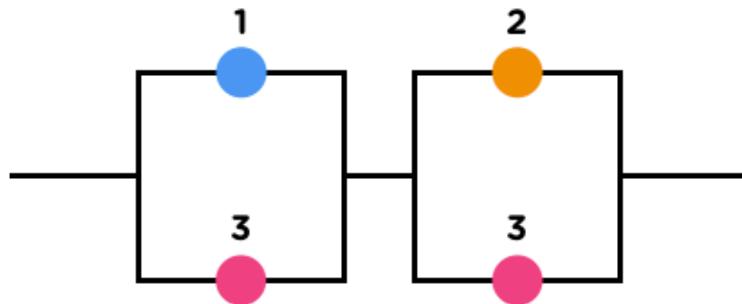
$$\begin{aligned}\phi(\mathbf{x}) &= \min \left(\max_{i \in C_1} x_i, \max_{i \in C_2} x_i, \dots, \max_{i \in C_m} x_i \right) \\ &= \prod_{j=1}^m \max_{i \in C_j} x_i\end{aligned}\tag{1.5.3.5}$$

Consider the same system with three independent components.



Determine the structure function of the system using the minimal cut sets approach.

The minimal cut sets are $C_1 = \{1, 3\}$ and $C_2 = \{2, 3\}$. Notice that the system can be represented using minimal cut sets as shown in the diagram below.



Thus, the structure function is:

$$\begin{aligned}
 \phi(\mathbf{x}) &= \max(x_1, x_3) \cdot \max(x_2, x_3) \\
 &= [1 - (1 - x_1)(1 - x_3)] \cdot [1 - (1 - x_2)(1 - x_3)] \\
 &= [1 - (1 - x_3 - x_1 + x_1 x_3)] \cdot [1 - (1 - x_3 - x_2 + x_2 x_3)] \\
 &= [x_3 + x_1 - x_1 x_3] \cdot [x_3 + x_2 - x_2 x_3] \\
 &= x_3^2 + x_2 x_3 - x_2 x_3^2 + x_1 x_3 + x_1 x_2 - x_1 x_2 x_3 - x_1 x_3^2 - x_1 x_2 x_3 + x_1 x_2 x_3^2 \\
 &= \mathbf{x}_1 \mathbf{x}_2 + \mathbf{x}_3 - \mathbf{x}_1 \mathbf{x}_2 \mathbf{x}_3
 \end{aligned}$$

Note that $x_i^2 = x_i$ due to Equation 1.5.3.3. The structure function is derived using both Equation 1.5.3.5 and Equation 1.5.3.1.

Example 1.5.3.1

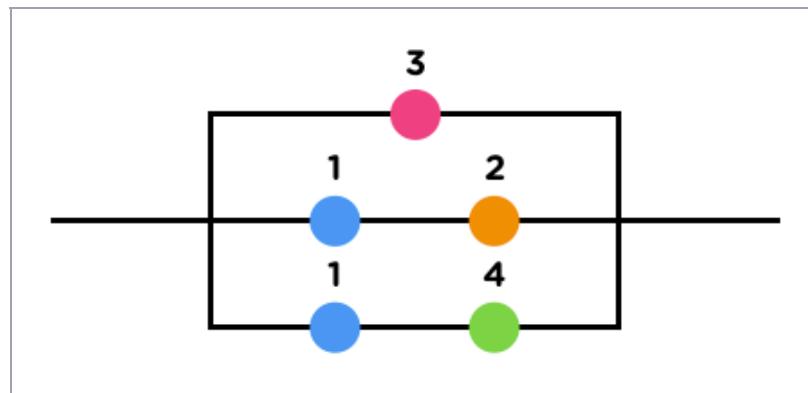
You are given the following information about a system of four components:

- The minimal path sets are $A_1 = \{3\}$, $A_2 = \{1, 2\}$, and $A_3 = \{1, 4\}$.
- All components in the system are independent.

Determine the structure function of the system.

Solution

Since we are given the minimal path sets, let's use the minimal path sets approach. The system can be illustrated as shown in the diagram below.



Using Equation 1.5.3.4,

$$\begin{aligned}
 \phi(\mathbf{x}) &= \max(x_3, x_1x_2, x_1x_4) \\
 &= 1 - (1 - x_3)(1 - x_1x_2)(1 - x_1x_4) \\
 &= 1 - (1 - x_1x_2 - x_3 + x_1x_2x_3)(1 - x_1x_4) \\
 &= 1 - (1 - x_1x_2 - x_3 + x_1x_2x_3 - x_1x_4 + x_1x_2x_4 + x_1x_3x_4 - x_1x_2x_3x_4) \\
 &= x_3 + x_1x_2 + x_1x_4 - x_1x_2x_3 - x_1x_3x_4 - x_1x_2x_4 + x_1x_2x_3x_4
 \end{aligned}$$

It is not necessary to expand the structure function. However, it will be helpful in the next subsection.



Example 1.5.3.2

You are given a three-out-of-four system with independent components.

Determine the structure function of the system.

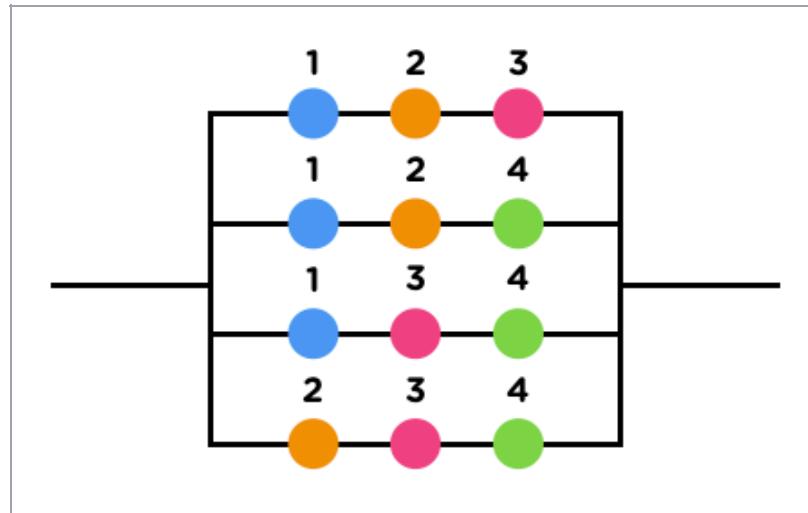
Solution

We can derive the structure function from the minimal path sets. A three-out-of-four system requires at least three components to be functioning for the system to function. So, the minimal path sets are all possible combinations of three components: $A_1 = \{1, 2, 3\}$, $A_2 = \{1, 2, 4\}$, $A_3 = \{1, 3, 4\}$, and $A_4 = \{2, 3, 4\}$.

Therefore, the structure function is:

$$\begin{aligned}\phi(\mathbf{x}) &= \max(x_1x_2x_3, x_1x_2x_4, x_1x_3x_4, x_2x_3x_4) \\ &= 1 - (1 - x_1x_2x_3)(1 - x_1x_2x_4)(1 - x_1x_3x_4)(1 - x_2x_3x_4)\end{aligned}$$

Note that the representation of the system using minimal path sets is:



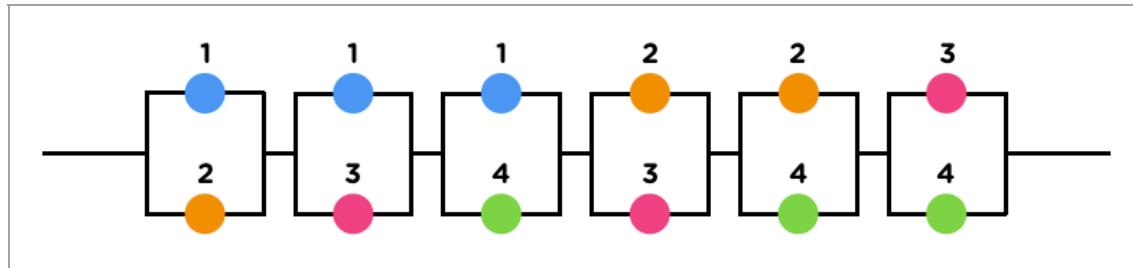
Alternative Solution

A three-out-of-four system will fail if at least two components fail. Therefore, the minimal cut sets are all possible combinations of two components: $C_1 = \{1, 2\}$, $C_2 = \{1, 3\}$, $C_3 = \{1, 4\}$, $C_4 = \{2, 3\}$, $C_5 = \{2, 4\}$, and $C_6 = \{3, 4\}$.

So, the structure function of a three-out-of-four system is:

$$\phi(\mathbf{x}) = \max(x_1, x_2) \cdot \max(x_1, x_3) \cdot \max(x_1, x_4) \cdot \max(x_2, x_3) \cdot \max(x_2, x_4) \cdot \max(x_3, x_4)$$

Note that the representation of the system using minimal cut sets is:



Expanding the structure functions will be messy, but doing so will prove that both methods yield the same structure function.



1.5.4 Reliability of Systems

🕒 35m

Now, we will focus on the reliability of a system and its reliability function. The reliability of a component is the probability that the component is functioning. Likewise, the *reliability of a system* is the probability that the system is functioning. The reliability of a system can be expressed as a function of the reliability of its components. This function is called the *reliability function*.

Let X_i be a Bernoulli random variable representing the state of component i , and let p_i be the probability that component i functions, i.e. $\Pr(X_i = 1)$. Then, we define two vectors:

- $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is the vector of the random variables for the components. Notice that this is similar to the state vector, just with random variables in place of the states.
- $\mathbf{p} = (p_1, p_2, \dots, p_n)$ is the vector of the reliabilities of the components in a system.

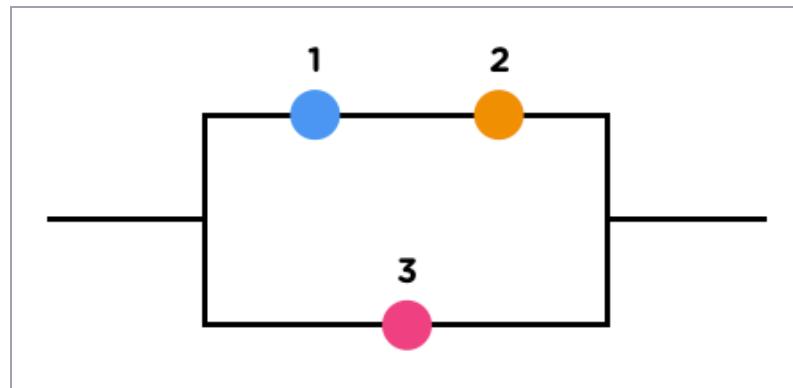
The reliability function of a system is denoted as $r(\mathbf{p})$. It is the sum of the probabilities of the path vectors. Equivalently, it is one minus the sum of the probabilities of the cut vectors.

$$r(\mathbf{p}) = \Pr[\phi(\mathbf{X}) = 1] = 1 - \Pr[\phi(\mathbf{X}) = 0] \quad (1.5.4.1)$$

If we know the structure function of a system, $\phi(\mathbf{x})$, we can determine its reliability function by calculating the expectation of $\phi(\mathbf{X})$. This is because $\phi(\mathbf{X})$ is a Bernoulli random variable. Note that \mathbf{x} is the state vector and \mathbf{X} is the random variable vector.

$$r(\mathbf{p}) = E[\phi(\mathbf{X})] \quad (1.5.4.2)$$

Consider the same system with three independent components.



Find the reliability function of the system.

Let's express the reliability function in terms of the sum of the probabilities of the path vectors. Recall from previous examples that this system has five path vectors.

$$\begin{aligned}
 r(\mathbf{p}) &= \Pr[\phi(\mathbf{X}) = 1] \\
 &= \Pr[\mathbf{X} = (0, 0, 1)] + \Pr[\mathbf{X} = (0, 1, 1)] + \Pr[\mathbf{X} = (1, 0, 1)] + \Pr[\mathbf{X} = (1, 1, 0)] + \Pr[\mathbf{X} = (1, 1, 1)] \\
 &= (1 - p_1)(1 - p_2)p_3 + \dots + p_1p_2p_3 \\
 &= p_3 - p_1p_3 - p_2p_3 + p_1p_2p_3 + \dots + p_1p_2p_3 \\
 &= \mathbf{p}_1\mathbf{p}_2 + \mathbf{p}_3 - \mathbf{p}_1\mathbf{p}_2\mathbf{p}_3
 \end{aligned}$$

We can also derive the reliability function by calculating the complement, i.e. one minus the probability of the three cut vectors.

$$\begin{aligned}
 r(\mathbf{p}) &= \Pr[\phi(\mathbf{X}) = 1] \\
 &= 1 - \Pr[\phi(\mathbf{X}) = 0] \\
 &= 1 - \Pr[\mathbf{X} = (0, 0, 0)] - \Pr[\mathbf{X} = (0, 1, 0)] - \Pr[\mathbf{X} = (1, 0, 0)] \\
 &= 1 - (1 - p_1)(1 - p_2)(1 - p_3) - \dots - p_1(1 - p_2)(1 - p_3) \\
 &= p_1p_2 + p_3 - p_1p_2p_3
 \end{aligned}$$

We know that the structure function of this system is:

$$\phi(\mathbf{x}) = x_1x_2 + x_3 - x_1x_2x_3$$

Since we know the structure function, another way to derive the reliability function is by computing the expectation of $\phi(\mathbf{X})$.

$$\begin{aligned}
 r(\mathbf{p}) &= E[\phi(\mathbf{X})] \\
 &= E[X_1X_2 + X_3 - X_1X_2X_3] \\
 &= E[X_1X_2] + E[X_3] - E[X_1X_2X_3] \\
 &= E[X_1] \cdot E[X_2] + E[X_3] - E[X_1] \cdot E[X_2] \cdot E[X_3] \\
 &= p_1p_2 + p_3 - p_1p_2p_3
 \end{aligned}$$

Note that since the components are all independent, $E[X_iX_j] = E[X_i] \cdot E[X_j]$ for $i \neq j$.

As can be seen with this example, the reliability function can often be derived in multiple ways. Choose the best approach depending on the information given in the question.

If you are given the reliability of each component, we can compute the reliability of the system.

If we are given the reliability of each component, we can compute the reliability of the system.

Suppose the reliability of each component is 0.7. The reliability of the system is

$$\begin{aligned} r &= 0.7(0.7) + 0.7 - 0.7(0.7)(0.7) \\ &= 0.847 \end{aligned}$$

If $r(\mathbf{p})$ is the reliability function of a system of independent components, then $r(\mathbf{p})$ is an increasing function of \mathbf{p} . This means that as the reliability of any given component increases, the reliability of the system increases as well. This type of system is called *monotone*.

Let's explore a few examples.

Example 1.5.4.1

You are given the following information about a system of four components:

- The minimal path sets are $A_1 = \{3\}$, $A_2 = \{1, 2\}$, and $A_3 = \{1, 4\}$.
- All components in the system are independent.
- All components have a reliability of 0.8.

Calculate the reliability of the system.

Solution

Let's solve this using basic probability. With four components, there are $2^4 = 16$ possible state vectors. We could list all of the path vectors and sum their probabilities, but this is not the most efficient way.

The system will function as long as component 3 functions. If component 3 fails, the system will function only when component 1 functions and at least one of the remaining components (2 and 4) functions. Therefore, the reliability is:

$$\begin{aligned} r &= 0.8 + 0.2(0.8)(0.8^2 + 0.2(0.8) + 0.8(0.2)) \\ &= 0.9536 \end{aligned}$$

where

- the term in red includes all path vectors with component 3 functioning.
- the term in blue represents the probability that component 3 fails.
- the term in green represents the probability that component 1 functions.
- the remaining terms represent the probabilities that components 2 and 4 function, component 2 fails and component 4 functions, and component 2 functions and component 4 fails, respectively. Note that this can also be calculated as the complement of the probability that components 2 and 4 both fail, or $1 - 0.2^2$.



Alternative Solution

Apply Equation 1.5.4.2 to calculate the reliability of the system. From Example 1.5.3.1, we know that the structure function is:

$$\phi(\mathbf{x}) = x_3 + x_1x_2 + x_1x_4 - x_1x_2x_3 - x_1x_3x_4 - x_1x_2x_4 + x_1x_2x_3x_4$$

Therefore, the reliability function is:

$$\begin{aligned} r(\mathbf{p}) &= E[\phi(\mathbf{X})] \\ &= E[X_3 + X_1X_2 + X_1X_4 - X_1X_2X_3 - X_1X_3X_4 - X_1X_2X_4 + X_1X_2X_3X_4] \\ &= p_3 + p_1p_2 + p_1p_4 - p_1p_2p_3 - p_1p_3p_4 - p_1p_2p_4 + p_1p_2p_3p_4 \end{aligned}$$

The reliability is:

$$\begin{aligned} r &= 0.8 + 0.8^2 + 0.8^2 - 0.8^3 - 0.8^3 - 0.8^3 + 0.8^4 \\ &= \mathbf{0.9536} \end{aligned}$$



Coach's Remarks

One advantage of deriving the structure function of a system is that we can then easily derive the reliability function as the expected value of the structure function. For this, make sure you don't just replace each x_i in a given structure function with p_i to get the reliability function. This will not always work. Consider the structure function we had for the series in Examples 1.5.3.1 and 1.5.4.1 before expanding:

$$\phi(\mathbf{x}) = 1 - (1 - x_3)(1 - x_1x_2)(1 - x_1x_4)$$

If we replace each x_i with p_i and evaluate the function at $p = 0.8$, we would get an answer of:

$$\begin{aligned} r(\mathbf{p}) &= 1 - (1 - p_3)(1 - p_1p_2)(1 - p_1p_4) \\ &= 1 - (1 - 0.8)(1 - 0.8^2)(1 - 0.8^2) \\ &= 0.9741 \end{aligned}$$

We can see that this reliability function does not give us the same result as the one in Example 1.5.4.1. This is because

$$E[(1 - X_3)(1 - X_1X_2)(1 - X_1X_4)] \neq (1 - E[X_3])(1 - E[X_1]E[X_2])(1 - E[X_1]E[X_4])$$

This would only hold if $E[X_1^2] = (E[X_1])^2$, but that is not the case here. To summarize, make sure to find the reliability function as the expected value of the structure function, not by replacing the x_i 's in the structure function with p_i 's. However, it is not always easy to derive the structure function, but the reliability of the system can often still be calculated using basic probability.

Example 1.5.4.2

You are given the following information about a three-out-of-four system:

You are given the following information about a three-out-of-four system.

- All components in the system are independent.
- Components 1 and 4 have a reliability of 0.8.
- Components 2 and 3 have a reliability of 0.9.

Calculate the reliability of the system.

Solution

From Example 1.5.3.2, we know that it is tedious to express the structure function of this system such that the reliability function would be straightforward to determine. So, we do not recommend calculating the reliability from the structure function. Instead, solve this question using basic probability. A three-out-of-four system requires at least three functioning components for the system to function. Thus, the system will function in any of these five scenarios:

- Components 1, 2, and 3 are functioning, and component 4 is not.
- Components 1, 2, and 4 are functioning, and component 3 is not.
- Components 1, 3, and 4 are functioning, and component 2 is not.
- Components 2, 3, and 4 are functioning, and component 1 is not.
- All components are functioning.

Thus, the reliability of the system, which is the probability that the system is functioning, is the sum of the probabilities of the five scenarios.

Scenario	Probability of Scenario
1	$p_1 p_2 p_3 (1 - p_4) = 0.1296$
2	$p_1 p_2 p_4 (1 - p_3) = 0.0576$
3	$p_1 p_3 p_4 (1 - p_2) = 0.0576$
4	$p_2 p_3 p_4 (1 - p_1) = 0.1296$
5	$p_1 p_2 p_3 p_4 = 0.5184$
Total	$r = 0.8928$



Coach's Remarks

Note that we have indirectly derived one form of the reliability function. So, we can express the reliability function of a three-out-of-four system as:

$$\begin{aligned} r(\mathbf{p}) &= p_1 p_2 p_3 (1 - p_4) + p_1 p_2 p_4 (1 - p_3) + p_1 p_3 p_4 (1 - p_2) + p_2 p_3 p_4 (1 - p_1) + p_1 p_2 p_3 p \\ &= p_1 p_2 p_3 + p_1 p_2 p_4 + p_1 p_3 p_4 + p_2 p_3 p_4 - 3p_1 p_2 p_3 p_4 \end{aligned}$$

Then, the structure function can be derived from the reliability function. Thus, the structure function is:

$$\phi(\mathbf{x}) = x_1 x_2 x_3 + x_1 x_2 x_4 + x_1 x_3 x_4 + x_2 x_3 x_4 - 3x_1 x_2 x_3 x_4$$

Example 1.5.4.3

Assume you want to build a series system with three types of components, where each type of component has a reliability of 0.6. You have two of each type of component, and all components in the system are independent.

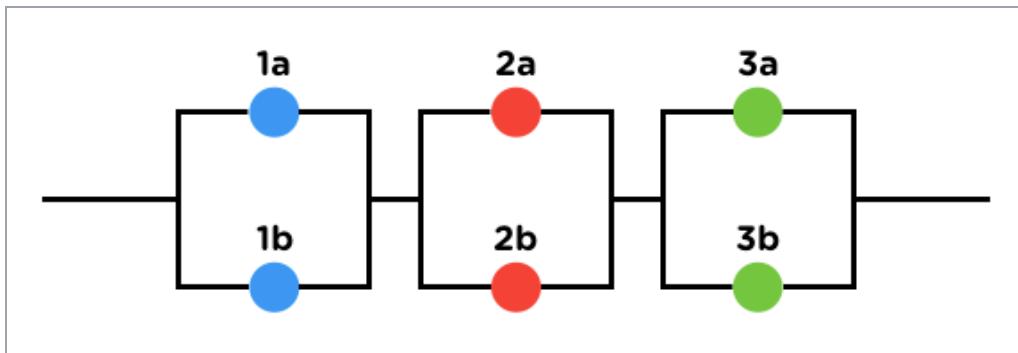
Calculate the probability of having a functioning system if

1. each component type has a backup (i.e. the system is a series of three parallel systems).
2. the entire series system has a backup (i.e. the system is a parallel of two series systems).

Solution to (1)

Denote the two components of the first type as 1a and 1b, the two components of the second type as 2a and 2b, and the two components of the third type as 3a and 3b.

With a system where each component has a backup, the system will function as long as at least one of each type of component functions. In other words, the system will fail if both components of the same type fail.



Therefore, the minimal cut sets are $\{1a, 1b\}$, $\{2a, 2b\}$, and $\{3a, 3b\}$. Use the minimal cut sets approach to determine the structure function.

$$\begin{aligned}
 \phi(\mathbf{x}) &= \max(x_{1a}, x_{1b}) \cdot \max(x_{2a}, x_{2b}) \cdot \max(x_{3a}, x_{3b}) \\
 &= [1 - (1 - x_{1a})(1 - x_{1b})] \cdot [1 - (1 - x_{2a})(1 - x_{2b})] \cdot [1 - (1 - x_{3a})(1 - x_{3b})] \\
 &= (x_{1a} + x_{1b} - x_{1a}x_{1b}) \cdot (x_{2a} + x_{2b} - x_{2a}x_{2b}) \cdot (x_{3a} + x_{3b} - x_{3a}x_{3b})
 \end{aligned}$$

Finally, the probability that the system functions is

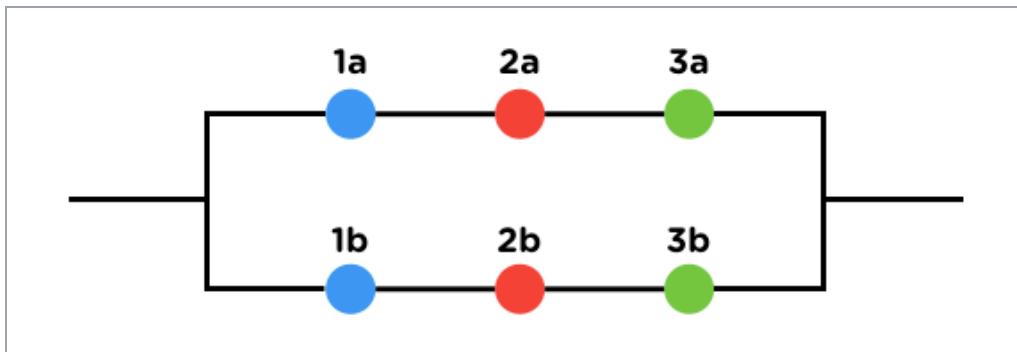
$$\begin{aligned}
 r(\mathbf{p}) &= E[\phi(\mathbf{X})] \\
 &= E[(X_{1a} + X_{1b} - X_{1a}X_{1b}) \cdot (X_{2a} + X_{2b} - X_{2a}X_{2b}) \cdot (X_{3a} + X_{3b} - X_{3a}X_{3b})] \\
 &= (p + p - p^2)^3 \\
 &= (0.6 + 0.6 - 0.6^2)^3 \\
 &= \mathbf{0.5927}
 \end{aligned}$$

■

Solution to (2)

Let 1a, 2a, and 3a denote the components for the first system and 1b, 2b, and 3b

denote the components for the backup system. With a parallel of two series systems, we need all three components for at least one of the systems to be functioning in order to have at least one functioning system.



Therefore, the minimal path sets are $\{1a, 2a, 3a\}$ and $\{1b, 2b, 3b\}$. Use the minimal path sets approach to determine the structure function.

$$\begin{aligned}\phi(\mathbf{x}) &= \max(x_{1a}x_{2a}x_{3a}, x_{1b}x_{2b}x_{3b}) \\ &= 1 - (1 - x_{1a}x_{2a}x_{3a})(1 - x_{1b}x_{2b}x_{3b}) \\ &= x_{1a}x_{2a}x_{3a} + x_{1b}x_{2b}x_{3b} - x_{1a}x_{2a}x_{3a}x_{1b}x_{2b}x_{3b}\end{aligned}$$

Finally, the probability that at least one of the systems functions is

$$\begin{aligned}r(\mathbf{p}) &= E[\phi(\mathbf{X})] \\ &= E[X_{1a}X_{2a}X_{3a} + X_{1b}X_{2b}X_{3b} - X_{1a}X_{2a}X_{3a}X_{1b}X_{2b}X_{3b}] \\ &= p^3 + p^3 - p^6 \\ &= 0.6^3 + 0.6^3 - 0.6^6 \\ &= \mathbf{0.3853}\end{aligned}$$

■

Coach's Remarks

Notice that the reliability is higher in the system where each component has a backup, rather than where the entire system has a backup. This is true in general for a system with n components. It is better to have a backup for each component than a backup for the whole system.

1.5.5 Bounds of Reliability Functions

(L) 30m

The previous subsection explained how to derive the exact reliability function of a system. However, it is sometimes impractical or unnecessary to derive the exact reliability function. If that is the case, we can approximate the actual reliability function of a system by determining the upper and lower bounds of that reliability function.

There are two methods to determine the bounds of a reliability function. Note that the proofs of these formulas are beyond the scope of this exam, but it is still important to understand the formulas.

Method of Inclusion and Exclusion

Recall that the probability of the union of two events, E_1 and E_2 , is:

$$\Pr(E_1 \cup E_2) = \Pr(E_1) + \Pr(E_2) - \Pr(E_1 \cap E_2)$$

The probability of the union of three events, E_1 , E_2 , and E_3 , is:

$$\begin{aligned} \Pr(E_1 \cup E_2 \cup E_3) &= \Pr(E_1) + \Pr(E_2) + \Pr(E_3) \\ &\quad - \Pr(E_1 \cap E_2) - \Pr(E_1 \cap E_3) - \Pr(E_2 \cap E_3) \\ &\quad + \Pr(E_1 \cap E_2 \cap E_3) \end{aligned}$$

If we generalize the probability of the union of n events, we have:

$$\begin{aligned} \Pr\left(\bigcup_{j=1}^n E_j\right) &= \sum_{j=1}^n \Pr(E_j) \\ &\quad - \sum_{j=1}^n \sum_{k>j} \Pr(E_j \cap E_k) \\ &\quad + \sum_{j=1}^n \sum_{k>j} \sum_{l>k} \Pr(E_j \cap E_k \cap E_l) \\ &\quad - \dots + (-1)^{n+1} \Pr(E_1 \cap E_2 \cap \dots \cap E_n) \end{aligned}$$

This formula will be the basis for determining the bounds of reliability functions under the *method of inclusion and exclusion*.

Notice that if we only include the first term $\sum_{j=1}^n \Pr(E_j)$, we are overestimating the probability of the union. Similarly, if we include only $\sum_{j=1}^n \Pr(E_j) - \sum_{j=1}^n \sum_{k>j} \Pr(E_j \cap E_k)$, we are underestimating the probability of the union. However, these two estimates together will create bounds around the actual reliability. Specifically, the actual reliability will be between $\sum_{j=1}^n \Pr(E_j) - \sum_{j=1}^n \sum_{k>j} \Pr(E_j \cap E_k)$ and $\sum_{j=1}^n \Pr(E_j)$.

In general, this formula will result in a set of inequalities, which we will be using to determine the bounds. The method of inclusion and exclusion can be based on either minimal path sets or minimal cut sets.

MINIMAL PATH SETS

Let A_1, \dots, A_s be the minimal path sets and p_i be the reliability of component i . The bounds of the system's reliability are:

$$\begin{aligned}
 r(\mathbf{p}) &\leq \sum_{j=1}^s \left(\prod_{i \in A_j} p_i \right) \\
 r(\mathbf{p}) &\geq \sum_{j=1}^s \left(\prod_{i \in A_j} p_i \right) - \sum_{j=1}^s \sum_{k>j} \left(\prod_{i \in A_j \cup A_k} p_i \right) \\
 r(\mathbf{p}) &\leq \sum_{j=1}^s \left(\prod_{i \in A_j} p_i \right) - \sum_{j=1}^s \sum_{k>j} \left(\prod_{i \in A_j \cup A_k} p_i \right) + \sum_{j=1}^s \sum_{k>j} \sum_{l>k} \left(\prod_{i \in A_j \cup A_k \cup A_l} p_i \right) \\
 &\vdots
 \end{aligned} \tag{1.5.5.1}$$

The bounds of the reliability will converge to the exact reliability as more terms are included. This means that the upper bound from the third line above is more precise than the upper bound from the first line above.

Coach's Remarks

Including all possible terms in the formula will result in the exact reliability function of the system. So, this can be another method to derive the reliability function. This technique will be demonstrated in Example 1.5.5.1.

This approximation method using minimal path sets is most effective when the reliabilities of the components are low. If the reliabilities are high, the reliability interval will widen and become less useful.

Note that the reliabilities for each component do not have to be the same, but examples will often assume the same reliability for all components in order to simplify calculations.

Consider a three-out-of-four system where all components have a reliability of n .

Determine the bounds of the reliability of the system using the first two inclusion-exclusion bounds based on the minimal path sets.

From Example 1.5.3.2, we know the minimal path sets of a three-out-of-four system are:

$$A_1 = \{1, 2, 3\}, \quad A_2 = \{1, 2, 4\}, \quad A_3 = \{1, 3, 4\}, \quad A_4 = \{2, 3, 4\}$$

The first term of the bounds formula is:

$$\begin{aligned} \sum_{j=1}^4 \left(\prod_{i \in A_j} p_i \right) &= \prod_{i \in A_1} p_i + \prod_{i \in A_2} p_i + \prod_{i \in A_3} p_i + \prod_{i \in A_4} p_i \\ &= p_1 p_2 p_3 + p_1 p_2 p_4 + p_1 p_3 p_4 + p_2 p_3 p_4 \\ &= 4p^3 \end{aligned}$$

The second term is:

$$\begin{aligned} \sum_{j=1}^4 \sum_{k>j} \left(\prod_{i \in A_j \cup A_k} p_i \right) &= \prod_{i \in A_1 \cup A_2} p_i + \prod_{i \in A_1 \cup A_3} p_i + \prod_{i \in A_1 \cup A_4} p_i + \prod_{i \in A_2 \cup A_3} p_i + \prod_{i \in A_2 \cup A_4} p_i + \prod_{i \in A_3 \cup A_4} p_i \\ &= 6p_1 p_2 p_3 p_4 \\ &= 6p^4 \end{aligned}$$

Since \mathbf{p} consists of p only, we will let $r(p)$ denote the system's reliability function. Thus, the first two terms yield the following inclusion-exclusion bounds:

$$4p^3 - 6p^4 \leq r(p) \leq 4p^3$$

Coach's Remarks

If $p = 0.2$, the reliability bounds would be:

$$\begin{aligned} 4(0.2)^3 - 6(0.2)^4 &\leq r(0.2) \leq 4(0.2)^3 \\ 0.0224 &\leq r(0.2) \leq 0.032 \end{aligned}$$

The exact reliability of a three-out-of-four system with $p = 0.2$ is:

$$\begin{aligned} r(0.2) &= p_1 p_2 p_3 + p_1 p_2 p_4 + p_1 p_3 p_4 + p_2 p_3 p_4 - 3p_1 p_2 p_3 p_4 \\ &= 4p^3 - 3p^4 \\ &= 0.0272 \end{aligned}$$

Notice 0.0272 is within the bounds.

If we increase the reliability of the components to $p = 0.6$, the bounds become:

$$\begin{aligned} 4(0.6)^3 - 6(0.6)^4 &\leq r(0.6) \leq 4(0.6)^3 \\ 0.0864 &\leq r(0.6) \leq 0.864 \end{aligned}$$

Compare this to the interval of 0.0224 to 0.032 for $p = 0.2$. Notice that when p is high, the interval is significantly wider and does not provide much insight into the reliability of the system. This shows that the minimal path sets approach is more effective for small values of p . However, the width of the interval can be shortened by including more terms in the bounds formula.

MINIMAL CUT SETS

Let C_1, \dots, C_m be the minimal cut sets and p_i be the reliability of component i . The bounds of the system's reliability are:

$$\begin{aligned} 1 - r(\mathbf{p}) &\leq \sum_{j=1}^m \left(\prod_{i \in C_j} (1 - p_i) \right) \\ 1 - r(\mathbf{p}) &\geq \sum_{j=1}^m \left(\prod_{i \in C_j} (1 - p_i) \right) - \sum_{j=1}^m \sum_{k>j} \left(\prod_{i \in C_j \cup C_k} (1 - p_i) \right) \\ 1 - r(\mathbf{p}) &\leq \sum_{j=1}^m \left(\prod_{i \in C_j} (1 - p_i) \right) - \sum_{j=1}^m \sum_{k>j} \left(\prod_{i \in C_j \cup C_k} (1 - p_i) \right) + \sum_{j=1}^m \sum_{k>j} \sum_{l>k} \left(\prod_{i \in C_j \cup C_k \cup C_l} (1 - p_i) \right) \\ &\vdots \end{aligned}$$

The bounds of the reliability will converge to the exact reliability as more terms are included, similar to the minimal path sets approach.

This approximation method using minimal cut sets is most effective when the reliabilities of the components are very high. If the reliabilities are low, the reliability interval will widen and become less useful.

Consider a three-out-of-four system where all components have a reliability of p .

Determine the bounds of the reliability of the system using the first two inclusion-exclusion bounds based on the minimal cut sets.

From Example 1.5.3.2, we know the minimal cut sets of a three-out-of-four system are:

$$\begin{aligned} C_1 &= \{1, 2\}, & C_2 &= \{1, 3\}, & C_3 &= \{1, 4\} \\ C_4 &= \{2, 3\}, & C_5 &= \{2, 4\}, & C_6 &= \{3, 4\} \end{aligned}$$

The first term of the bounds formula is:

$$\begin{aligned} \sum_{j=1}^6 \left(\prod_{i \in C_j} (1-p_i) \right) &= \prod_{i \in C_1} (1-p_i) + \prod_{i \in C_2} (1-p_i) + \prod_{i \in C_3} (1-p_i) + \dots + \prod_{i \in C_6} (1-p_i) \\ &= 6(1-p)^2 \end{aligned}$$

The second term is:

$$\begin{aligned} \sum_{j=1}^6 \sum_{k>j} \left(\prod_{i \in C_j \cup C_k} (1-p_i) \right) &= \prod_{i \in C_1 \cup C_2} (1-p_i) + \prod_{i \in C_1 \cup C_3} (1-p_i) + \dots + \prod_{i \in C_5 \cup C_6} (1-p_i) \\ &= 12(1-p)^3 + 3(1-p)^4 \end{aligned}$$

Note that there are $\binom{6}{2} = 15$ terms in total for the second term.

Let $r(p) = r(p, p, p, p)$. Thus, the first two inclusion-exclusion bounds yield:

$$6(1-p)^2 - 12(1-p)^3 - 3(1-p)^4 \leq 1 - r(p) \leq 6(1-p)^2$$

Coach's Remarks

If $p = 0.9$, the bounds would be:

$$6(1-0.9)^2 - 12(1-0.9)^3 - 3(1-0.9)^4 \leq 1 - r(0.9) \leq 6(1-0.9)^2$$

$0.0177 < 1 - r(0.9) < 0.06$

$$0.94 \leq r(0.9) \leq 0.9523$$

The exact reliability of a three-out-of-four system is:

$$r(0.9) = 0.9477$$

Notice 0.9477 is within the bounds.

If $p = 0.6$, the bounds would be:

$$\begin{aligned} 6(1 - 0.6)^2 - 12(1 - 0.6)^3 - 3(1 - 0.6)^4 &\leq 1 - r(0.6) \leq 6(1 - 0.6)^2 \\ 0.1152 &\leq 1 - r(0.6) \leq 0.96 \\ 0.04 &\leq r(0.6) \leq 0.8848 \end{aligned}$$

Notice that the interval is significantly wider and does not provide much insight on the reliability. This shows the minimal cut sets approach is more effective for large values of p .

Example 1.5.5.1

You are given the following information about a system of four components:

- The minimal path sets are $A_1 = \{3\}$, $A_2 = \{1, 2\}$, and $A_3 = \{1, 4\}$.
- All components in the system are independent.
- The reliability for all components is 0.4.

Calculate the lower bound reliability of the system using the first two inclusion-exclusion bounds based on the minimal path sets.

Solution

The first term is:

$$\begin{aligned} \sum_{j=1}^3 \left(\prod_{i \in A_j} p_i \right) &= \prod_{i \in A_1} p_i + \prod_{i \in A_2} p_i + \prod_{i \in A_3} p_i \\ &= p_3 + p_1 p_2 + p_1 p_4 \\ &= 0.79 \end{aligned}$$

... -

The second term is:

$$\begin{aligned} \sum_{j=1}^3 \sum_{k>j} \left(\prod_{i \in A_j \cup A_k} p_i \right) &= \prod_{i \in A_1 \cup A_2} p_i + \prod_{i \in A_1 \cup A_3} p_i + \prod_{i \in A_2 \cup A_3} p_i \\ &= p_1 p_2 p_3 + p_1 p_3 p_4 + p_1 p_2 p_4 \\ &= 0.192 \end{aligned}$$

Thus, the lower bound (using two terms) is:

$$\sum_{j=1}^3 \left(\prod_{i \in A_j} p_i \right) - \sum_{j=1}^3 \sum_{k>j} \left(\prod_{i \in A_j \cup A_k} p_i \right) = \mathbf{0.528}$$

■

Coach's Remarks

With only three minimal path sets, we can only go up to the third term in the bounds formula. Let's calculate the third term.

$$\begin{aligned} \sum_{j=1}^3 \sum_{k>j} \sum_{l>k} \left(\prod_{i \in A_j \cup A_k \cup A_l} p_i \right) &= \prod_{i \in A_1 \cup A_2 \cup A_3} p_i \\ &= p_1 p_2 p_3 p_4 \end{aligned}$$

The bound (using three terms) is:

$$p_3 + p_1 p_2 + p_1 p_4 - p_1 p_2 p_3 - p_1 p_3 p_4 - p_1 p_2 p_4 + p_1 p_2 p_3 p_4$$

Notice that the expression above is the reliability function of the system as shown in Example 1.5.4.1.

Method of Intersection

Instead of the union of events, this method is based on the intersection of events.

Let A_1, \dots, A_s be the minimal path sets, C_1, \dots, C_m be the minimal cut sets, and p_i be the reliability of component i . The bounds of the system's reliability under the *method of intersection* are:

$$\prod_{j=1}^m \left[1 - \prod_{i \in C_j} (1 - p_i) \right] \leq r(\mathbf{p}) \leq 1 - \prod_{j=1}^s \left[1 - \prod_{i \in A_j} p_i \right] \quad (1.5.5.3)$$

The lower bound represents the probability that at least one component in each minimal cut set is functioning, and the upper bound represents the probability that all of the components in at least one minimal path are functioning.

The method of intersection works best when the values of p_i are all very large or very small.

Consider a three-out-of-four system where all components have a reliability of p .

Determine the bounds of the reliability of the system using the method of intersection.

The minimal path sets of a three-out-of-four system are:

$$A_1 = \{1, 2, 3\}, \quad A_2 = \{1, 2, 4\}, \quad A_3 = \{1, 3, 4\}, \quad A_4 = \{2, 3, 4\}$$

The upper bound of the reliability function is:

$$\begin{aligned} 1 - \prod_{j=1}^4 \left[1 - \prod_{i \in A_j} p_i \right] &= 1 - \left[1 - \prod_{i \in A_1} p_i \right] \left[1 - \prod_{i \in A_2} p_i \right] \left[1 - \prod_{i \in A_3} p_i \right] \left[1 - \prod_{i \in A_4} p_i \right] \\ &= 1 - (1 - p^3)^4 \end{aligned}$$

The minimal cut sets of a three-out-of-four system are:

$$\begin{aligned} C_1 &= \{1, 2\}, \quad C_2 = \{1, 3\}, \quad C_3 = \{1, 4\} \\ C_4 &= \{2, 3\}, \quad C_5 = \{2, 4\}, \quad C_6 = \{3, 4\} \end{aligned}$$

The lower bound of the reliability function is:

THE LOWER BOUND OF THE RELIABILITY FUNCTION IS.

$$\begin{aligned} \prod_{j=1}^6 \left[1 - \prod_{i \in C_j} (1 - p_i) \right] &= \left[1 - \prod_{i \in C_1} (1 - p_i) \right] \dots \left[1 - \prod_{i \in C_6} (1 - p_i) \right] \\ &= \left(1 - (1 - p)^2 \right)^6 \end{aligned}$$

Let $r(p) = r(p, p, p, p)$. Thus, the bounds are:

$$\left(1 - (1 - p)^2 \right)^6 \leq r(p) \leq 1 - (1 - p^3)^4$$

Coach's Remarks

If $p = 0.2$, the bounds would be:

$$0.0022 \leq r(0.2) \leq 0.0316$$

where $r(0.2) = 0.0272$.

If $p = 0.9$, the bounds would be:

$$0.9415 \leq r(0.9) \leq 0.9946$$

where $r(0.9) = 0.9477$.

Both intervals are tight; hence, they approximate the reliability well.

1.5.6 Random Graphs

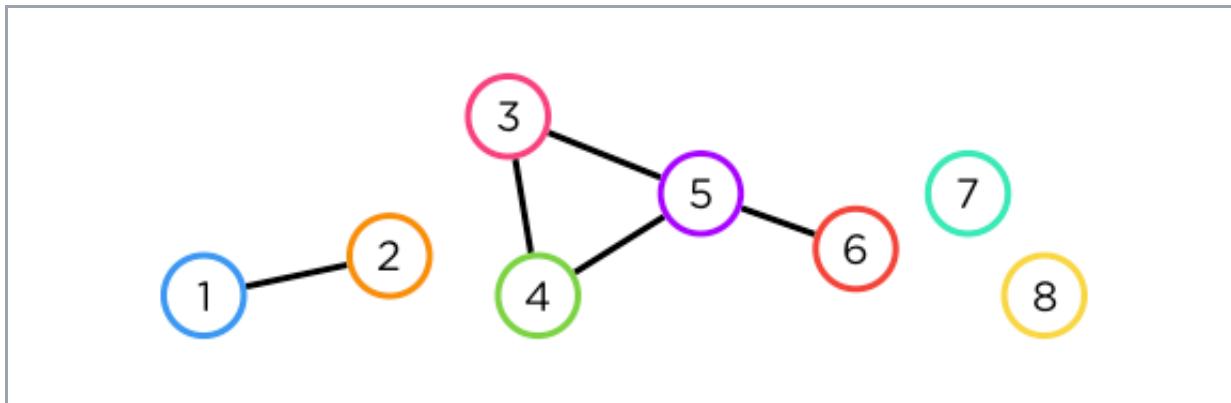
🕒 15m

Many of the concepts discussed in the previous subsections can be applied to random graphs. But before considering random graphs, it is important to understand a graph.

A graph consists of a set of *nodes* and a set of *arcs* connecting the nodes. Let N be the set of nodes and A be the set of arcs. A graph with n nodes has a maximum of $\binom{n}{2}$ arcs. This is because there are $\binom{n}{2}$ different pairs of nodes. A graph can be subdivided into subgraphs, called *components*, where each component consists of non-overlapping connected nodes.

A graph is *connected* if there is only one component. In other words, given any two nodes, there must be a path from one node to the other for the graph to be connected.

Consider the graph below.



List all the nodes, arcs, and components of this graph.

For this graph, there are

- eight nodes: $N = \{1, 2, 3, 4, 5, 6, 7, 8\}$.
- five arcs: $A = \{\{1, 2\}, \{3, 4\}, \{3, 5\}, \{4, 5\}, \{5, 6\}\}$.

- four components: $\{\{1, 2\}, \{3, 4, 5, 6\}, \{7\}, \{8\}\}$.

Since this graph has more than one component, it is not connected.

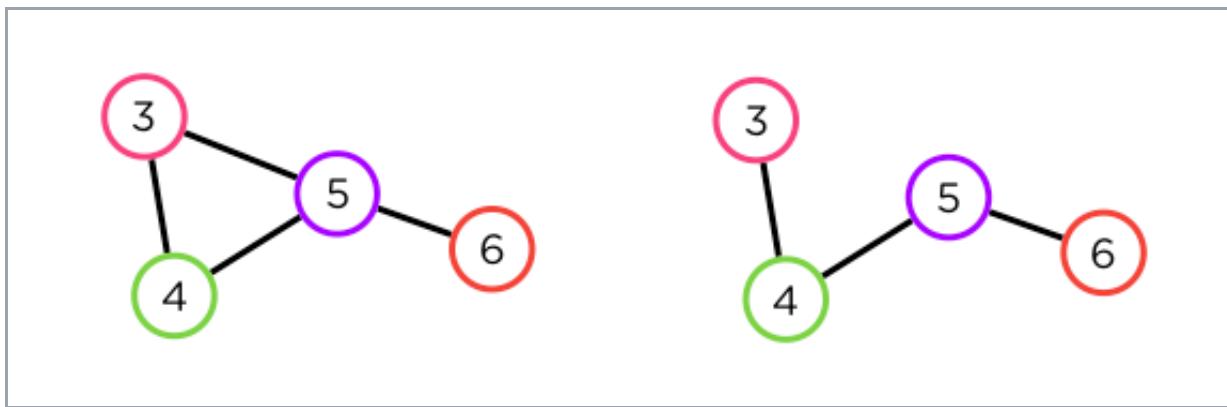
Now, consider a graph with n nodes, where any two nodes (node i and node j) are not connected with certainty but instead with a probability of $P_{i,j}$. This graph is known as a *random graph*. Let $X_{i,j}$ be a random variable that represents the existence of an arc between node i and node j .

$$X_{i,j} = \begin{cases} 1, & \text{if } \{i, j\} \text{ is an arc} \\ 0, & \text{otherwise} \end{cases}$$

Note that $\Pr(X_{i,j} = 1) = P_{i,j}$ and $\Pr(X_{i,j} = 0) = 1 - P_{i,j}$.

Suppose we are interested in the probability that a random graph is connected. Notice that a graph does not need to have every node connected to every other node in order for the graph to be connected.

Consider the graphs below.



Determine which graphs, if any, are connected.

Even though the arc $\{3, 5\}$ does not exist in the graph on the right, **both graphs are**

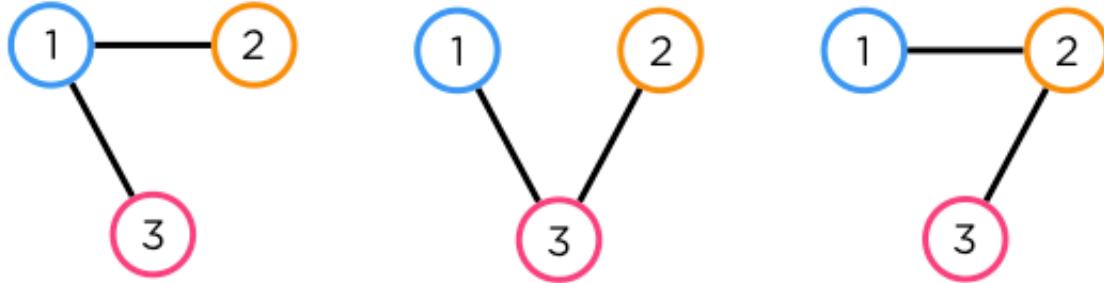
connected.

The connectedness of a graph is similar to the reliability of a system. The reliability of a system depends on the reliability of its components. A system will function if and only if all components of at least one minimal path set function.

Like a system, a graph has minimal path sets and minimal cut sets. A random graph is connected as long as all arcs of at least one minimal path set exist. A random graph with n nodes has n^{n-2} minimal path sets and $2^{n-1} - 1$ minimal cut sets.

Determine the minimal path sets for a random graph with three nodes.

A random graph with three nodes has $3^{3-2} = 3$ minimal path sets, as shown below.



Let's assume each random variable is independent and identically distributed, i.e. all $P_{i,j} = p$. Let P_n be the probability that a random graph with n nodes is connected. Then,

$$P_n = 1 - \sum_{k=1}^{n-1} \binom{n-1}{k-1} q^{k(n-k)} P_k, \quad n = 2, 3, \dots \quad (1.5.6.1)$$

where $P_1 = 1$, $P_2 = p$, and $q = 1 - p$. Note that this is a recursive formula. The value of P_n depends on the values of all P_i from $i = 1$ to $n - 1$. As n increases, the formula above gets more and more complicated.

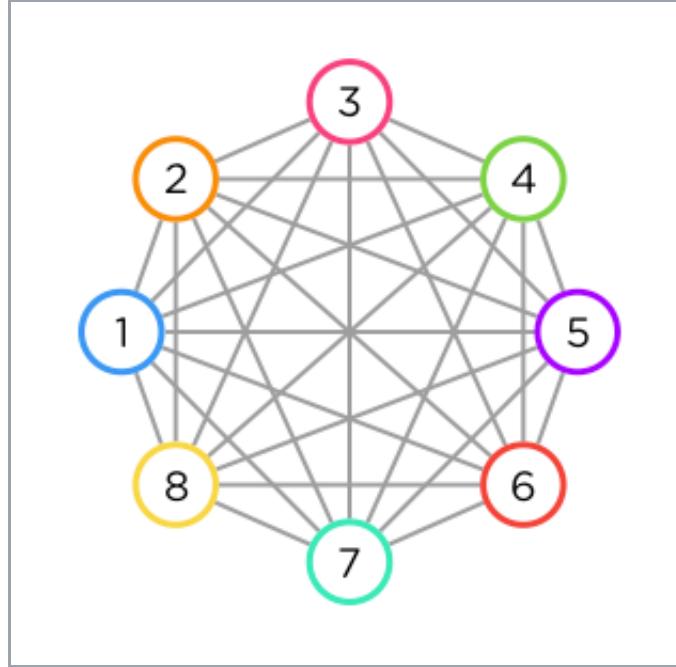
To avoid the tedious task of calculating the exact probability, we set up bounds similar to those for the reliability of a system. The corresponding bounds for P_n are:

$$nq^{n-1} - \binom{n}{2}q^{2n-3} \leq 1 - P_n \leq (n + 1)q^{n-1} \quad (1.5.6.2)$$

For a large n , it can be shown that:

$$P_n \approx 1 - nq^{n-1}$$

Consider a random graph with eight nodes. There are $\binom{8}{2} = 28$ random variables for the possible arcs. So, if every node is connected to every other node, there are 28 arcs, as shown in the diagram below.



Calculate the lower and upper bounds for P_8 at $p = 0.5$.

Use Equation 1.5.6.2 to calculate the bounds.

$$\begin{aligned} 8q^{8-1} - \binom{8}{2}q^{2(8)-3} &= 8(0.5^7) - 28(0.5^{13}) \\ &= 0.0591 \end{aligned}$$

$$\begin{aligned} (8+1)q^{8-1} &= 9(0.5^7) \\ &= 0.0703 \end{aligned}$$

Thus,

$$\begin{aligned} 0.0591 &\leq 1 - P_8 \leq 0.0703 \\ -0.9409 &\leq -P_8 \leq -0.9297 \\ \mathbf{0.9297} &\leq P_8 \leq \mathbf{0.9409} \end{aligned}$$

Note that the exact value of P_8 is 0.9371.

Example 1.5.6.1

You are given the following information on a graph:

- The graph has four nodes.
- The probability that there is an arc between any two nodes is 0.6.
- Every arc is independent.

Calculate the probability that the graph is connected.

Solution

Use the recursive formula to calculate P_4 . We know that:

$$\begin{aligned}P_1 &= 1 \\P_2 &= p = 0.6\end{aligned}$$

Let's start by calculating P_3 .

$$\begin{aligned}P_3 &= 1 - \sum_{k=1}^2 \binom{2}{k-1} q^{k(3-k)} P_k \\&= 1 - \left[\binom{2}{0} q^2 P_1 + \binom{2}{1} q^2 P_2 \right] \\&= 1 - q^2 - 2q^2 p \\&= 1 - 0.4^2 - 2(0.4^2)(0.6) \\&= 0.648\end{aligned}$$

The probability that the graph is connected is:

$$\begin{aligned}P_4 &= 1 - \sum_{k=1}^3 \binom{3}{k-1} q^{k(4-k)} P_k \\&= 1 - \left[\binom{3}{0} q^3 P_1 + \binom{3}{1} q^4 P_2 + \binom{3}{2} q^3 P_3 \right] \\&= 1 - q^3 - 3q^4 p - 3q^3 P_3 \\&= 1 - 0.4^3 - 3(0.4^4)(0.6) - 3(0.4^3)(0.648) \\&= \mathbf{0.7655}\end{aligned}$$



1.5.7 Lifetime of Systems

🕒 20m

In previous subsections, we examined how the reliability of a system is affected by the reliability of its components. Now, we will shift our attention to the lifetime of a system and examine how the lifetime of a system is impacted by the lifetime of its components.

Let T_i represent the lifetime of component i . Then,

- $S_i(t)$ is the survival function of the lifetime of component i .
- $\mathbf{S}(t) = (S_1(t), S_2(t), \dots, S_n(t))$ is the vector of the survival functions of the n components.

If we know the reliability function of a system, we can easily calculate the moments and probabilities for the lifetime of a system. Let T be the lifetime of a system. Then, the probability that the system lifetime exceeds t is:

$$\Pr(T > t) = r[\mathbf{S}(t)] \quad (1.5.7.1)$$

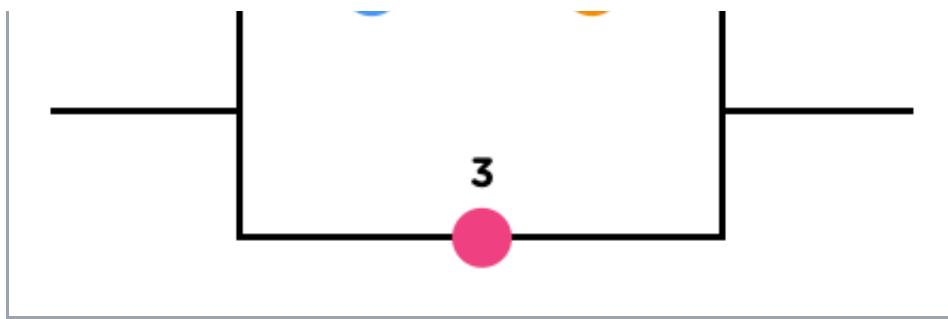
Note that $r[\mathbf{S}(t)]$ is the reliability function in terms of the components' survival functions instead of their reliabilities. In other words, we set $\mathbf{p} = \mathbf{S}(t)$, or equivalently, $p_i = S_i(t)$ for all i .

The expected value of the system lifetime is:

$$E[T] = \int_0^\infty r[\mathbf{S}(t)] dt \quad (1.5.7.2)$$

Consider the same system with three independent components. Assume the lifetimes of all components are uniformly distributed over $(0, 2)$.





Calculate the expected lifetime of the system.

The survival function for each component is:

$$S_i(t) = 1 - \frac{t}{2}$$

The reliability function of this three-component system is:

$$r(\mathbf{p}) = p_1 p_2 + p_3 - p_1 p_2 p_3$$

Therefore, the system lifetime has the following survival function:

$$\begin{aligned} \Pr(T > t) &= r[\mathbf{S}(t)] \\ &= \left(1 - \frac{t}{2}\right)^2 + 1 - \frac{t}{2} - \left(1 - \frac{t}{2}\right)^3 \\ &= \left(\frac{2-t}{2}\right)^2 + \frac{2-t}{2} - \left(\frac{2-t}{2}\right)^3 \\ &= \frac{t^3 - 4t^2 + 8}{8} \end{aligned}$$

To calculate the expected system lifetime, integrate the survival function from 0 to 2.

$$t^2 - 4t^2 + 8$$

$$\begin{aligned}
 E[T] &= \int_0^{\infty} \frac{t^4}{8} dt \\
 &= \frac{1}{8} \left[\frac{t^4}{4} - \frac{4t^3}{3} + 8t \right]_0^{\infty} \\
 &= 1.1667
 \end{aligned}$$

Coach's Remarks

Note that for a uniform distribution with interval $[a, b]$, the survival function is:

$$S(x) = \frac{b-x}{b-a}, \quad a \leq x \leq b$$

The hazard function is:

$$h(x) = \frac{1}{b-x}, \quad a \leq x \leq b$$

Structure Functions, Reliability Functions, and Survival Functions

As you may have noticed, there is a relationship between the structure function of a system, its reliability function, and the survival function of its lifetime. We can derive the reliability function of a system from its structure function and then derive the survival function of its lifetime from the reliability function.

To calculate probabilities and expected values involving the lifetime of systems, follow these steps:

1. Determine the structure function using minimal path sets or minimal cut sets.

2. **Derive the reliability function** from the structure function or using basic probability.
3. **Develop the survival function** from the reliability function.
4. **Calculate the desired probability or expected value.**

Let's revisit a few examples and calculate probabilities and expected lifetimes for those systems.

Example 1.5.7.1

You are given the following information about a system with four components:

- The minimal path sets are $A_1 = \{3\}$, $A_2 = \{1, 2\}$, and $A_3 = \{1, 4\}$.
- All components in the system are independent.
- The lifetimes of all components are uniformly distributed from 0 to 3.

Calculate the probability that the lifetime of the system is less than 1.5.

Solution

From Example 1.5.4.1, the reliability function of the system is:

$$r(\mathbf{p}) = p_3 + p_1p_2 + p_1p_4 - p_1p_2p_3 - p_1p_3p_4 - p_1p_2p_4 + p_1p_2p_3p_4$$

The survival function for each component is:

$$S_i(t) = 1 - \frac{t}{3}$$

Thus, the survival function for the lifetime of the system is:

$$\begin{aligned}S(t) &= r[\mathbf{S}(t)] \\&= \left(1 - \frac{t}{3}\right) + 2\left(1 - \frac{t}{3}\right)^2 - 3\left(1 - \frac{t}{3}\right)^3 + \left(1 - \frac{t}{3}\right)^4\end{aligned}$$

The probability that the lifetime of the system is less than 1.5 is:

$$\begin{aligned}\Pr(T < 1.5) &= 1 - \Pr(T \geq 1.5) \\&= 1 - S(1.5) \\&= 1 - 0.6875 \\&= \mathbf{0.3125}\end{aligned}$$



Example 1.5.7.2

You are given the following information about a three-out-of-four system:

- All components in the system are independent.
- The reliability for each component is exponentially distributed with mean 1.

Calculate the expected lifetime of the system.

Solution

From Example 1.5.4.2, the reliability function of a three-out-of-four system is:

$$r(\mathbf{p}) = p_1 p_2 p_3 + p_1 p_2 p_4 + p_1 p_3 p_4 + p_2 p_3 p_4 - 3p_1 p_2 p_3 p_4$$

The lifetime of each component follows an exponential distribution with mean 1. So, the survival function for each component is:

$$S_i(t) = e^{-t}$$

Thus, the survival function of the lifetime of the system is:

$$\begin{aligned} S(t) &= r[\mathbf{S}(t)] \\ &= 4e^{-3t} - 3e^{-4t} \end{aligned}$$

The expected lifetime of the system is:

$$\begin{aligned} E[T] &= \int_0^{\infty} 4e^{-3t} - 3e^{-4t} dt \\ &= \left[\frac{4e^{-3t}}{-3} - \frac{3e^{-4t}}{-4} \right]_0^{\infty} \\ &= \mathbf{0.5833} \end{aligned}$$



Coach's Remarks

For a k -out-of- n system of independent and identically distributed exponential components, each with mean θ , the mean lifetime of the system is

$$\mathbb{E}[T] = \theta \sum_{i=k}^n \frac{1}{i}$$

This formula comes from the order statistics of exponential random variables, covered in Section 2.7.

Example 1.5.7.3 (Adapted from CAS S F2016 8)

You are given the following information about a series system with two independent machines:

- The hazard rate for each machine (with x in months) is

$$\mu_x = \frac{1}{50-x}, \quad 0 \leq x < 50$$

- One machine has worked for 15 months, and the other machine has worked for 20 months.

Calculate the probability that the system will function for ten more months.

Solution

From the hazard rate, infer that each component's lifetime is uniformly

distributed from 0 to 50. Because this is a series system, the reliability function is the probability of both components functioning, i.e. $p_1 p_2$.

Thus, the probability that the system will function for another ten months is the probability that each component will function for 10 more months given their respective ages.

$$\begin{aligned}
 & \Pr(T_1 > 25 \mid T_1 > 15) \cdot \Pr(T_2 > 30 \mid T_2 > 20) \\
 &= \frac{\Pr(T_1 > 25)}{\Pr(T_1 > 15)} \cdot \frac{\Pr(T_2 > 30)}{\Pr(T_2 > 20)} \\
 &= \frac{\frac{50-25}{50}}{\frac{50-15}{50}} \cdot \frac{\frac{50-30}{50}}{\frac{50-20}{50}} \\
 &= \mathbf{0.4762}
 \end{aligned}$$



Coach's Remarks

If you don't recognize from the hazard rate that each component's lifetime is uniform, you can determine the survival function from its relationship to the hazard rate.

$$\begin{aligned}
 S(x) &= \exp \left(- \int_{-\infty}^x h(t) dt \right) \\
 &= \exp \left(- \int_0^x \frac{1}{50-t} dt \right) \\
 &= \frac{50-x}{50}
 \end{aligned}$$

1.5.8 Miscellaneous

🕒 5m

Material covered in this subsection is not as important as material covered in the other Reliability Theory subsections. Skip it if you are pressed for time.

Recall the failure rate function and cumulative hazard function from Section 1.0.2.

$$h(x) = \frac{f(x)}{S(x)}$$

$$H(x) = \int_{-\infty}^x h(t) dt$$

Let's define a few terms related to the failure rate function.

- If $h(x)$ is an increasing function of x , then the distribution is an *increasing failure rate (IFR)* distribution.
- If $h(x)$ is a decreasing function of x , then the distribution is a *decreasing failure rate (DFR)* distribution.
- If $h(x)$ is constant, then the distribution is both an IFR and DFR distribution.
- If $\frac{H(x)}{x} = \frac{\int_0^x h(s) ds}{x}$ increases in x for $x \geq 0$, then the distribution is an *increasing failure rate on the average (IFRA)* distribution.
- If the distribution is IFR, then it is also IFRA.
- For a monotone system with independent components, if each component has an IFRA lifetime distribution, then the distribution of the system lifetime is also IFRA.

Weibull Distribution

A Weibull distribution with parameters τ and θ is:

- IFR if $\tau \geq 1$.

- DFR if $0 < \tau \leq 1$.
- both IFR and DFR if $\tau = 1$.

Gamma Distribution

A Gamma distribution with parameters α and θ is:

- IFR if $\alpha \geq 1$.
- DFR if $0 < \alpha \leq 1$.
- both IFR and DFR if $\alpha = 1$.

1.5 Summary

🕒 5m

- A parallel system will function as long as one of the components is functioning.
- A series system will function only when all components are functioning.
- A k -out-of- n system functions if and only if at least k out of the n components are functioning.

Minimal Path Sets and Minimal Cut Sets

- A minimal path set, A_i , is a minimal set of components whose functioning guarantees the functioning of the system.
- A minimal cut set, C_i , is a minimal set of components whose failure guarantees the failure of the system.
- For a k -out-of- n system, there are $\binom{n}{k}$ minimal path sets and $\binom{n}{n-k+1}$ minimal cut sets.

Structure Functions

The structure function of any system can be derived using minimal path sets or minimal cut sets.

	Structure Function, $\phi(\mathbf{x})$
Parallel	$1 - \prod_{i=1}^n (1 - x_i)$
Series	$\prod_{i=1}^n x_i$
Minimal Path Sets	$\max_j \prod_{i \in A_j} x_i$

Minimal Cut Sets	$\prod_{j=1}^m \max_{i \in C_j} x_i$
------------------	--------------------------------------

Reliability of Systems

The reliability of a system is a function of the reliability of its components.

$$\begin{aligned} r(\mathbf{p}) &= \Pr[\phi(\mathbf{X}) = 1] \\ &= \mathbb{E}[\phi(\mathbf{X})] \end{aligned}$$

Bounds on Reliability Function

- Method of Inclusion and Exclusion
 - First two inclusion-exclusion bounds using minimal path sets:

$$\sum_{j=1}^s \left(\prod_{i \in A_j} p_i \right) - \sum_{j=1}^s \sum_{k>j} \left(\prod_{i \in A_j \cup A_k} p_i \right) \leq r(\mathbf{p}) \leq \sum_{j=1}^s \left(\prod_{i \in A_j} p_i \right)$$

- First two inclusion-exclusion bounds using minimal cut sets:

$$\sum_{j=1}^m \left(\prod_{i \in C_j} (1 - p_i) \right) - \sum_{j=1}^m \sum_{k>j} \left(\prod_{i \in C_j \cup C_k} (1 - p_i) \right) \leq 1 - r(\mathbf{p}) \leq \sum_{j=1}^m \left(\prod_{i \in C_j} (1 - p_i) \right)$$

- Method of Intersection

$$\prod_{j=1}^m \left[1 - \prod_{i \in C_j} (1 - p_i) \right] \leq r(\mathbf{p}) \leq 1 - \prod_{j=1}^s \left[1 - \prod_{i \in A_j} p_i \right]$$

Random Graphs

- A random graph with n nodes has n^{n-2} minimal path sets and $2^{n-1} - 1$ minimal cut sets.
- Probability that a random graph with n nodes and with all $P_{i,j} = p$ is connected:

$$P_n = 1 - \sum_{k=1}^{n-1} \binom{n-1}{k-1} q^{k(n-k)} P_k, \quad n = 2, 3, \dots$$

- Bounds for P_n :

$$nq^{n-1} - \binom{n}{2} q^{2n-3} \leq 1 - P_n \leq (n+1)q^{n-1}$$

Lifetime of Systems

- Probability that a system's lifetime exceeds t :

$$\Pr(T > t) = r[\mathbf{S}(t)]$$

- Expected value of a system's lifetime:

$$\mathbb{E}[T] = \int_0^\infty r[\mathbf{S}(t)] dt$$

1.6.0 Overview

 5m

In this subsection, we will look at discrete Markov chains, as well as some of their properties and applications. Markov chains are used when there is a process that has a specific value in each time period. For example, a Markov chain could be used when considering the presence of rain for each day or the population growth with each generation. In general, a Markov chain is a process where the value at time $m + 1$ will depend only on the value at time m .

1.6.1 Introduction

🕒 10m

As mentioned, a Markov chain is used when there is a process that has a specific value in each time period, like the stock price at the end of each trading day. Each possible value for the process is referred to as a **state** of the Markov chain. The states of a given Markov chain will typically be denoted using integer values. Then, the process can only take on one value (i.e. be in one state) in each time period. But we are often interested in the probability of transitioning from one state to another. This is sometimes referred to as **entering** a state j from a state i (even if $i = j$).

Before getting into the details of this, let's formally define a Markov chain. Let $\{X_m, m = 0, 1, 2, \dots\}$ be a stochastic process that can take on a countable (finite or infinite) set of possible values, like non-negative integers. Recall that a stochastic process is a collection (or set) of random variables. Then, $X_m = i$ indicates that the process is in state i at time m .

In addition, assume that whenever the process is in state i , there is a **fixed** probability of transitioning to state j . This results in a **homogeneous** Markov chain, which can also be referred to as a **stationary** Markov chain. If the probabilities are not fixed, the Markov chain is **non-homogeneous**. We will denote the probability of transitioning from state i to state j as $P_{i,j}$. More specifically, $P_{i,j}$ is the conditional probability of any future state X_{m+1} given its present and past states.

A **Markov chain** is a stochastic process where the conditional distribution of any future state X_{m+1} given its present and past states is dependent only on the present state X_m . In other words, any future state is independent of past states. This is the memoryless property of a stochastic process, or the **Markov property**. Then, the conditional distribution is:

$$P_{i,j} = \Pr \left(\underbrace{X_{m+1} = j}_{\text{future state}} \middle| \underbrace{X_m = i}_{\text{present state}} \right)$$

We can represent all transition probabilities in a matrix. Let \mathbf{P} be the one-step transition probability matrix for a discrete Markov chain. Then, each entry $P_{i,j}$ in the matrix represents the probability of being in state i now and transitioning to state j in the next period.

$$\mathbf{P} = \begin{bmatrix} P_{1,1} & P_{1,2} & P_{1,3} & \cdots \\ P_{2,1} & P_{2,2} & P_{2,3} & \cdots \\ \vdots & \vdots & \vdots & \\ P_{i,1} & P_{i,2} & P_{i,3} & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix}$$

Note that the sum of the transition probabilities for each row i must equal 1.

To better understand the concept of Markov chains, consider the following example.

Assume that the weather for a given day can be sunny, cloudy, or rainy, and assume that tomorrow's weather depends only on today's weather. Specifically,

- If it is sunny today, then the probability that it will be sunny tomorrow is 0.4, and the probability that it will be cloudy tomorrow is 0.5.
- If it is cloudy today, then the probability that it will be rainy tomorrow is 0.7, and the probability that it will be cloudy tomorrow is 0.2.
- If it is rainy today, then the probability that it will be sunny tomorrow is 0.6, and the probability that it will be rainy tomorrow is 0.4.

Determine the one-step transition probability matrix for this Markov chain.

We can model this with a three-state Markov chain with states 1 (sunny), 2 (cloudy), and 3 (rainy).

$$\mathbf{P} = \begin{bmatrix} P_{1,1} & P_{1,2} & P_{1,3} \\ P_{2,1} & P_{2,2} & P_{2,3} \\ P_{3,1} & P_{3,2} & P_{3,3} \end{bmatrix}$$

$P_{1,1}$ is the probability that it is sunny tomorrow given that it is sunny today, so $P_{1,1} = 0.4$. And $P_{1,2}$ is the probability that it is cloudy tomorrow given that it is sunny today, so $P_{1,2} = 0.5$. Then, recall that we need the probabilities for a given starting state to sum to 1. So, we can deduce that the probability that it is rainy tomorrow given that it is sunny today is $P_{1,3} = 1 - 0.4 - 0.5 = 0.1$.

Apply similar logic to determine the remaining transition probabilities. So, the one-step transition probability matrix for this Markov chain is

$$\mathbf{P} = \begin{bmatrix} 0.4 & 0.5 & 0.1 \\ 0.1 & 0.2 & 0.7 \\ 0.6 & 0.0 & 0.4 \end{bmatrix}$$

Coach's Remarks

In the textbook by Ross, examples will often begin with state 0, rather than state 1. Exam questions will sometimes define the starting state, but when they do not, the choice is irrelevant.

1.6.2 Multiple-Step Transition Probabilities

🕒 50m

Suppose we are interested in the weather for the next few days given today's weather. If it is sunny today, what is the probability that it will be rainy two days from today? If it is rainy today, what is the probability that it will be cloudy five days from today?

In the previous section, we defined $P_{i,j}$ as the probability of moving from state i to state j with one transition. This is the one-step transition probability from state i to state j .

Now, we are interested in the n -step transition probability, which is the probability of moving from state i to state j after n transitions. To find this probability, we need to consider all possible transition sequences from state i to state j within n periods. The n -step transition probability is denoted as $P_{i,j}^n$. Note that $P_{i,j}^n$ does not equal $P_{i,j}$ to the n^{th} power.

Consider the previous example.

Assume that the weather for a given day can be sunny, cloudy, or rainy, and assume that tomorrow's weather depends only on today's weather. Specifically,

- If it is sunny today, then the probability that it will be sunny tomorrow is 0.4, and the probability that it will be cloudy tomorrow is 0.5.
- If it is cloudy today, then the probability that it will be rainy tomorrow is 0.7, and the probability that it will be cloudy tomorrow is 0.2.
- If it is rainy today, then the probability that it will be sunny tomorrow is 0.6, and the probability that it will be rainy tomorrow is 0.4.

Calculate the probability that it is rainy two days from now given that it is sunny today.

We can list all possible transition sequences with two transitions from state 1 to state 3.

- 1 → 1 → 3

- . - . -

- $1 \rightarrow 2 \rightarrow 3$
- $1 \rightarrow 3 \rightarrow 3$

Then, calculate the transition probabilities for all three possible scenarios.

- $P_{1,1}P_{1,3} = 0.4(0.1) = 0.04$
- $P_{1,2}P_{2,3} = 0.5(0.7) = 0.35$
- $P_{1,3}P_{3,3} = 0.1(0.4) = 0.04$

Thus, the probability that it will be rainy two days from today given it is sunny today is $P_{1,3}^2 = 0.04 + 0.35 + 0.04 = \mathbf{0.43}$.

What we just did was apply the *Chapman-Kolmogorov equations*. These equations break down the n -step transition probabilities and add them up to calculate $P_{i,j}^n$. The Chapman-Kolmogorov equations are:

$$P_{i,j}^{n+m} = \sum_{k=1}^{\infty} P_{i,k}^n P_{k,j}^m \quad (1.6.2.1)$$

We can use this approach to calculate the conditional distribution of any future state after multiple transitions. However, as the number of additional transitions increases, the number of possible transition sequences will increase tremendously. A more systematic approach to calculating multiple-transition probabilities involves the use of matrices.

Let $\mathbf{P}^{(n)}$ be the n -step transition matrix. Then, from the Chapman-Kolmogorov equations, we have:

$$\mathbf{P}^{(n+m)} = \mathbf{P}^{(n)} \cdot \mathbf{P}^{(m)} \quad (1.6.2.2)$$

The n -step transition matrix is calculated by multiplying the one-step transition probability matrix by itself $n - 1$ times. Note that the dot in Equation 1.6.2.2 represents matrix multiplication.

Matrix Multiplication Review

Let \mathbf{A} be an $m \times n$ matrix and \mathbf{B} be a $p \times q$ matrix.

Matrix multiplication in the form of \mathbf{AB} is only possible if $n = p$. The resulting matrix \mathbf{AB} will be an $m \times q$ matrix. Note that \mathbf{AB} and \mathbf{BA} will not be equal unless $\mathbf{A} = \mathbf{B}$.

To calculate the "row i , column j " entry of the matrix product, \mathbf{AB} , multiply the entries in row i of matrix \mathbf{A} by the corresponding entries in column j of matrix \mathbf{B} , then add them together.

As an example, consider the following 2×2 matrices.

$$\mathbf{A} = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{bmatrix}$$

The product of these two matrices is:

$$\mathbf{AB} = \begin{bmatrix} a_{1,1}b_{1,1} + a_{1,2}b_{2,1} & a_{1,1}b_{1,2} + a_{1,2}b_{2,2} \\ a_{2,1}b_{1,1} + a_{2,2}b_{2,1} & a_{2,1}b_{1,2} + a_{2,2}b_{2,2} \end{bmatrix}$$

Let's calculate the two-step transition probability matrix for our three-state Markov chain for weather.

$$\begin{aligned} \mathbf{P}^{(2)} &= \mathbf{P}^{(1)} \cdot \mathbf{P}^{(1)} \\ &= \mathbf{P} \cdot \mathbf{P} \\ &= \begin{bmatrix} 0.4 & 0.5 & 0.1 \\ 0.1 & 0.2 & 0.7 \\ 0.6 & 0.0 & 0.4 \end{bmatrix} \cdot \begin{bmatrix} 0.4 & 0.5 & 0.1 \\ 0.1 & 0.2 & 0.7 \\ 0.6 & 0.0 & 0.4 \end{bmatrix} \\ &= \begin{bmatrix} 0.27 & 0.30 & 0.43 \\ 0.10 & 0.00 & 0.19 \end{bmatrix} \end{aligned}$$

$$= \begin{bmatrix} 0.40 & 0.09 & 0.43 \\ 0.48 & 0.30 & 0.22 \end{bmatrix}$$

The two-step transition probability from state 1 (sunny) to state 3 (rainy) is 0.43, as calculated earlier. Note that the sum of the transition probabilities for each row is 1. Use this property to check if there are any miscalculations.

Example 1.6.2.1

You are given the following four-state (1, 2, 3, 4) transition matrix:

$$\mathbf{P} = \begin{bmatrix} 0.5 & 0.5 & 0.0 & 0.0 \\ 0.0 & 0.8 & 0.2 & 0.0 \\ 0.6 & 0.3 & 0.0 & 0.1 \\ 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}$$

Calculate the four-step transition probability from state 3 to state 4.

Solution

To find the four-step transition matrix, begin by calculating the two-step transition matrix.

$$\begin{aligned} \mathbf{P}^{(2)} &= \mathbf{P} \cdot \mathbf{P} \\ &= \begin{bmatrix} 0.25 & 0.65 & 0.10 & 0.00 \\ 0.12 & 0.70 & 0.16 & 0.02 \\ 0.30 & 0.54 & 0.06 & 0.10 \\ 0.00 & 0.00 & 0.00 & 1.00 \end{bmatrix} \end{aligned}$$

Calculate the four-step transition matrix by multiplying the two-step transition matrix by itself.

$$\begin{aligned}\mathbf{P}^{(4)} &= \mathbf{P}^{(2)} \cdot \mathbf{P}^{(2)} \\ &= \begin{bmatrix} 0.1705 & 0.6715 & 0.1350 & 0.0230 \\ 0.1620 & 0.6544 & 0.1336 & 0.0500 \\ 0.1578 & 0.6054 & 0.1200 & 0.1168 \\ 0.0000 & 0.0000 & 0.0000 & 1.0000 \end{bmatrix}\end{aligned}$$

Therefore, the four-step transition probability from state 3 to state 4 is

$$P_{3,4}^4 = \mathbf{0.1168}$$



Coach's Remarks

Note that it is not necessary to calculate the entire n -step transition matrix in order to get $P_{i,j}^n$. A shortcut for finding $P_{i,j}^n$ is to multiply together the i^{th} row of \mathbf{P} , $\mathbf{P}^{(n-2)}$, and the j^{th} column of \mathbf{P} .

So, for this example, we can calculate the four-step transition probability from state 3 to state 4 as

$$\begin{aligned}P_{3,4}^4 &= [0.6 \quad 0.3 \quad 0.0 \quad 0.1] \begin{bmatrix} 0.25 & 0.65 & 0.10 & 0.00 \\ 0.12 & 0.70 & 0.16 & 0.02 \\ 0.30 & 0.54 & 0.06 & 0.10 \\ 0.00 & 0.00 & 0.00 & 1.00 \end{bmatrix} \begin{bmatrix} 0.0 \\ 0.0 \\ 0.1 \\ 1.0 \end{bmatrix} \\ &= [0.186 \quad 0.600 \quad 0.108 \quad 0.106] \begin{bmatrix} 0.0 \\ 0.0 \\ 0.1 \\ 1.0 \end{bmatrix}\end{aligned}$$

$$\begin{bmatrix} 0.1 \\ 1.0 \end{bmatrix} = 0.1168$$

Example 1.6.2.2

An auto insurance company models their policyholders using a Markov chain with the following three states:

- 1: Preferred
- 2: Standard
- 3: Risky

Policyholders are classified into one of the three states depending on the number of claims filed in the previous year. All new policyholders are classified as standard for the first policy year. The transitions occur at the beginning of every year.

- If a preferred policyholder files at least one claim this year, they will be classified as standard next year. Otherwise, they remain as preferred.
- If a standard policyholder files no claims this year, they will be classified as preferred next year. If they file more than one claim, they are classified as risky. Otherwise, they remain as standard.
- If a risky policyholder files no claims this year, they are classified as standard. Otherwise, they remain as risky.

Each policyholder, regardless of state, files claims at a Poisson rate of 1 per year. What is the probability that a new policyholder will be classified as risky in the fourth policy year? Assume the new policyholder's inception date is at the beginning of the

year.

Solution

Before we calculate the desired probability, we need to derive the transition probability matrix.

$$\mathbf{P} = \begin{bmatrix} P_{1,1} & P_{1,2} & P_{1,3} \\ P_{2,1} & P_{2,2} & P_{2,3} \\ P_{3,1} & P_{3,2} & P_{3,3} \end{bmatrix}$$

The transition probability from state 1 to state 1 is the probability of filing no claims in one year. The annual number of claims N is a Poisson random variable with mean 1. Therefore,

$$\begin{aligned} P_{1,1} &= \Pr(N = 0) \\ &= e^{-1} \end{aligned}$$

A preferred policyholder will transition to standard if at least one claim is filed. Therefore,

$$\begin{aligned} P_{1,2} &= \Pr(N \geq 1) \\ &= 1 - e^{-1} \end{aligned}$$

The transition probability from state 2 to state 1 is the probability of filing no claims. This means $P_{2,1} = e^{-1}$. The transition probability from state 2 to state 2 is the probability of filing one claim, $P_{2,2} = e^{-1}$. The transition probability from state 2 to 3 is the probability of filing more than one claim.

$$\begin{aligned} P_{2,3} &= \Pr(N > 1) \\ &= 1 - \Pr(N = 0) - \Pr(N = 1) \\ &= 1 - e^{-1} - e^{-1} \end{aligned}$$

$$\begin{array}{cccc}
 & \downarrow & \downarrow & \downarrow \\
 = 1 - 2e^{-1}
 \end{array}$$

The transition probability from state 3 to state 2 is the probability of filing no claims. The transition probability from state 3 to state 3 is the probability of filing at least one claim.

Therefore, we have the following transition probability matrix:

$$\mathbf{P} = \begin{bmatrix} e^{-1} & 1 - e^{-1} & 0 \\ e^{-1} & e^{-1} & 1 - 2e^{-1} \\ 0 & e^{-1} & 1 - e^{-1} \end{bmatrix}$$

Next, calculate the probability that a new policyholder will be risky in their fourth policy year. Recall that new policyholders start as standard policyholders, i.e. they start in state 2. There are three transitions between the first and fourth year. Thus, we need to calculate the three-step transition probability matrix.

The two-step transition matrix is:

$$\mathbf{P}^{(2)} = \mathbf{P} \cdot \mathbf{P} = \begin{bmatrix} 0.3679 & 0.4651 & 0.1670 \\ 0.2707 & 0.4651 & 0.2642 \\ 0.1353 & 0.3679 & 0.4968 \end{bmatrix}$$

The three-step transition matrix is:

$$\mathbf{P}^{(3)} = \mathbf{P}^{(2)} \cdot \mathbf{P} = \begin{bmatrix} 0.3064 & 0.4651 & 0.2285 \\ 0.2707 & 0.4394 & 0.2899 \\ 0.1851 & 0.4036 & 0.4112 \end{bmatrix}$$

Thus, the probability that a new policyholder will be risky in the fourth policy year is

$$P_{2,3}^3 = \mathbf{0.2899}$$

Coach's Remarks

Pay attention to the wording of the question. The " n^{th} year" does not necessarily mean that there are n transitions.

For instance, in this example, there are only three transitions from the first to the fourth policy year classification. Thus, it would be incorrect to calculate the four-step transition probability.

Probabilities Involving Absorbing States

Suppose we are interested in the probability that the Markov chain, starting in state i , enters state j at time m without ever being in any states in some set \mathcal{A} . To calculate this probability, we define another Markov chain that contains all states **not** in set \mathcal{A} and an absorbing state to represent all states in set \mathcal{A} . An *absorbing* state is a state that cannot be left once it is entered.

All states in set \mathcal{A} can be combined and treated as one absorbing state because we are only interested in transition probabilities that exclude entering states in set \mathcal{A} . Note that the states in set \mathcal{A} are not necessarily absorbing states, but the collection of states in set \mathcal{A} as a whole is treated as one absorbing state.

Let A represent the absorbing state, and $Q_{i,j}$ represent the transition probability from state i to state j for the new Markov chain. Construct the probability transition matrix \mathbf{Q} from \mathbf{P} using the following rules:

$$\begin{aligned}
 Q_{i,j} &= P_{i,j}, && \text{if } i \notin \mathcal{A}, \quad j \notin \mathcal{A} \\
 Q_{i,A} &= \sum_{j \in \mathcal{A}} P_{i,j}, && \text{if } i \notin \mathcal{A} \\
 Q_{A,i} &= 0, && \text{if } i \notin \mathcal{A} \\
 Q_{A,A} &= 1
 \end{aligned} \tag{1.6.2.3}$$

Note that $Q_{i,A}$ represents the one-step transition probability from state i to state A . It is the sum of the probabilities of entering any of the states that are in set \mathcal{A} from state i . Note that each row in \mathbf{Q} must also sum to 1.

To better understand these formulas, let's explore an example.

Example 1.6.2.3

You are given the following transition matrix for a four-state (1, 2, 3, 4) Markov chain:

$$\mathbf{P} = \begin{bmatrix} 0.5 & 0.3 & 0.2 & 0.0 \\ 0.0 & 0.7 & 0.2 & 0.1 \\ 0.6 & 0.2 & 0.0 & 0.2 \\ 0.8 & 0.1 & 0.1 & 0.0 \end{bmatrix}$$

The Markov chain is in state 1 at time 0.

Calculate the probability of entering state 2 at time 4, without ever entering states 3 and 4.

Solution

States 3 and 4 need to be collapsed into an absorbing state, so $\mathcal{A} = \{3, 4\}$. Using the rules specified by Equation 1.6.2.3, we have the following matrix \mathbf{Q} .

$$\begin{aligned}
 \mathbf{Q} &= \begin{bmatrix} Q_{1,1} & Q_{1,2} & Q_{1,A} \\ Q_{2,1} & Q_{2,2} & Q_{2,A} \\ Q_{A,1} & Q_{A,2} & Q_{A,A} \end{bmatrix} \\
 &= \begin{bmatrix} P_{1,1} & P_{1,2} & P_{1,3} + P_{1,4} \\ P_{2,1} & P_{2,2} & P_{2,3} + P_{2,4} \\ 0.0 & 0.0 & 1.0 \end{bmatrix} \\
 &= \begin{bmatrix} 0.5 & 0.3 & 0.2 \\ 0.0 & 0.7 & 0.3 \\ 0.0 & 0.0 & 1.0 \end{bmatrix}
 \end{aligned}$$

Our goal is to calculate the transition probability from state 1 to state 2 after four transitions, i.e. $Q_{1,2}^4$. This is the $Q_{1,2}$ entry in the four-step transition probability matrix for \mathbf{Q} .

First, compute the two-step transition matrix, $\mathbf{Q}^{(2)} = \mathbf{Q} \cdot \mathbf{Q}$. Then, find the four-step transition matrix, $\mathbf{Q}^{(4)} = \mathbf{Q}^{(2)} \cdot \mathbf{Q}^{(2)}$.

$$\mathbf{Q}^{(4)} = \begin{bmatrix} 0.0625 & 0.2664 & 0.6711 \\ 0.0000 & 0.2401 & 0.7599 \\ 0.0000 & 0.0000 & 1.0000 \end{bmatrix}$$

The desired probability is:

$$Q_{1,2}^4 = \mathbf{0.2664}$$



Alternative Solution

An alternative approach is to use logic and basic probability. We want the four-step transition probability from state 1 to state 2 without entering states 3 and 4.

State 1 can access state 2, but state 2 cannot access state 1, as $P_{2,1} = 0$. This means that once state 1 enters state 2, the chain has to stay in state 2 without transitioning to states 3 and 4. Thus, the possible transition sequences are as follows:

- $1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 2$
- $1 \rightarrow 1 \rightarrow 1 \rightarrow 2 \rightarrow 2$
- $1 \rightarrow 1 \rightarrow 2 \rightarrow 2 \rightarrow 2$
- $1 \rightarrow 2 \rightarrow 2 \rightarrow 2 \rightarrow 2$

Calculate the probability for each transition sequence.

- $0.5 (0.5) (0.5) (0.3) = 0.0375$
- $0.5 (0.5) (0.3) (0.7) = 0.0525$
- $0.5 (0.3) (0.7) (0.7) = 0.0735$
- $0.3 (0.7) (0.7) (0.7) = 0.1029$

The final answer is the sum of these probabilities.

$$0.0375 + 0.0525 + 0.0735 + 0.1029 = \mathbf{0.2664}$$



Coach's Remarks

As demonstrated in the example above, instead of calculating the probabilities using the matrix \mathbf{Q} , we can calculate and sum the probabilities of all possible transition sequences.

This approach will be easier if some of the one-step transition probabilities are zero or if the number of possible transition sequences is small.

So far, we have assumed the starting state i and the ending state j are not in set \mathcal{A} . Now, suppose we are interested in the probability that the chain starts in state i and enters state j at time m , where the starting and ending states can either be in set \mathcal{A} or not, as long as the in-between states are not in \mathcal{A} .

To calculate this probability, we will use entries from both \mathbf{P} and \mathbf{Q} , depending on the starting and ending states. The matrix \mathbf{P} will only be needed for transferring out of a state in set \mathcal{A} in the first step or transferring into a state in set \mathcal{A} in the last step. Otherwise, matrix \mathbf{Q} will be used to calculate transition probabilities.

The table below summarizes the desired transition probability for each situation.

State i	State j	Desired Probability
$i \notin \mathcal{A}$	$j \notin \mathcal{A}$	$Q_{i,j}^m$
$i \notin \mathcal{A}$	$j \in \mathcal{A}$	$\sum_{r \notin \mathcal{A}} Q_{i,r}^{m-1} P_{r,j}$
$i \in \mathcal{A}$	$j \notin \mathcal{A}$	$\sum_{r \notin \mathcal{A}} P_{i,r} Q_{r,j}^{m-1}$
$i \in \mathcal{A}$	$j \in \mathcal{A}$	$\sum_{r \notin \mathcal{A}} \sum_{k \notin \mathcal{A}} P_{i,r} Q_{r,k}^{m-2} P_{k,j}$

Note that the term $Q_{i,r}^{m-1}$ represents the $(m - 1)$ -step transition probabilities without entering any state in set \mathcal{A} , and that $P_{r,j}$ represents the one-step transition probabilities into state j .

For example, suppose state j is in set \mathcal{A} , and we are interested in the probability that the chain starts in state i and enters state j after m transitions without entering any states in set \mathcal{A} before that. First, use the matrix \mathbf{Q} to calculate the probabilities of transitioning into a state $r \notin \mathcal{A}$ on the $(m - 1)^{\text{st}}$ transition. Then, sum the transition probabilities from each possible state r into state j using the Chapman-Kolmogorov equations.

To apply this methodology, follow these steps:

1. Determine if the starting and ending states, i and j , are in set \mathcal{A} .
2. Determine the correct formula to use based on the table above.
3. Depending on the formula, find the m -, $(m - 1)$ -, or $(m - 2)$ -step transition probabilities.
4. Calculate the desired probability.

Example 1.6.2.4

You are given the following transition matrix for a four-state (1, 2, 3, 4) Markov chain:

$$\mathbf{P} = \begin{bmatrix} 0.2 & 0.2 & 0.5 & 0.1 \\ 0.1 & 0.6 & 0.2 & 0.1 \\ 0.1 & 0.7 & 0.1 & 0.1 \\ 0.2 & 0.1 & 0.3 & 0.4 \end{bmatrix}$$

Calculate the three-step transition probability from state 4 to state 3 without entering state 2 and without re-entering state 4.

Solution

States 2 and 4 need to be collapsed into an absorbing state, so $\mathcal{A} = \{2, 4\}$. Using the rules specified by Equation 1.6.2.3, we have the following matrix \mathbf{Q} .

$$\begin{aligned}\mathbf{Q} &= \begin{bmatrix} Q_{1,1} & Q_{1,3} & Q_{1,A} \\ Q_{3,1} & Q_{3,3} & Q_{3,A} \\ Q_{A,1} & Q_{A,3} & Q_{A,A} \end{bmatrix} \\ &= \begin{bmatrix} 0.2 & 0.5 & 0.3 \\ 0.1 & 0.1 & 0.8 \\ 0.0 & 0.0 & 1.0 \end{bmatrix}\end{aligned}$$

State 4 is in \mathcal{A} , while state 3 is not. Thus, the transition probability is:

$$\sum_{r \notin \mathcal{A}} P_{i,r} Q_{r,j}^{m-1} = P_{4,1} Q_{1,3}^2 + P_{4,3} Q_{3,3}^2$$

To find $Q_{1,3}^2$ and $Q_{3,3}^2$, compute the two-step transition probability matrix for \mathbf{Q} .

$$\begin{aligned}\mathbf{Q}^{(2)} &= \begin{bmatrix} Q_{1,1}^2 & Q_{1,3}^2 & Q_{1,A}^2 \\ Q_{3,1}^2 & Q_{3,3}^2 & Q_{3,A}^2 \\ Q_{A,1}^2 & Q_{A,3}^2 & Q_{A,A}^2 \end{bmatrix} \\ &= \begin{bmatrix} 0.09 & 0.15 & 0.76 \\ 0.03 & 0.06 & 0.91 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}\end{aligned}$$

Therefore, the desired probability is:

$$\begin{aligned}P_{4,1} Q_{1,3}^2 + P_{4,3} Q_{3,3}^2 &= 0.2(0.15) + 0.3(0.06) \\ &= \mathbf{0.048}\end{aligned}$$

Note that because states 2 and 4 have been combined into a single

Note that because state 2 and state 4 have been combined into a single absorbing state, $Q_{1,3}^2$ and $Q_{3,3}^2$ are not the "1st row, 3rd column" and "3rd row, 3rd column" entries of the matrix $\mathbf{Q}^{(2)}$. They are the two-step transition probabilities from state 1 to state 3 and from state 3 to state 3, respectively.



Alternative Solution

Once again, we can use logic and basic probability as an alternative approach. We want the three-step transition probability from state 4 to state 3 without entering state 2 or re-entering state 4. Thus, the possible transition sequences are as follows:

- $4 \rightarrow 1 \rightarrow 1 \rightarrow 3$
- $4 \rightarrow 1 \rightarrow 3 \rightarrow 3$
- $4 \rightarrow 3 \rightarrow 1 \rightarrow 3$
- $4 \rightarrow 3 \rightarrow 3 \rightarrow 3$

Calculate the probabilities of each transition sequence and add them up.

- $0.2(0.2)(0.5) = 0.020$
- $0.2(0.5)(0.1) = 0.010$
- $0.3(0.1)(0.5) = 0.015$
- $0.3(0.1)(0.1) = 0.003$

$$0.020 + 0.010 + 0.015 + 0.003 = \mathbf{0.048}$$

Unconditional Probabilities

The transition probabilities that we have been working with so far are conditional on the initial states. Specifically, $P_{i,j}^n$ is the probability that the process is in state j at time n given the initial state is i . To calculate the unconditional probability of being in state j at time n , we need the distribution of the initial state.

Let α_i be the probability of being in state i at time 0, $\Pr(X_0 = i)$. Then, the unconditional probability of being in state j at time n is:

$$\Pr(X_n = j) = \sum_{i=1}^{\infty} \alpha_i P_{i,j}^n$$

Recall the three-state weather Markov chain with states 1 (sunny), 2 (cloudy), and 3 (rainy).

$$\mathbf{P} = \begin{bmatrix} 0.4 & 0.5 & 0.1 \\ 0.1 & 0.2 & 0.7 \\ 0.6 & 0.0 & 0.4 \end{bmatrix}$$

Suppose the probabilities that it is sunny, cloudy, and rainy on the first day are 0.3, 0.5, and 0.2, respectively. What is the probability that it is cloudy on the third day?

There are two transitions from the first day to the third day. So, we need the two-step transition probability matrix, which we calculated previously.

$$\mathbf{P}^{(2)} = \begin{bmatrix} 0.27 & 0.30 & 0.43 \\ 0.48 & 0.09 & 0.43 \\ 0.48 & 0.30 & 0.22 \end{bmatrix}$$

Thus, the unconditional probability that it is cloudy on the third day is:

$$0.3(0.30) + 0.5(0.09) + 0.2(0.30) = \mathbf{0.195}$$

1.6.3 Classification of States

🕒 15m

In this subsection, we will discuss the classifications and properties of states.

A state j is **accessible** from state i if the following is satisfied:

$$P_{i,j}^n > 0, \quad \text{for some } n \geq 0$$

This means there is at least one point in time (including the current time) where going from state i to state j is possible. This relationship is written as $i \rightarrow j$. If two states are accessible to each other, then the two states **communicate**. This relationship is written as $i \leftrightarrow j$. Note that every state communicates with itself because every state is accessible to itself. By definition, if $j = i$, $n = 0$, and m is the current time:

$$P_{i,i}^0 = \Pr(X_m = i \mid X_m = i) = 1 > 0$$

There are three properties of communication:

- State i communicates with itself.
- If state i communicates with state j , then state j communicates with state i .
- If state i communicates with state j and state j communicates with state k , then state i communicates with state k .

Two states that communicate are in the same **class**. All states in a class communicate with one another. A **class property** is a property that applies to all states in the class.

A Markov chain with only one class is an **irreducible** Markov chain. This means all states in an irreducible Markov chain communicate with one another.

A **finite** Markov chain is a Markov chain with a finite number of states. An **infinite** Markov chain is a Markov chain with an infinite number of states.

Let f_i be the probability of re-entering state i at any point in the future, given that the process starts in state i .

A state is **recurrent** if the probability of ever re-entering itself is 1, i.e. if $f_i = 1$. This

means that it is always possible to re-enter state i from any state in the future. If state i is recurrent, and state i communicates with state j , then state j is also recurrent.

If a process starts in a recurrent state i , the process will re-enter state i infinitely often. In this case, the expected number of time periods the process is in state i is infinite. In general, we can determine that a state i is recurrent if $\sum_{n=1}^{\infty} P_{i,i}^n = \infty$.

A state is *transient* if the probability of ever re-entering itself is less than 1, i.e. if $f_i < 1$. A transient state is a state where there is a non-zero probability of not re-entering the state ever.

If a process starts in a transient state i , there is a probability of $1 - f_i$ that the process will never re-enter state i . This means the probability that the process is in state i exactly n times, given it starts in state i , equals $f_i^{n-1} (1 - f_i)$, for $n \geq 1$. So, for a process starting in transient state i , the number of times the process is in state i has a geometric distribution with mean $\frac{1}{1-f_i}$. We can determine that a state i is transient if $\sum_{n=1}^{\infty} P_{i,i}^n < \infty$.

Coach's Remarks

In order to align with the parameterization provided in the MAS-I exam tables for the geometric distribution, we can rephrase the distribution representing the number of times the process is in state i . Given that a process starts in a transient state i , the number of times the process re-enters state i , for $n \geq 0$, has a geometric distribution with $\beta = \frac{f_i}{1-f_i}$.

Because all states in a class are communicating, if one of the states is recurrent, then all of them are recurrent. If one of the states is transient, then all of them are transient. Thus, recurrence and transience are both class properties.

In a finite Markov chain, there must be at least one recurrent state. Since an irreducible Markov chain has only one class, this means that all states in a finite irreducible Markov chain are recurrent.

To apply these concepts, let's look at two Markov chains with the following transition

matrices.

$$\mathbf{A} = \begin{bmatrix} 0.3 & 0.7 & 0.0 & 0.0 \\ 0.8 & 0.1 & 0.1 & 0.0 \\ 0.0 & 0.5 & 0.0 & 0.5 \\ 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0.4 & 0.5 & 0.1 \\ 0.1 & 0.2 & 0.7 \\ 0.6 & 0.0 & 0.4 \end{bmatrix}$$

A is the transition probability matrix for a four-state (1, 2, 3, 4) Markov chain. State 4 is absorbing and recurrent. Once the process enters state 4, it remains in state 4. The probability of re-entering state 4 is 1 because no other states can be entered once the chain is in state 4.

State 1 can access state 2 directly and can access state 3 through state 2. State 2 can access states 1 or 3. State 3 can access state 2 directly and can access state 1 through state 2. This means that states 1, 2, and 3 communicate. In addition, they are also transient. This is because there is a possibility of never re-entering the states if there is a transition from state 3 to state 4, which is an absorbing state.

Note that this Markov chain has two classes, {1, 2, 3} and {4}.

B is the transition probability matrix for a three-state (1, 2, 3) Markov chain. States 1, 2, and 3 are communicating states; thus, they are in the same class. This means that this is an irreducible Markov chain. Since the chain is also finite, this means the states are all recurrent.

Example 1.6.3.1 (Adapted from CAS S F2015 9)

You are given the following Markov chain transition matrix:

$$\mathbf{P} = \begin{bmatrix} 0.5 & 0.5 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.7 & 0.3 & 0.0 & 0.0 \\ 0.0 & 0.8 & 0.2 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.3 & 0.5 & 0.2 \\ 0.0 & 0.0 & 0.0 & 0.9 & 0.1 \end{bmatrix}$$

Determine the number of recurrent states in this Markov chain.

Solution

Let the states be 1, 2, 3, 4, and 5.

State 1 is not accessible from any other states besides itself. This means there is a possibility of never re-entering the state, which occurs if the process leaves state 1. Therefore, state 1 is transient.

States 2 and 3 are communicating states. In addition, they are recurrent because it is always possible to re-enter these states, given that the process starts there. This is because when starting in either state 2 or 3, the process can only transition to states 2 and 3.

States 4 and 5 are communicating states as well. However, they are transient because there is a possibility of never re-entering these states. This occurs if the chain enters state 3 from state 4 (or from state 5 via state 4), as state 4 and state 5 can never be re-entered from state 3.

Therefore, there are **two recurrent states**, states 2 and 3, in this Markov chain.



1.6.4 Long-Run Proportions and Limiting Probabilities

🕒 20m

Before discussing long-run proportions, let's expand on the concept of recurrence. For this subsection, let state j be a recurrent state.

Recurrence can be classified into two types: positive recurrence and null recurrence. Let m_j be the expected number of transitions for a Markov chain to return to state j given the chain starts in state j . A state is *positive recurrent* if $m_j < \infty$, and a state is *null recurrent* if $m_j = \infty$.

Recall that recurrence is a class property. This means that if a state in a class is positive recurrent, then all other states in the same class are positive recurrent as well. The same applies to null recurrence.

Now, let π_j be the *long-run proportion* of time that an irreducible Markov chain is in state j . This is also known as the *stationary probability* because if the initial state is selected according to the probabilities π_j , then the probability of being in state j at any time n is also equal to π_j . The long-run proportion for a state j is the reciprocal of the expected number of transitions it takes to return to state j given the chain starts in that state.

$$\pi_j = \frac{1}{m_j} \quad (1.6.4.1)$$

To determine the long-run proportions, first note that $\pi_i P_{i,j}$ is the long-run proportion of transitions from state i to state j . In other words, of all the possible transitions over time, the proportion of transitions that are from state i to state j equals $\pi_i P_{i,j}$. Then, for an irreducible and positive recurrent Markov chain, the long-run proportions are solved using the following two equations.

$$\begin{aligned} \pi_j &= \sum_{i=1}^{\infty} \pi_i P_{i,j} \\ \sum_{j=1}^{\infty} \pi_j &= 1 \end{aligned} \quad (1.6.4.2)$$

If no unique solution exists for the equations above, the Markov chain is either transient or null recurrent, i.e. not positive recurrent, in which case all $\pi_i = 0$. For an irreducible

finite-state Markov chain, all states will be positive recurrent.

Recall the three-state weather Markov chain with states 1 (sunny), 2 (cloudy), and 3 (rainy).

$$\mathbf{P} = \begin{bmatrix} 0.4 & 0.5 & 0.1 \\ 0.1 & 0.2 & 0.7 \\ 0.6 & 0.0 & 0.4 \end{bmatrix}$$

Calculate the long-run proportions for all states.

The Markov chain in this example is irreducible since all states communicate with each other, which means they are all in the same class. Since this chain is also finite, all states are positive recurrent.

Then, using Equation 1.6.4.2, we have the following system of equations:

$$\pi_1 = 0.4\pi_1 + 0.1\pi_2 + 0.6\pi_3 \quad (1)$$

$$\pi_2 = 0.5\pi_1 + 0.2\pi_2 \quad (2)$$

$$\pi_3 = 0.1\pi_1 + 0.7\pi_2 + 0.4\pi_3 \quad (3)$$

$$\pi_1 + \pi_2 + \pi_3 = 1 \quad (4)$$

Using (2), express π_2 in terms of π_1 .

$$\begin{aligned} 0.8\pi_2 &= 0.5\pi_1 \\ \pi_2 &= \frac{5}{8}\pi_1 \end{aligned} \quad (5)$$

Using (3), express π_3 in terms of π_1 and π_2 , and substitute (5) into the equation.

$$\begin{aligned}
 0.6\pi_3 &= 0.1\pi_1 + 0.7\pi_2 \\
 \pi_3 &= \frac{1}{6}\pi_1 + \frac{7}{6}\pi_2 \\
 &= \frac{1}{6}\pi_1 + \frac{7}{6}\left(\frac{5}{8}\pi_1\right) \\
 &= \frac{43}{48}\pi_1
 \end{aligned} \tag{6}$$

Substitute (5) and (6) into (4).

$$\begin{aligned}
 \pi_1 + \frac{5}{8}\pi_1 + \frac{43}{48}\pi_1 &= 1 \\
 \pi_1 &= \mathbf{0.3967}
 \end{aligned} \tag{7}$$

Substitute (7) into (5) and (6) to solve for the other long-run proportions.

$$\begin{aligned}
 \pi_2 &= \mathbf{0.2479} \\
 \pi_3 &= \mathbf{0.3554}
 \end{aligned}$$

So, in the long run, it is sunny 40% of the time, cloudy 25% of the time, and rainy 35% of the time.

Coach's Remarks

Solving a system of equations can be intimidating. The trick is to start with the equations that have the least number of variables (the columns with the most zeros). Also, choose one of the variables and express all other equations in terms of that variable. In the example above, we expressed (2) in terms of π_1 and used that to express the remaining equations in terms of π_1 .

Example 1.6.4.1

You are given the following transition probability matrix for a three-state (1, 2, 3) Markov chain:

$$\mathbf{P} = \begin{bmatrix} 0.4 & 0.5 & 0.1 \\ 0.3 & 0.3 & 0.4 \\ 0.0 & 0.6 & 0.4 \end{bmatrix}$$

Calculate the stationary probability of being in state 3.

Solution

Here, we have an irreducible Markov chain since all states communicate with each other, which means they are all in the same class. So, since this chain is also finite, we can apply Equation 1.6.4.2.

Form an equation using the column with the most zeros. Using the first column:

$$\begin{aligned}\pi_1 &= 0.4\pi_1 + 0.3\pi_2 \\ 0.6\pi_1 &= 0.3\pi_2 \\ \pi_1 &= 0.5\pi_2\end{aligned}$$

We have an equation in terms of π_2 . Form another equation using a different column, and express the equation in terms of π_2 . Using the third column:

$$\begin{aligned}\pi_3 &= 0.1\pi_1 + 0.4\pi_2 + 0.4\pi_3 \\ 0.6\pi_3 &= 0.1\pi_1 + 0.4\pi_2 \\ \pi_3 &= \frac{1}{6}\pi_1 + \frac{2}{3}\pi_2\end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{6}(0.5\pi_2) + \frac{2}{3}\pi_2 \\
 &= 0.75\pi_2
 \end{aligned}$$

Using $\sum_j \pi_j = 1$, we have:

$$\begin{aligned}
 \pi_1 + \pi_2 + \pi_3 &= 1 \\
 0.5\pi_2 + \pi_2 + 0.75\pi_2 &= 1 \\
 2.25\pi_2 &= 1 \\
 \pi_2 &= 0.4444
 \end{aligned}$$

Thus, the stationary probability of being in state 3 is:

$$\begin{aligned}
 \pi_3 &= 0.75\pi_2 \\
 &= \mathbf{0.3333}
 \end{aligned}$$



We can expand on the concept of calculating the long-run proportion of time spent in state j . Suppose we receive a reward, $r(j)$, every time the process is in state j . On average, the reward received will be:

$$\sum_{j=1}^{\infty} r(j)\pi_j$$

Example 1.6.4.2

An auto insurance company models their policyholders using a Markov chain with the following three states:

- 1: Preferred
- 2: Standard
- 3: Risky

The transition probability matrix is given as follows:

$$\mathbf{P} = \begin{bmatrix} e^{-1} & 1 - e^{-1} & 0 \\ e^{-1} & e^{-1} & 1 - 2e^{-1} \\ 0 & e^{-1} & 1 - e^{-1} \end{bmatrix}$$

The company requires the policyholders to pay an annual premium each year depending on their states.

State	Annual Premium
Preferred	600
Standard	900
Risky	1,200

Calculate the average annual premium paid by a policyholder.

Solution

To calculate the average annual premium, we have to calculate the long-run proportions for each state.

Using the first column:

$$\begin{aligned}\pi_1 &= e^{-1}\pi_1 + e^{-1}\pi_2 \\ \pi_1 &= \frac{e^{-1}}{1 - e^{-1}}\pi_2\end{aligned}$$

Using the third column:

$$\begin{aligned}\pi_3 &= (1 - 2e^{-1})\pi_2 + (1 - e^{-1})\pi_3 \\ \pi_3 &= \frac{1 - 2e^{-1}}{e^{-1}}\pi_2\end{aligned}$$

Then, using $\sum_{j=1}^3 \pi_j = 1$, we have:

$$\begin{aligned}\pi_1 + \pi_2 + \pi_3 &= 1 \\ \frac{e^{-1}}{1 - e^{-1}}\pi_2 + \pi_2 + \frac{1 - 2e^{-1}}{e^{-1}}\pi_2 &= 1 \\ \pi_2 &= 0.4347\end{aligned}$$

$$\begin{aligned}\pi_1 &= 0.2530 \\ \pi_3 &= 0.3123\end{aligned}$$

Therefore, the average annual premium paid by a policyholder is:

$$\begin{aligned}\sum_{j=1}^3 r(j)\pi_j &= 600(0.2530) + 900(0.4347) + 1,200(0.3123) \\ &= \mathbf{917.7769}\end{aligned}$$



Limiting Probabilities

Suppose we calculate the four-step and eight-step transition matrices for the three-state Markov chain for weather conditions.

$$\mathbf{P} = \begin{bmatrix} 0.4 & 0.5 & 0.1 \\ 0.1 & 0.2 & 0.7 \\ 0.6 & 0.0 & 0.4 \end{bmatrix}$$

$$\mathbf{P}^{(4)} = \begin{bmatrix} 0.4233 & 0.2370 & 0.3397 \\ 0.3792 & 0.2811 & 0.3397 \\ 0.3792 & 0.2370 & 0.3838 \end{bmatrix}$$

$$\mathbf{P}^{(8)} = \begin{bmatrix} 0.3979 & 0.2475 & 0.3547 \\ 0.3959 & 0.2494 & 0.3547 \\ 0.3959 & 0.2475 & 0.3566 \end{bmatrix}$$

Notice the n -step transition probabilities within each column get closer and closer together over time, e.g. $P_{1,1}^8 \approx P_{2,1}^8 \approx P_{3,1}^8$. In fact, for this Markov chain, each column converges to some value as n approaches infinity. These values are the *limiting probabilities* of the corresponding states.

Notice the 8-step transition probabilities are relatively close to the long-run proportions for each state, e.g. $P_{1,1}^8 \approx P_{2,1}^8 \approx P_{3,1}^8 \approx \pi_1 = 0.3967$. In fact, the limiting probabilities are equal to the long-run proportions if the Markov chain is aperiodic.

A Markov chain is *aperiodic* if it has limiting probabilities. In contrast, a Markov chain that does not have limiting probabilities is *periodic*.

Consider a two-state Markov chain with the following transition matrix:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

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

The Markov chain alternates between states 1 and 2 at every step. This means that the process stays in state 1 half of the time and in state 2 the other half of the time. Thus, the long-run proportions for both states are 0.5. However, the n -step transition probabilities do not converge. Consider the following matrices:

$$\mathbf{A} = \mathbf{A}^{(3)} = \dots = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$
$$\mathbf{A}^{(2)} = \mathbf{A}^{(4)} = \dots = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

The n -step transition probabilities alternate between 0 and 1 indefinitely. So, this Markov chain does not have limiting probabilities. Thus, it is periodic. To check if a chain is aperiodic or periodic, we have to calculate the n -step transition probabilities for multiple values of n and see if the probabilities converge.

A Markov chain that is irreducible, positive recurrent, and aperiodic is called *ergodic*.

1.6.5 Time Spent in Transient States

🕒 20m

Suppose we have a set of transient states, and we are interested in the expected number of time periods spent in the transient states. Let \mathbf{P}_T be the transition probability matrix for transient states only. Thus, \mathbf{P}_T is a matrix with only transition probabilities from one transient state to another transient state. Note that the rows in \mathbf{P}_T do not necessarily sum to 1.

$s_{i,j}$ is the expected number of time periods the process is in transient state j given it started in transient state i . Let \mathbf{S} be the matrix for the values of $s_{i,j}$. Then,

$$\mathbf{S} = (\mathbf{I} - \mathbf{P}_T)^{-1} \quad (1.6.5.1)$$

where \mathbf{I} is the identity matrix. An identity matrix is a matrix with ones on the main diagonal and zeros elsewhere.

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & \dots \\ 0 & 1 & 0 & \dots \\ 0 & 0 & 1 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

We can express the values of the identity matrix as a binary variable.

$$\delta_{i,j} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$$

Calculating \mathbf{S} requires inverting a matrix. The process to invert a 2×2 matrix is shown below. It is unlikely that you will need to invert a 3×3 matrix. However, the result is provided in the appendix at the end of this section.

Coach's Remarks

Inverse Matrices

Here are a couple of refreshers on inverse matrices:

- \mathbf{A}^{-1} is the *inverse* of \mathbf{A} , such that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$, an identity matrix with ones on the main diagonal and zeros elsewhere.
- For a 2×2 matrix \mathbf{A} , where

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

its inverse is

$$\mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$f_{i,j}$ is the probability of ever making a transition into transient state j given the process started in transient state i .

$$f_{i,j} = \frac{s_{i,j} - \delta_{i,j}}{s_{j,j}} \quad (1.6.5.2)$$

Example 1.6.5.1

You are given the following transition matrix for a three-state (1, 2, 3) Markov chain:

$$\mathbf{P} = \begin{bmatrix} 0.5 & 0.5 & 0.0 \\ 0.0 & 0.8 & 0.2 \end{bmatrix}$$

$$\begin{bmatrix} 0.0 & 0.0 & 1.0 \end{bmatrix}$$

Calculate the expected number of time periods spent in state 2 given the initial state is 1.

Solution

State 3 is an absorbing state. States 1 and 2 are transient. Thus, the transition probability matrix for transient states is:

$$\mathbf{P}_T = \begin{bmatrix} 0.5 & 0.5 \\ 0.0 & 0.8 \end{bmatrix}$$

Thus, $\mathbf{I} - \mathbf{P}_T$ is:

$$\begin{aligned} \mathbf{I} - \mathbf{P}_T &= \begin{bmatrix} 1 - 0.5 & 0 - 0.5 \\ 0 - 0.0 & 1 - 0.8 \end{bmatrix} \\ &= \begin{bmatrix} 0.5 & -0.5 \\ 0.0 & 0.2 \end{bmatrix} \end{aligned}$$

The inverse of $\mathbf{I} - \mathbf{P}_T$ is:

$$\begin{aligned} \mathbf{S} &= (\mathbf{I} - \mathbf{P}_T)^{-1} \\ &= \frac{1}{0.5(0.2) - 0.0(-0.5)} \begin{bmatrix} 0.2 & 0.5 \\ 0.0 & 0.5 \end{bmatrix} \\ &= 10 \begin{bmatrix} 0.2 & 0.5 \\ 0.0 & 0.5 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 5 \\ 0 & 5 \end{bmatrix} \end{aligned}$$

This is the expected number of time periods spent in state 2 given the initial state is 1.

Thus, the expected number of time periods spent in state 2 given the initial state is 1 is:

$$s_{1,2} = 5$$



Example 1.6.5.2

You are given the following transition matrix for a four-state (1, 2, 3, 4) Markov chain:

$$\mathbf{P} = \begin{bmatrix} 0.2 & 0.8 & 0.0 & 0.0 \\ 0.6 & 0.2 & 0.1 & 0.1 \\ 0.0 & 0.7 & 0.2 & 0.1 \\ 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}$$

Calculate the probability of ever entering state 3 given the process started in state 1.

Solution

The transition matrix for the transient states is:

$$\mathbf{P}_T = \begin{bmatrix} 0.2 & 0.8 & 0.0 \\ 0.6 & 0.2 & 0.1 \\ 0.0 & 0.7 & 0.2 \end{bmatrix}$$



Thus, $\mathbf{I} - \mathbf{P}_T$ is:

$$\mathbf{I} - \mathbf{P}_T = \begin{bmatrix} 0.8 & -0.8 & 0.0 \\ -0.6 & 0.8 & -0.1 \\ 0.0 & -0.7 & 0.8 \end{bmatrix}$$

The inverse of $\mathbf{I} - \mathbf{P}_T$ is:

$$\begin{aligned} \mathbf{S} &= (\mathbf{I} - \mathbf{P}_T)^{-1} \\ &= \begin{bmatrix} \frac{95}{12} & \frac{80}{9} & \frac{10}{9} \\ \frac{20}{3} & \frac{80}{9} & \frac{10}{9} \\ \frac{35}{6} & \frac{70}{9} & \frac{20}{9} \end{bmatrix} \end{aligned}$$

The probability of ever entering state 3 given the process started in state 1 is:

$$\begin{aligned} f_{1,3} &= \frac{s_{1,3} - \delta_{1,3}}{s_{3,3}} \\ &= \frac{\frac{10}{9} - 0}{\frac{20}{9}} \\ &= \mathbf{0.5} \end{aligned}$$

This approach requires the need to invert a 3x3 matrix. The alternative solution avoids matrix inversion.



Alternative Solution

This approach requires a good grasp of probability. We are interested in the probability of state 1 ever transitioning to state 3, $f_{1,3}$. We can write $f_{1,3}$ as the sum of the probabilities of entering state 3 from each possible state after the first transition step.

$$f_{1,3} = P_{1,1} f_{1,3} + P_{1,2} f_{2,3}$$

- $P_{1,1} f_{1,3}$ is the probability of transitioning from state 1 to state 1, and then, eventually transitioning to state 3.
- $P_{1,2} f_{2,3}$ is the probability of transitioning from state 1 to state 2, and then, eventually transitioning to state 3.

Note that the terms $P_{1,3}$ and $P_{1,4} f_{4,3}$ are omitted from the equation because they are both 0.

We can do the same thing for $f_{2,3}$.

$$f_{2,3} = P_{2,1} f_{1,3} + P_{2,2} f_{2,3} + P_{2,3}$$

- $P_{2,1} f_{1,3}$ is the probability of transitioning from state 2 to state 1, and then, eventually transitioning to state 3.
- $P_{2,2} f_{2,3}$ is the probability of transitioning from state 2 to state 2, and then, eventually transitioning to state 3.
- $P_{2,3}$ is the probability of transitioning from state 2 to state 3 directly.

Solve the two equations:

$$\begin{aligned} f_{1,3} &= P_{1,1} f_{1,3} + P_{1,2} f_{2,3} \\ f_{1,3} &= 0.2 f_{1,3} + 0.8 f_{2,3} \\ f_{1,3} &= f_{2,3} \end{aligned}$$

$$\begin{aligned}f_{2,3} &= P_{2,1} f_{1,3} + P_{2,2} f_{2,3} + P_{2,3} \\f_{2,3} &= 0.6 f_{1,3} + 0.2 f_{2,3} + 0.1 \\f_{1,3} &= 0.6 f_{1,3} + 0.2 f_{1,3} + 0.1 \\0.2 f_{1,3} &= 0.1 \\f_{1,3} &= \mathbf{0.5}\end{aligned}$$

■

Example 1.6.5.3

You are given the following transition matrix for a four-state (1, 2, 3, 4) Markov chain:

$$\mathbf{P} = \begin{bmatrix} 0.2 & 0.2 & 0.5 & 0.1 \\ 0.1 & 0.6 & 0.2 & 0.1 \\ 0.1 & 0.7 & 0.1 & 0.1 \\ 0.2 & 0.1 & 0.3 & 0.4 \end{bmatrix}$$

Calculate the expected time spent before entering states 2 or 4 given the process started in state 1.

Solution

There are no absorbing states in this Markov chain. However, since we want to

calculate the expected number of time periods before entering states 2 or 4, we can treat states 2 and 4 as absorbing. Thus, we have the following transition matrix for transient states.

$$\mathbf{P}_T = \begin{bmatrix} 0.2 & 0.5 \\ 0.1 & 0.1 \end{bmatrix}$$

Thus, $\mathbf{I} - \mathbf{P}_T$ is:

$$\mathbf{I} - \mathbf{P}_T = \begin{bmatrix} 0.8 & -0.5 \\ -0.1 & 0.9 \end{bmatrix}$$

The inverse of $\mathbf{I} - \mathbf{P}_T$ is:

$$\begin{aligned} \mathbf{S} &= (\mathbf{I} - \mathbf{P}_T)^{-1} \\ &= \frac{1}{0.8(0.9) - (-0.1)(-0.5)} \begin{bmatrix} 0.9 & 0.5 \\ 0.1 & 0.8 \end{bmatrix} \\ &= \frac{1}{0.67} \begin{bmatrix} 0.9 & 0.5 \\ 0.1 & 0.8 \end{bmatrix} \\ &= \begin{bmatrix} \frac{90}{67} & \frac{50}{67} \\ \frac{10}{67} & \frac{80}{67} \end{bmatrix} \end{aligned}$$

The expected number of time periods spent before entering states 2 or 4 given the process started in state 1 is:

$$\begin{aligned} s_{1,1} + s_{1,3} &= \frac{90}{67} + \frac{50}{67} \\ &= \mathbf{2.0896} \end{aligned}$$



1.6.6 Time Reversibility

🕒 20m

Recall that an ergodic Markov chain is irreducible, positive recurrent, and aperiodic. Suppose we have a stationary ergodic Markov chain, $\{X_m, m \geq 0\}$, and we want to trace the sequence of states going back in time, i.e. $X_m, X_{m-1}, X_{m-2}, \dots$. This is called the reversed process of a Markov chain.

Then, it turns out that the reversed process of an ergodic Markov chain is itself a Markov chain with the following transition probabilities:

$$R_{i,j} = \frac{\pi_j P_{j,i}}{\pi_i} \quad (1.6.6.1)$$

If $R_{i,j} = P_{i,j}$ for every i and j , then the Markov chain is *time reversible*. Thus, for a time reversible Markov chain, the following must be true:

$$\pi_i P_{i,j} = \pi_j P_{j,i} \quad (1.6.6.2)$$

This means that for a time reversible Markov chain starting in state i , any path back to state i has the same probability as the reversed path. For example, $i \rightarrow j \rightarrow k \rightarrow i$ has the same probability as $i \rightarrow k \rightarrow j \rightarrow i$, and therefore

$$P_{i,j} P_{j,k} P_{k,i} = P_{i,k} P_{k,j} P_{j,i}$$

Recall the three-state Markov chain for modeling weather with states 1 (sunny), 2 (cloudy), and 3 (rainy).

$$\mathbf{P} = \begin{bmatrix} 0.4 & 0.5 & 0.1 \\ 0.1 & 0.2 & 0.7 \\ 0.6 & 0.0 & 0.4 \end{bmatrix}$$

Calculate the transition probabilities for the reversed process.

Let \mathbf{R} be the probability transition matrix for the reversed Markov chain. We previously calculated the long-run proportions for each state.

$$\pi_1 = 0.3967$$

$$\pi_2 = 0.2479$$

$$\pi_3 = 0.3554$$

Then, using Equation 1.6.6.1, calculate $R_{1,2}$.

$$\begin{aligned} R_{1,2} &= \frac{\pi_2 P_{2,1}}{\pi_1} \\ &= \frac{0.2479}{0.3967} (0.1) \\ &= 0.0625 \end{aligned}$$

Using the same formula, calculate the other transition probabilities. The transition probability matrix for the reversed process is:

$$\mathbf{R} = \begin{bmatrix} 0.4000 & 0.0625 & 0.5375 \\ 0.8000 & 0.2000 & 0.0000 \\ 0.1116 & 0.4884 & 0.4000 \end{bmatrix}$$

This Markov chain is not time reversible because $R_{i,j}$ does not equal $P_{i,j}$ for every i and j . However, notice that the main diagonal of the matrix remains unchanged because $R_{i,i} = P_{i,i}$ is a natural consequence of Equation 1.6.6.1.

Coach's Remarks

An ergodic Markov chain is time reversible if and only if $P_{i,j} = 0$ whenever $P_{j,i} = 0$ and any path back to state i starting in state j has the same probability as the reversed path.

From the example on weather conditions, we can immediately tell that the Markov chain (which is ergodic) is not time reversible because $P_{3,2} = 0$ but $P_{2,3} \neq 0$.

Example 1.6.6.1

You are given an incomplete transition matrix for a time-reversible three-state (1, 2, 3) Markov chain:

$$\mathbf{P} = \begin{bmatrix} 0.40 & 0.30 & P_{1,3} \\ P_{2,1} & 0.30 & 0.60 \\ 0.06 & P_{3,2} & P_{3,3} \end{bmatrix}$$

Calculate $P_{3,3}$.

Solution

Every row must sum to 1. Thus, $P_{1,3} = 0.30$ and $P_{2,1} = 0.10$.

Calculate the other transition probabilities using the fact that the Markov chain is time reversible.

$$\pi_2 = \frac{\pi_1 P_{1,2}}{P_{2,1}}$$

$$\pi_3 = \frac{\pi_1 P_{1,3}}{P_{3,1}}$$

Using $\sum_j \pi_j = 1$, we can solve for π_1 .

$$\begin{aligned}\pi_1 + \pi_2 + \pi_3 &= 1 \\ \pi_1 + \frac{\pi_1 P_{1,2}}{P_{2,1}} + \frac{\pi_1 P_{1,3}}{P_{3,1}} &= 1 \\ \pi_1 \left(1 + \frac{P_{1,2}}{P_{2,1}} + \frac{P_{1,3}}{P_{3,1}}\right) &= 1 \\ \pi_1 \left(1 + \frac{0.30}{0.10} + \frac{0.30}{0.06}\right) &= 1 \\ \pi_1 &= \frac{1}{9}\end{aligned}$$

Then, calculate π_2 and π_3 .

$$\pi_2 = \frac{1}{3}$$

$$\pi_3 = \frac{5}{9}$$

Calculate $P_{3,2}$.

$$\begin{aligned}P_{3,2} &= \frac{\pi_2 P_{2,3}}{\pi_3} \\ &= 0.36\end{aligned}$$

Thus, the transition probability from state 3 to state 2 is:

$$P_{3,3} = 1 - 0.36 - 0.06 = \mathbf{0.58}$$



Example 1.6.6.2

You are given the following transition probability matrix for a three-state (1, 2, 3) Markov chain:

$$\mathbf{P} = \begin{bmatrix} 0.3 & 0.4 & 0.3 \\ 0.4 & 0.0 & 0.6 \\ 0.0 & 1.0 & 0.0 \end{bmatrix}$$

Determine if the Markov chain is

1. irreducible.
2. positive recurrent.
3. aperiodic.
4. time reversible.

Solution to (1)

In order to be irreducible, a Markov chain must only have one class. If a chain has only one class, that means all states communicate with each other.

State 1 can access states 2 and 3 directly. State 2 can access states 1 and 3. State 3 can access state 2 directly and can access state 1 through state 2. So,

states 1, 2, and 3 communicate, which means there is only one class. Thus, **the Markov chain is irreducible.**



Solution to (2)

Since we have an irreducible finite-state Markov chain, all states are positive recurrent. Thus, **the Markov chain is positive recurrent.**

We can also conclude that the states are positive recurrent by solving the following system of equations and seeing that it has a unique solution:

$$\pi_1 = 0.3\pi_1 + 0.4\pi_2 \quad (1)$$

$$\pi_2 = 0.4\pi_1 + \pi_3 \quad (2)$$

$$\pi_3 = 0.3\pi_1 + 0.6\pi_2 \quad (3)$$

$$\pi_1 + \pi_2 + \pi_3 = 1 \quad (4)$$

Using (1), express π_1 in terms of π_2 .

$$0.7\pi_1 = 0.4\pi_2$$

$$\pi_1 = \frac{4}{7}\pi_2 \quad (5)$$

Next, using (2) and (5), express π_3 in terms of π_2 .

$$\pi_3 = \pi_2 - 0.4 \left(\frac{4}{7}\pi_2 \right)$$

$$\pi_3 = \frac{27}{35}\pi_2 \quad (6)$$

Substitute (5) and (6) into (4) and solve for π_2 .

$$\begin{aligned} \frac{4}{7}\pi_2 + \pi_2 + \frac{27}{35}\pi_2 &= 1 \\ \pi_2 &= \frac{35}{82} = 0.4268 \end{aligned} \quad (7)$$

Substitute (7) into (5) and (6) to solve for the other long-run proportions.

$$\begin{aligned} \pi_1 &= \frac{10}{41} = 0.2439 \\ \pi_3 &= \frac{27}{82} = 0.3293 \end{aligned}$$

Since we found a unique solution to the system of equations, this confirms that the Markov chain is positive recurrent.



Solution to (3)

For a Markov chain to be aperiodic, it must have limiting probabilities. If the chain is aperiodic, its limiting probabilities will equal the long-run proportions found above. We can calculate the two-step, four-step, and eight-step transition matrices to determine if the transition probabilities within each column are

converging to those proportions:

$$\mathbf{P}^{(2)} = \mathbf{P} \cdot \mathbf{P} = \begin{bmatrix} 0.25 & 0.42 & 0.33 \\ 0.12 & 0.76 & 0.12 \\ 0.40 & 0.00 & 0.60 \end{bmatrix}$$

$$\mathbf{P}^{(4)} = \mathbf{P}^{(2)} \cdot \mathbf{P}^{(2)} = \begin{bmatrix} 0.2449 & 0.4242 & 0.3309 \\ 0.1692 & 0.6280 & 0.2028 \\ 0.3400 & 0.1680 & 0.4920 \end{bmatrix}$$

$$\mathbf{P}^{(8)} = \mathbf{P}^{(4)} \cdot \mathbf{P}^{(4)} = \begin{bmatrix} 0.2443 & 0.4259 & 0.3299 \\ 0.2166 & 0.5002 & 0.2831 \\ 0.2790 & 0.3324 & 0.3886 \end{bmatrix}$$

As can be seen from $\mathbf{P}^{(8)}$, the values in the first column are converging to $\pi_1 = 0.2439$. Likewise, the values in the second column are converging to $\pi_2 = 0.4268$, and the values in the third column are converging to $\pi_3 = 0.3293$. Therefore, **the Markov chain is aperiodic.**



Coach Remarks

We can use the Excel function =MMULT() to perform the matrix multiplications.

C14	:	=MMULT(C10:E12,C10:E12)		
1	P	0.3	0.4	0.3
2				

		0.0	0.1	0.2
3		0.4	0	0.6
4		0	1	0
5				
6	P^(2)	0.25	0.42	0.33000
7		0.12	0.76	0.12
8		0.4	0	0.6
9				
10	P^(4)	0.24490	0.42420	0.33090
11		0.16920	0.628	0.20280
12		0.34000	0.168	0.492
13				
14	P^(8)	0.24426	0.42588	0.32987
15		0.21665	0.50023	0.28312
16		0.27897	0.33239	0.38864
17				
18	P^(16)	0.24395	0.42670	0.32935
19		0.24027	0.43660	0.32313
20		0.24857	0.41426	0.33717
21				
22	P^(32)	0.24390	0.42683	0.32927
23		0.24384	0.42700	0.32916
24		0.24399	0.42661	0.32941
25				
26	P^(64)	0.24390	0.42683	0.32927
27		0.24390	0.42683	0.32927
28		0.24390	0.42683	0.32927
...				

Solution to (4)

Because the Markov chain is irreducible, positive recurrent, and aperiodic, conclude that it is also ergodic. Recall that in order for an ergodic Markov chain to be time reversible, $P_{i,j}$ must equal 0 whenever $P_{j,i}$ equals 0, and vice versa.

Here, $P_{3,1} = 0$. However, $P_{1,3} = 0.3 \neq 0$. Thus, conclude that **the Markov chain is not time reversible**.



1.6.7 Applications of Markov Chains

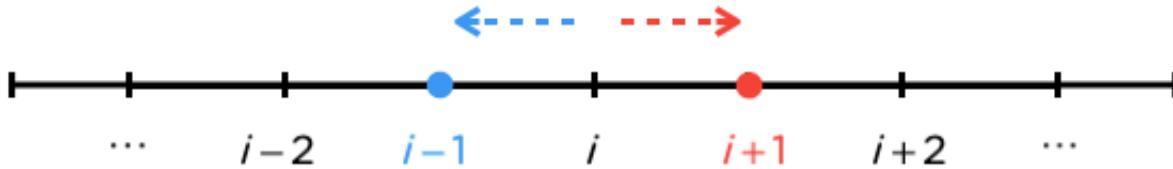
🕒 10m

Random Walk Model

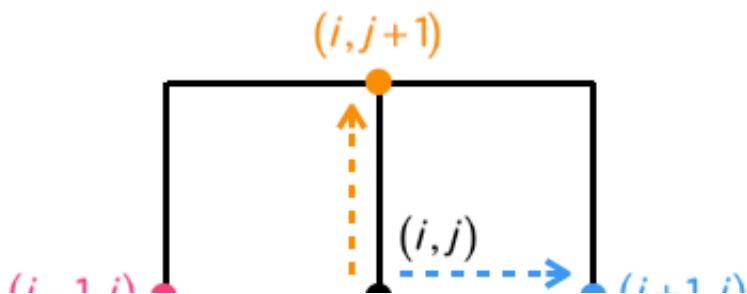
Imagine an infinite number of states arranged on a number line (i.e. one-dimensional). A one-dimensional *random walk* is a Markov chain that can only transition to state $i + 1$ or $i - 1$ at every step, given that the chain is currently at state i . The transition probability to state $i + 1$ given the current state i is p , where $0 < p < 1$.

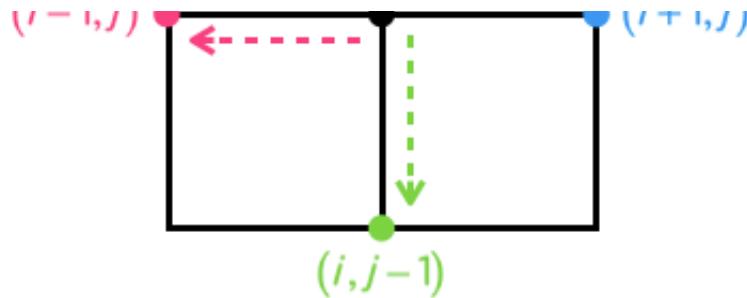
$$\begin{aligned} P_{i, i+1} &= p \\ P_{i, i-1} &= 1 - p \end{aligned}$$

A one-dimensional random walk where $p = 0.5$ is *symmetric*.



Now, imagine a two-dimensional space, where each state is represented by a pair of integers (i, j) . Then, a two-dimensional random walk is a Markov chain that can transition to state $(i + 1, j)$, state $(i - 1, j)$, state $(i, j + 1)$, or state $(i, j - 1)$ at every step, given that the chain is currently at state (i, j) . We can think of a two-dimensional random walk as a process that can move one step up, down, left, or right at each transition. If the transition probability for each direction is equal, the random walk is a symmetric two-dimensional random walk.





TRANSIENCE VS. RECURRENCE

One-dimensional and two-dimensional symmetric random walks are recurrent, while all higher-dimensional symmetric random walks are transient. In addition, all non-symmetric random walks are transient.

Gambler's Ruin Problem

Suppose that in each round of a game, a gambler wins a chip with a probability of p and loses a chip with a probability of $1 - p$. One might want to know the probability that a gambler who starts with i chips will end with j chips. This is known as the *gambler's ruin problem*.

A *gambling model* is used to model the gambler's ruin problem. It is similar to a random walk. However, instead of having an infinite number of states, a gambling model has a finite number of states. Thus, it is a finite Markov chain. A gambling model has the following properties:

- $P_{0,0} = P_{j,j} = 1$. This is because the gambler stops when he/she has either 0 or j chips. So, 0 and j are absorbing states.
- $P_{i,i+1} = p = 1 - P_{i,i-1}$, where $i = 1, 2, \dots, j-1$
- There are three classes: $\{0\}, \{1, 2, \dots, j-1\}, \{j\}$.
- $\{0\}$ and $\{j\}$ are recurrent, and $\{1, 2, \dots, j-1\}$ is transient.

Let P_i be the probability of starting with i chips and ending with j , and let $q = 1 - p$. So, the complement, $1 - P_i$, is the probability of starting with i chips and ending with 0.

$$P_i = \begin{cases} \frac{1 - \left(\frac{q}{p}\right)^i}{1 - \left(\frac{q}{p}\right)^j}, & p \neq \frac{1}{2} \\ \frac{i}{j}, & p = \frac{1}{2} \end{cases} \quad (1.6.7.1)$$

Let X be a random variable representing the number of chips the gambler has at the end. Since the gambler will play until he/she has 0 or j chips, X follows a distribution with two possible values, j and 0, which have probabilities P_i and $1 - P_i$, respectively.

Example 1.6.7.1 (Adapted from CAS S F2015 8)

Lauren and James decide to wager pieces of candy against each other on flips of a coin until one of them runs out of candy.

Lauren agrees to receive one piece of candy from James for every heads that is flipped and to lose one piece of candy to James for every tails.

At the beginning, Lauren has 25 candies and James has 10 candies.

Calculate

1. the variance of the final count of Lauren's candies, assuming the coin is fair.
2. the average final count of James' candies, assuming the coin is biased such that there is a 55% chance of flipping a tails.

Solution to (1)

This is an example of a gambler's ruin problem. Lauren receives one piece of candy for every heads. Since the coin is fair, the probability of heads is:

$$p = \Pr(\text{Heads}) = \frac{1}{2}$$

Lauren's final candy count will either be 0 or 35. Since Lauren starts with 25 pieces of candy, the probability that she will end with all 35 is:

$$P_{25} = \frac{25}{35}$$

Then, using the Bernoulli shortcut, the variance of the final count of Lauren's candies is:

$$(35 - 0)^2 \left(\frac{25}{35} \right) \left(\frac{10}{35} \right) = \mathbf{250}$$



Solution to (2)

James receives one piece of candy for every tails. In this case, the coin is biased such that the probability of tails is:

$$p = \Pr(\text{Tails}) = 0.55$$

James' final candy count will either be 0 or 35. Since James starts with 10 pieces of candy, the probability that he will end with all 35 is:

$$P_{10} = \frac{1 - \left(\frac{0.45}{0.55}\right)^{10}}{1 - \left(\frac{0.45}{0.55}\right)^{35}} = 0.8663$$

Thus, the average final count of James' candies is:

$$0 \cdot (1 - 0.8663) + 35 \cdot 0.8663 = \mathbf{30.3219}$$



1.6.8 Branching Processes

(L) 10m

Suppose we have a population of individuals, where each individual produces j new offspring by the end of their lifetime with a probability of P_j . Then, μ is the average number of new offspring an individual produces.

$$\mu = \sum_{j=0}^{\infty} j P_j \quad (1.6.8.1)$$

Let σ^2 be the variance of the number of new offspring an individual produces. Then,

$$\sigma^2 = \sum_{j=0}^{\infty} (j - \mu)^2 P_j \quad (1.6.8.2)$$

Let X_n be the size of the n^{th} generation. Then, X_0 is the size of the zeroth generation. When $X_0 = 1$, the mean and variance of the size of the n^{th} generation are:

$$\mathbb{E}[X_n] = \mu^n \quad (1.6.8.3)$$

$$\text{Var}[X_n] = \begin{cases} \sigma^2 \mu^{n-1} \left(\frac{1 - \mu^n}{1 - \mu} \right), & \mu \neq 1 \\ n\sigma^2, & \mu = 1 \end{cases} \quad (1.6.8.4)$$

If $X_0 = k$, the mean and variance of the size of the n^{th} generation are $k\mathbb{E}[X_n]$ and $k\text{Var}[X_n]$, respectively.

Let π_0 be the probability that the population will eventually die out when $X_0 = 1$.

$$\pi_0 = \begin{cases} 1, & \mu \leq 1 \\ \sum_{j=0}^{\infty} \pi_0^j P_j, & \mu > 1 \end{cases} \quad (1.6.8.5)$$

χ ↗

Note that for $\mu > 1$, π_0 is dependent on π_0^j . Since we have to solve a polynomial equation to determine π_0 , there may be more than one solution for π_0 . Therefore, the correct solution is the **smallest positive value** of π_0 that satisfies the equation,

$$\pi_0 = \sum_{j=0}^{\infty} \pi_0^j P_j.$$

When $X_0 = k$, the probability that the population will eventually die out is π_0^k .

Coach's Remarks

In the cases where individuals are guaranteed to produce at least 1 offspring (i.e., $P_0 = 0$), Equation (1.6.8.5) is not applicable, and $\pi = 0$.

Example 1.6.8.1

An isolated island begins with a population of 10 individuals. The number of offspring each individual produces has the following distribution:

Number of Offspring	Probability
0	0.1
1	0.5
2	0.4

Calculate:

- the expected size of the 20th generation.
- the probability that the population will eventually die out.

Solution to (1)

Let's begin with calculating the average number of offspring an individual produces.

$$\mu = 0(0.1) + 1(0.5) + 2(0.4) = 1.3$$

Thus,

$$\begin{aligned} E[X_{20}] &= 1.3^{20} \\ &= 190.0496 \end{aligned}$$

The expected size of the 20th generation is:

$$10(190.0496) \approx 1,900$$



Solution to (2)

Since $\mu > 1$, the probability that a population of one individual will eventually die out is:

$$\pi_0 = \pi_0^0 P_0 + \pi_0^1 P_1 + \pi_0^2 P_2$$

$$\pi_0 = 0.1 + 0.5\pi_0 + 0.4\pi_0^2$$
$$0.4\pi_0^2 - 0.5\pi_0 + 0.1 = 0$$

Solve for π_0 using the quadratic formula.

$$\pi_0 = \frac{-(-0.5) \pm \sqrt{(-0.5)^2 - 4(0.4)(0.1)}}{2(0.4)}$$
$$= 0.25 \quad \text{or} \quad 1$$

Recall that we choose the solution that is the smallest positive number, so $\pi_0 = 0.25$.

Then, the probability that a population with 10 individuals will eventually die out is approximately zero.

$$0.25^{10} \approx 0$$



1.6 Summary

(L) 10m

A Markov chain is a stochastic process where the conditional distribution of any future state given its present and past states is dependent only on the present state.

Multiple-Step Transition Probabilities

- The n -step transition matrix is calculated by multiplying the one-step transition probability matrix by itself $n - 1$ times.
- The probability that the Markov chain starts in state i and enters state j at time m without ever entering any states in a set \mathcal{A} depends on whether state i and state j are in set \mathcal{A} .

State i	State j	Desired Probability
$i \notin \mathcal{A}$	$j \notin \mathcal{A}$	$Q_{i,j}^m$
$i \notin \mathcal{A}$	$j \in \mathcal{A}$	$\sum_{r \notin \mathcal{A}} Q_{i,r}^{m-1} P_{r,j}$
$i \in \mathcal{A}$	$j \notin \mathcal{A}$	$\sum_{r \notin \mathcal{A}} P_{i,r} Q_{r,j}^{m-1}$
$i \in \mathcal{A}$	$j \in \mathcal{A}$	$\sum_{r \notin \mathcal{A}} \sum_{k \notin \mathcal{A}} P_{i,r} Q_{r,k}^{m-2} P_{k,j}$

- If α_i is the probability of being in state i at time 0, the unconditional probability of being in state j at time n is

$$\Pr(X_n = j) = \sum_{i=1}^{\infty} \alpha_i P_{i,j}^n$$

Classification of States

Terms	Descriptions
Absorbing	State that cannot be left once it is entered
Accessible	State that can be entered from another state
Communicating	Two states are accessible to each other
Class	A set of communicating states
Irreducible	Markov chain with only one class
Recurrent	Probability of re-entering state is 1
Transient	Probability of re-entering state is less than 1
Positive Recurrent	Finite expected number of transitions for a chain to return to state j given it started in that state
Null Recurrent	Infinite expected number of transitions for a chain to return to state j given it started in that state
Aperiodic	Markov chain that has limiting probabilities
Periodic	Markov chain that does not have limiting probabilities
Ergodic	Markov chain that is irreducible, positive recurrent, and aperiodic

Long-Run Proportions and Limiting Probabilities

- The long-run proportions (stationary probabilities) of states in an irreducible, positive recurrent Markov chain are calculated using:

$$\pi_j = \sum_i \pi_i P_{i,j}$$

$$\sum_j \pi_j = 1$$

- The long-run proportions are the limiting probabilities if the Markov chain is aperiodic.

Time Spent in Transient States

- \mathbf{S} is the matrix for the values of $s_{i,j}$, where $s_{i,j}$ is the expected number of time periods the process is in transient state j given it started in transient state i .

$$\mathbf{S} = (\mathbf{I} - \mathbf{P}_T)^{-1}$$

- $f_{i,j}$ is the probability of ever making a transition into transient state j given the process started in transient state i .

$$f_{i,j} = \frac{s_{i,j} - \delta_{i,j}}{s_{j,j}}$$

Time Reversibility

- The reversed process of an ergodic Markov chain is a Markov chain with the following transition probabilities:

$$R_{i,j} = \frac{\pi_j P_{j,i}}{\pi_i}$$

- A Markov chain is time reversible if $R_{i,j} = P_{i,j}$ for every i and j .

Random Walk

- A one-dimensional random walk is a Markov chain that can only transition to state $i + 1$ or $i - 1$, given the chain is currently in state i .
- All random walk models are transient except for one-dimensional symmetric random walks and two-dimensional symmetric random walks, which are recurrent.

Gambling Model (Gambler's Ruin Problem)

- The probability of reaching j starting with i , P_i , is:

$$P_i = \begin{cases} \frac{1 - \left(\frac{q}{p}\right)^i}{1 - \left(\frac{q}{p}\right)^j}, & p \neq \frac{1}{2} \\ \frac{i}{j}, & p = \frac{1}{2} \end{cases}$$

Branching Processes

- The mean and variance of the number of offspring an individual produces is:

$$\mu = \sum_{j=0}^{\infty} j P_j$$

$$\sigma^2 = \sum_{j=0}^{\infty} (j - \mu)^2 P_j$$

- If $X_0 = 1$, the mean and variance of the size of the n^{th} generation are:

$$\mathbb{E}[X_n] = \mu^n$$

$$\text{Var}[X_n] = \begin{cases} \sigma^2 \mu^{n-1} \left(\frac{1 - \mu^n}{1 - \mu} \right), & \mu \neq 1 \\ n\sigma^2, & \mu = 1 \end{cases}$$

- If $X_0 = 1$, the probability that the population will eventually die out, π_0 , is:

$$\pi_0 = \begin{cases} 1, & \mu \leq 1 \\ \sum_{j=0}^{\infty} \pi_0^j P_j, & \mu > 1 \end{cases}$$

- If $\mu > 1$, the probability that the population will eventually die out is the **smallest positive value** of π_0 that satisfies the equation above.
- If $X_0 = k$, the mean and variance of the size of the n^{th} generation are $kE[X_n]$ and $k\text{Var}[X_n]$, respectively. In addition, the probability that the population will eventually die out is π_0^k .

Appendix

⌚ 5m

Inverting a 3x3 Matrix

Let \mathbf{A} be an invertible matrix.

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Step 1: Create a matrix \mathbf{B} using determinants of sub-matrices.

$$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}$$

where each entry b_{ij} is calculated as the determinant of the sub-matrix which excludes the i^{th} row and j^{th} column.

For example, b_{11} is equal to the determinant of $\begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}$. Therefore,

$$b_{11} = a_{22}a_{33} - a_{23}a_{32}.$$

Similarly, b_{23} is equal to the determinant of $\begin{bmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{bmatrix}$. Therefore,

$$b_{23} = a_{11}a_{32} - a_{12}a_{31}.$$

Step 2: Apply alternating positive/negative signs.

Multiply the \mathbf{B} matrix with alternating positive signs and negative signs.

$$\mathbf{C} = \begin{bmatrix} + & - & + \\ - & + & - \\ + & - & + \end{bmatrix} \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} = \begin{bmatrix} v_{11} & -v_{12} & v_{13} \\ -b_{21} & b_{22} & -b_{23} \\ b_{31} & -b_{32} & b_{33} \end{bmatrix}$$

Step 3: Transpose the \mathbf{C} matrix.

$$\mathbf{C}^T = \begin{bmatrix} b_{11} & -b_{21} & b_{31} \\ -b_{12} & b_{22} & -b_{32} \\ b_{13} & -b_{23} & b_{33} \end{bmatrix}$$

Step 4: Calculate the determinant of the original matrix.

The determinant of a 3x3 matrix is:

$$d = a_{11}b_{11} - a_{12}b_{12} + a_{13}b_{13}$$

where each a_{ij} is from the matrix \mathbf{A} and each b_{ij} is from the matrix \mathbf{B} .

Step 5: Divide the \mathbf{C}^T matrix by the determinant to get the inverse of matrix \mathbf{A} .

The inverse of matrix \mathbf{A} is:

$$\mathbf{A}^{-1} = \frac{1}{a_{11}b_{11} - a_{12}b_{12} + a_{13}b_{13}} \begin{bmatrix} b_{11} & -b_{21} & b_{31} \\ -b_{12} & b_{22} & -b_{32} \\ b_{13} & -b_{23} & b_{33} \end{bmatrix}$$

1.7.0 Overview

 5m

The concept of life contingencies is the basis for life insurance. In life insurance, actuaries model the probability of survival and death and the payments of cash flows upon survival or death of human lives. Although it is generally uncommon to apply these concepts in the property and casualty industry due to the nature of the risks, the concepts can be helpful in certain property and casualty insurance calculations.

1.7.1 Probabilities

🕒 20m

Consider the following example.

Company X produces incandescent light bulbs. It is known that a light bulb has a 0.1 probability of failing in the first year, 0.5 probability of failing in the second year, 0.3 probability of failing in the third year, and 0.1 probability of failing in the fourth year. So, light bulbs have a maximum life of four years.

If company X produces a batch of 1,000 light bulbs, how many light bulbs are expected to fail in the first, second, third, and fourth years?

Using the probabilities given, we calculate that **100** light bulbs are expected to fail in the first year, **500** in the second year, **300** in the third year, and **100** in the fourth year, as shown in the table below.

Year, t	Number of Bulbs at Beginning of Year t	Number of Bulbs Failing in Year t
1	1,000	100
2	900	500
3	400	300
4	100	100

How does this relate to life contingencies? Rather than light bulbs, we can apply the concepts above to the survival and death of human lives.

Let's define notation related to these concepts. A life at exact age x is denoted as (x) . We use l_x to denote the number of lives at exact age x and d_x to denote the number of lives at exact age x who die before age $x + 1$. Thus,

$$d_x = l_x - l_{x+1} \quad (1.7.1.1)$$

The number of deaths during age x is the difference between the number of lives at age x and the number of lives at age $x + 1$.

For the example above, the number of light bulbs at age 0, l_0 , is 1,000. The number of deaths before age 1, d_0 , is 100.

Coach's Remarks

Note that the beginning of year t is the same as age $t - 1$. For example, the beginning of year 1 is the same as age 0.

Thus, the number of light bulbs at the beginning of year 1 is the same as the number of light bulbs at age 0, l_0 ,

The probability that a person age x dies before reaching age $x + 1$ is denoted as q_x .

$$q_x = \frac{d_x}{l_x}$$

The complement, p_x , is the probability of a person age x surviving to age $x + 1$.

$$\begin{aligned} p_x &= 1 - q_x \\ &= 1 - \frac{d_x}{l_x} \\ &= \frac{l_{x+1}}{l_x} \end{aligned}$$

Consider the previous example. What is probability of a light bulb age 1 surviving to

Consider the previous example. What is probability of a light bulb age 1 surviving to age 2, p_1 ?

For this calculation, divide the number of light bulbs at age 2 by the number of light bulbs at age 1.

$$p_1 = \frac{l_2}{l_1} = \frac{400}{900} = 0.44$$

Since the lifetime of the light bulbs can be modeled similarly to the lifetime of a human life, (x) , let's rewrite our table using the notation defined above.

x	l_x	d_x	q_x	p_x
0	1,000	100	0.10	0.90
1	900	500	0.56	0.44
2	400	300	0.75	0.25
3	100	100	1.00	0.00

We can also generalize the notation to extend the probability to cover any number of years, t . The probability of a person age x surviving to age $x + t$ is denoted as ${}_t p_x$. This is the same as a person age x surviving to age $x + 1$, and then that person at age $x + 1$ surviving to age $x + 2$, and so on until age $x + t$.

$$\begin{aligned} {}_t p_x &= p_x \cdot p_{x+1} \cdot \dots \cdot p_{x+t-1} \\ &= \frac{l_{x+1}}{l_x} \cdot \frac{l_{x+2}}{l_{x+1}} \cdot \dots \cdot \frac{l_{x+t}}{l_{x+t-1}} \\ &= \frac{l_{x+t}}{l_x} \end{aligned} \tag{1.7.1.2}$$

Likewise, ${}_t q_x$ is the probability that a person age x dies before reaching age $x + t$.

$${}_t q_x = 1 - {}_t p_x$$

$$= \frac{l_x - l_{x+t}}{l_x} \quad (1.7.1.3)$$

For the example above, calculate the probability of a light bulb surviving at least two years after production.

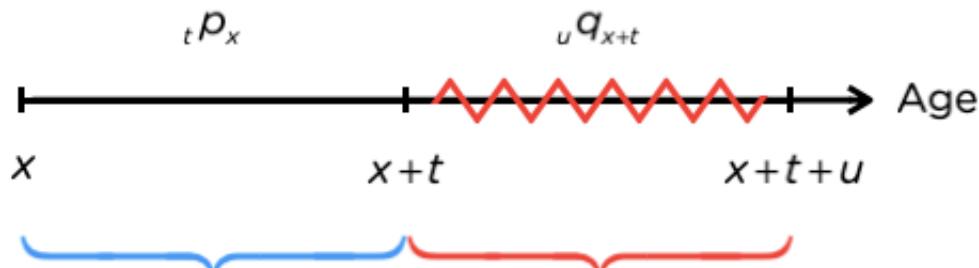
We can use Equation 1.7.1.2 for this calculation.

$$\begin{aligned} {}_2p_0 &= \frac{l_2}{l_0} \\ &= \frac{400}{1,000} \\ &= 0.40 \end{aligned}$$

You can also derive the same answer from basic principles:

$$\begin{aligned} {}_2p_0 &= p_0 \cdot p_1 \\ &= 0.9 (0.44) \\ &= 0.40 \end{aligned}$$

Below is an illustration of (x) , who survives for t years but dies within the next u years. The probability of this scenario is ${}_t p_x \cdot {}_u q_{x+t}$.



t *u*

Coach's Remarks

1. If $t = 0$, ${}_0 p_x = 1$ and ${}_0 q_x = 0$.
2. If $t = 1$, the probability of (x) surviving for one year is technically represented as ${}_1 p_x$. However, it is shortened to p_x . This is the same for q_x .
3. Also, note that:

$${}^t q_x \neq q_x \cdot q_{x+1} \cdot \dots \cdot q_{x+t-1}$$

Illustrative Life Table

The Illustrative Life Table is one part of the exam tables given to students. For each age, the Illustrative Life Table lists the number of lives remaining, the probability of death, and other relevant information related to life contingencies. Most questions from this section will require the use of the life table. Below is a snippet of the table.

Illustrative Life Table: Basic Functions and Single Benefit Premiums at $i = 0.06$

x	l_x	$1000 q_x$	\ddot{a}_x	$1000 A_x$	$1000 ({}^2 A_x)$	$1000 {}_5 E_x$	$1000 {}_{10} E_x$	$1000 {}_{20} E_x$	x
41	9,287,264	2.98	14.6864	168.69	52.01	734.40	534.99	271.12	41
42	9,259,571	3.20	14.5510	176.36	55.62	733.42	533.14	267.85	42
43	9,229,925	3.44	14.4102	184.33	59.48	732.34	531.12	264.31	43
44	9,198,149	3.71	14.2639	192.61	63.61	731.17	528.92	260.48	44
45	9,164,051	4.00	14.1121	201.20	68.02	729.88	526.52	256.34	45

At this point, you should be familiar with the first three columns. The remaining quantities will be explained in the rest of this subsection.

Example 1.7.1.1 (Adapted from CAS S F2016 16)

You are given the following information:

- There are three independent lives: (30), (40), and (60).
- Mortality follows the Illustrative Life Table.

Calculate the probability that (30) and (40) are the only people who are still alive after 30 years.

Solution

This is the same as the probability that (30) survives the next 30 years, (40) survives the next 30 years, and (60) does not survive the next 30 years. Because the lives are independent, we can multiply the probabilities of the individual lives together to calculate the joint probability. Look up the necessary values from the life table.

The probability that (30) survives the next 30 years is:

$${}_{30}p_{30} = \frac{l_{60}}{l_{30}} = \frac{8,188,074}{9,501,381} = 0.86178$$

The probability that (40) survives the next 30 years is:

$${}_{30}p_{40} = \frac{l_{70}}{l_{40}} = \frac{6,616,155}{9,313,166} = 0.71041$$

The probability that (60) does not survive the next 30 years, i.e. dies sometime in the next 30 years, is:

$$30q_{60} = \frac{l_{60} - l_{90}}{l_{60}} = \frac{8,188,074 - 1,058,491}{8,188,074} = 0.87073$$

Therefore, the probability that only (30) and (40) are alive after 30 years is:

$$(0.86178)(0.71041)(0.87073) = \mathbf{0.53307}$$



Using the probabilities calculated above, we can determine the expected value and variance of the number of deaths from a group of insureds age x in the next t years.

We will use D to represent the random variable for the number of deaths in the next t years from a group of n insureds that are all age x . Because the n lives are independent, D is a binomial random variable with a success probability of tq_x . Therefore, the expected number of deaths and the variance of the number of deaths are:

$$\mathbb{E}[D] = n \cdot t q_x$$

$$\text{Var}[D] = n \cdot t q_x (1 - t q_x)$$

Using these values, we can construct the $100k\%$ **confidence interval** of the number of deaths from this group in the next t years. Confidence intervals will be covered in detail in Section 2.6, but for now, note that a $100k\%$ confidence interval of (L, U) means that we are $100k\%$ confident that a generic parameter (in this case, the number of deaths) will fall between L and U .

Then, the $100k\%$ confidence interval for the number of deaths from a group of n insureds age x in the next t years is

$$\mathbb{E}[D] \pm z_{(1+k)/2} \sqrt{\text{Var}[D]}$$

where z_q is the $100q^{\text{th}}$ percentile of the standard normal distribution.

An insurance company is using the Illustrative Life Table to price a block of life insurance policies covering 10,000 people aged 35.

Construct the 95% confidence interval for the number of deaths in this block during the next 25 years, using a normal approximation.

For a given insured age 35, the probability of death in the next 25 years can be found using values from the Illustrative Life Table.

$$\begin{aligned} {}_{25}q_{35} &= 1 - \frac{l_{60}}{l_{35}} \\ &= 1 - \frac{8,188,074}{9,420,657} \\ &= 0.13084 \end{aligned}$$

Then, the mean and variance of the number of deaths are

$$\begin{aligned} \mathbb{E}[D] &= n \cdot {}_t q_x \\ &= 10,000 (0.13084) \\ &= 1,308.383 \end{aligned}$$

$$\begin{aligned} \text{Var}[D] &= n \cdot {}_t q_x (1 - {}_t q_x) \\ &= 10,000 (0.13084) (1 - 0.13084) \\ &= 1,137.197 \end{aligned}$$

•,••••••••

Finally, calculate the 95% confidence interval as:

$$\begin{aligned} \mathbb{E}[D] \pm z_{(1+k)/2} \sqrt{\text{Var}[D]} &= 1,308.383 \pm z_{0.975} \sqrt{1,137.197} \\ &= 1,308.383 \pm 1.960 \sqrt{1,137.197} \\ &= (1,242.29, 1,374.48) \end{aligned}$$

Life Expectancy

The life expectancy is the average number of years an individual is expected to live. The *curtate life expectancy* for an individual age x , denoted as e_x , is the expected number of full years the individual will live. This can be calculated from the life table for (x) as $\sum_{k=1}^{\infty} k p_x$.

The *complete expectation of life* is the actual life expectancy, rather than just the number of full years. For this calculation, we need to make an assumption about the time of death. One common assumption is that deaths will be uniformly distributed throughout a year. In this case, the complete expectation of life is equal to 0.5 plus the curtate life expectancy.

1.7.2 Life Insurance

(L) 30m

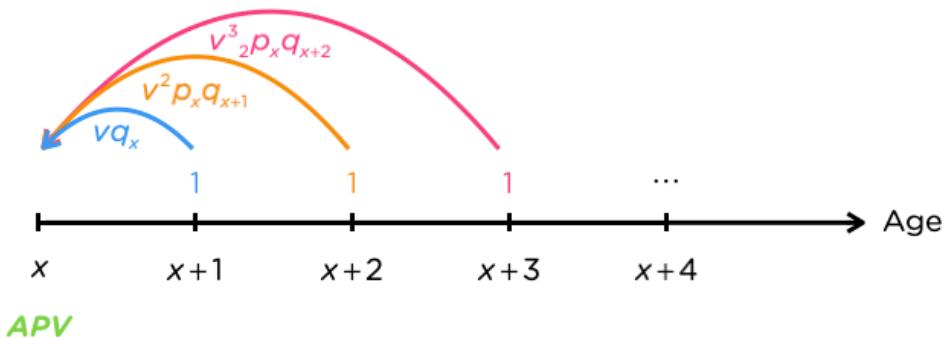
Consider a product that pays 1 at the end of the year of death for (x) . For example, if a person age x dies before reaching age $x + 1$, a payment will be made at the end of the year. If he/she dies in the second year, a payment will be made at the end of the second year, and so on. What is the present value of this payment structure?

Notice that the timing of the payment is uncertain. The present value is conditional on the survival of (x) . So, rather than discounting with interest only, we also need to discount with mortality. When we include mortality in the present value calculation, we are calculating the *actuarial present value* (APV). Let's calculate the APV of the product described above.

- A payment at the end of the first year will only be made if (x) dies in the first year. Therefore, the present value of a payment of 1 given that (x) dies in the first year is vq_x , where v is the discount factor.
- A payment at the end of the second year will only be made if (x) survives the first year but dies in the second. The present value of 1 paid at the end of two years then includes the probability of surviving the first year and dying in the second year, or $v^2 p_x q_{x+1}$.
- A payment at the end of the third year will only happen if (x) survives two years but dies in the third year. Thus, the present value is $v^3 2 p_x q_{x+2}$.

This goes on indefinitely because the product does not expire. Therefore, the actuarial present value of this product is:

$$\text{APV} = vq_x + v^2 p_x q_{x+1} + v^3 2 p_x q_{x+2} + \dots$$



Using this basis, we will discuss different types of life insurance and how to calculate the actuarial present value of each.

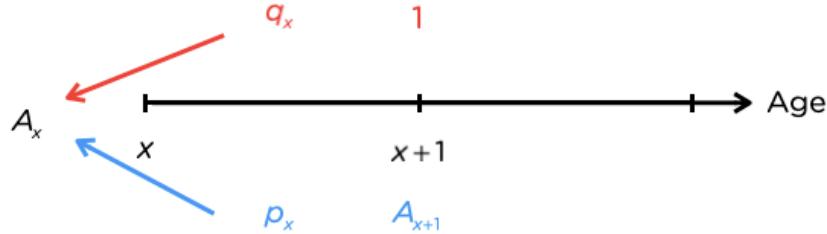
Whole Life Insurance

The product discussed above is whole life insurance. *Whole life insurance* is insurance that pays a death benefit at the end of the year of death. This insurance provides lifetime coverage. We use A_x to denote the actuarial present value of whole life insurance paying 1 for (x) . For a death benefit other than 1, simply multiply A_x by the value of the death benefit.

$$\begin{aligned}
 A_x &= \sum_{k=0}^{\infty} v^{k+1} k p_x q_{x+k} & (1.7.2.1) \\
 &= vq_x + v^2 p_x q_{x+1} + v^3 2 p_x q_{x+2} + \dots \\
 &= vq_x + vp_x (vq_{x+1} + v^2 p_{x+1} q_{x+2} + \dots)
 \end{aligned}$$

$$A_x = vq_x + vp_x A_{x+1} \quad (1.7.2.2)$$

Notice in Equation 1.7.2.2 that A_x can be expressed as a recursive formula. As shown in the illustration below, A_x is equivalent to receiving 1 at the end of the first year if (x) dies in the first year or receiving whole life insurance for $(x+1)$ at the end of the first year if (x) survives the first year.



Note that A_x is the expected value of the insurance payment, so it is the first moment of whole life insurance with a death benefit of 1.

You are given the following information:

- A whole life insurance policy has a benefit of 500 at the end of the year of death.
- Mortality follows the Illustrative Life Table.
- $i = 0.06$.

Calculate the actuarial present value of this insurance for (16).

The actuarial present value of this insurance is

$$\text{APV} = 500 A_{16}$$

Since mortality follows the Illustrative Life Table and $i = 0.06$, we are able to use any values in the Illustrative Life Table. Notice that A_{16} is not in the Illustrative Life Table, but we are able to solve for it by rearranging Equation 1.7.2.2.

$$A_{15} = vq_{15} + vp_{15}A_{16}$$

$$A_{16} = \frac{A_{15} - vq_{15}}{vp_{15}}$$

Using the values from the Illustrative Life Table, the APV of this insurance is

$$\begin{aligned} 500A_{16} &= 500 \cdot \frac{A_{15} - vq_{15}}{vp_{15}} \\ &= 500 \left(\frac{A_{15}}{vp_{15}} - \frac{q_{15}}{p_{15}} \right) \\ &= 500 \left(\frac{1.06(0.05255)}{1 - 0.00091} - \frac{0.00091}{1 - 0.00091} \right) \\ &= \mathbf{27.42} \end{aligned}$$

Coach's Remarks

In the Illustrative Life Table, 2A_x is the actuarial present value of a whole life paying 1 at twice the force of interest. Twice the force of interest is equivalent to replacing the discount factor v with v^2 .

$${}^2A_x = \sum_{k=0}^{\infty} v^{2(k+1)} k p_x q_{x+k}$$

For an effective annual interest rate of $i = 0.06$, twice the force of interest is equivalent to replacing i with $j = 0.1236$, as shown below.

$$\begin{aligned} e^\delta &= 1 + i \\ e^{2\delta} &= (1 + i)^2 = 1 + j \\ e^{2\delta} &= (1.06)^2 = 1.1236 \\ j &= 0.1236 \end{aligned}$$

For example, at an effective annual interest rate of $j = 0.1236$, the actuarial present value of whole life insurance paying 1 at the end of the year of death of a life age 20 is ${}^2A_{20}$ in the Illustrative Life Table, which is 0.0143.

2A_x also provides the second raw moment of whole life insurance with a death benefit of 1. So, this can be used along with A_x to calculate the variance for this type of insurance.

Deferred Whole Life Insurance

Deferred whole life insurance is insurance that only pays a death benefit if death occurs after t years. If death occurs during the first t years, no benefit is paid. For a death that occurs after t years, a death benefit is paid at the end of the year of death. Thus, this insurance acts like whole life insurance after the deferral period.

The actuarial present value of deferred whole life insurance is the actuarial present value of regular whole life insurance at age $x + t$ discounted with interest and mortality for t years to age x . This is because (x) has to survive t years to age $x + t$ to start getting insurance coverage.

$$\text{APV} = v^t {}_t p_x A_{x+t}$$

The term $v^t {}_t p_x$ can also be written as $_t E_x$. This represents the present value of 1 paid t years from now, conditional on surviving t years.

$${}_t E_x = v^t {}_t p_x \quad (1.7.2.3)$$

Therefore, the actuarial present value of deferred whole life insurance is:

$$\text{APV} = {}_t E_x A_{x+t} \quad (1.7.2.4)$$

Note that the term ${}_t E_x$ can be broken down into smaller components:

$${}_{t+h} E_x = {}_t E_x \cdot {}_h E_{x+t}$$

For example, ${}_{30} E_{20} = {}_{20} E_{20} \cdot {}_{10} E_{40}$.

Coach's Remarks

At this point, only the notation used in the Illustrative Life Table has been introduced. You are expected to know this notation.

The notation used in the remainder of this subsection is not required for the exam. However, you are expected to understand the different types of insurance and how to calculate their APVs using the Illustrative Life Table.

Term Life Insurance

Term life insurance is insurance that pays a death benefit at the end of the year of death if the death

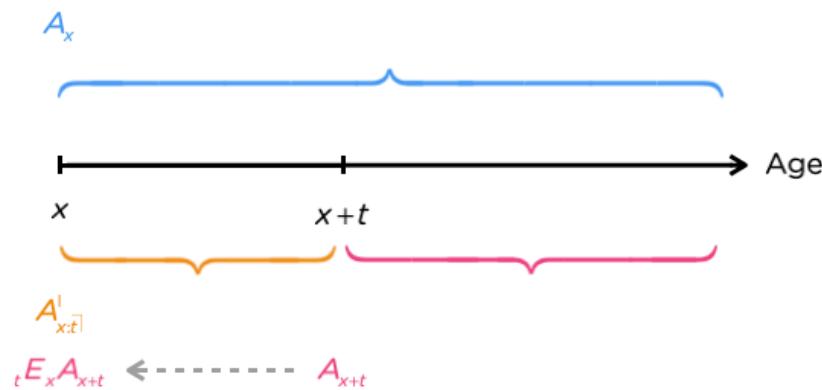
occurs in the first t years. Term life insurance can be viewed as whole life insurance with an expiration date. If death occurs after t years, no benefit is paid. The actuarial present value of term life insurance is:

$$\begin{aligned} \text{APV} = A_{x:\bar{t}}^1 &= \sum_{k=0}^{t-1} v^{k+1} k p_x q_{x+k} \\ &= v q_x + v^2 p_x q_{x+1} + \dots + v^t p_x q_{x+t-1} \end{aligned} \quad (1.7.2.5)$$

WHOLE LIFE, TERM LIFE, AND DEFERRED WHOLE LIFE

There is a relationship between whole life, term life, and deferred whole life insurances. Term life pays when deferred whole life does not and vice versa. Therefore, the actuarial present value of whole life is the sum of the actuarial present values of term life and deferred whole life, given that the length of the term is the same as the length of the deferral period and that all three are issued for the same life.

$$A_x = A_{x:\bar{t}}^1 + {}_t E_x A_{x+t} \quad (1.7.2.6)$$



That means the APVs of whole life insurance, term life insurance, and deferred whole life insurance can all be calculated using some combination of A_x , A_{x+t} , and ${}_t E_x$. Specifically,

- the APV of whole life is A_x ,
- the APV of deferred whole life is ${}_t E_x A_{x+t}$, and
- the APV of term life is $A_x - {}_t E_x A_{x+t}$.

Example 1.7.2.1

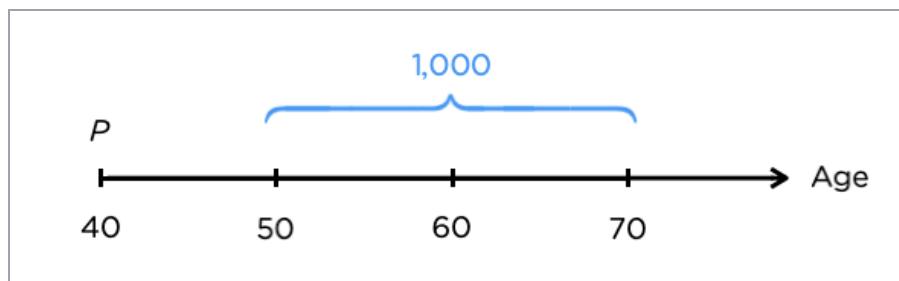
You are given the following information:

- An insurance policy pays nothing in the first 10 years, 1,000 if the insured dies in the next 20 years, and nothing thereafter.
- The benefit is paid at the end of the year of death.
- Mortality follows the Illustrative Life Table.
- $i = 0.06$.

Calculate the actuarial present value of this insurance for (40).

Solution

It helps to visualize the payment structure with a time diagram. Let P be the actuarial present value of the insurance.



This is deferred term life insurance. Start by calculating the actuarial present value at age 50 instead of 40.

Using Equation 1.7.2.6,

$$\begin{aligned} \text{APV}_{50} &= A_{50:20}^1 = A_{50} - {}_{20}E_{50}A_{70} \\ &= 0.24905 - 0.23047 (0.51495) \\ &= 0.13037 \end{aligned}$$

Next, discount the APV at age 50 ten years to get the APV at age 40.

Using Equation 1.7.2.4, we can multiply by ${}_{10}E_{40}$ to obtain the actuarial present value at age 40.

$$\begin{aligned} \text{APV}_{40} &= {}_{10}E_{40} \cdot \text{APV}_{50} \\ &= 0.53667 (0.13037) \\ &= 0.06997 \end{aligned}$$

Finally, multiply APV_{40} by 1,000 to get $P = 69.97$.

Coach's Remarks

- If mortality follows the Illustrative Life Table and $i = 0.06$, all values in the life table can be used.
- If mortality follows the Illustrative Life Table and $i \neq 0.06$, only the values in columns l_x and $1,000q_x$ can be used because all other columns are computed assuming $i = 0.06$.

Example 1.7.2.2

You are given the following information:

- The death benefit for a three-year term life insurance is 1,000 for the first year, 2,000 for the second year, and 3,000 for the third year.
- The benefit is paid at the end of the year of death.
- Mortality follows the Illustrative Life Table.
- $i = 0.05$.

Calculate the actuarial present value of the insurance for a life age 50.

Solution

Notice that the death benefit varies for each year. However, all of the formulas presented above assume a death benefit of 1 at the end of the year of death. So, use Equation 1.7.2.5 here, with a modification for the varying benefits:

$$\text{APV} = 1,000vq_{50} + 2,000v^2p_{50}q_{51} + 3,000v^3_2p_{50}q_{52}$$

Use the life table to obtain all necessary values.

x	q_x	p_x
50	0.00592	0.99408
51	0.00642	0.99358
52	0.00697	0.99303

The actuarial present value of the three-year term life insurance is:

$$\begin{aligned} \text{APV} &= 1,000 \left(\frac{1}{1.05} \right) (0.00592) + 2,000 \left(\frac{1}{1.05} \right)^2 (0.99408) (0.00642) + 3,000 \left(\frac{1}{1.05} \right)^3 (0.99408) (0.00642) \\ &= \mathbf{35.0560} \end{aligned}$$



Coach's Remarks

Note that p_x and q_x can also be obtained by applying Equation 1.7.1.2 and Equation 1.7.1.3, respectively, to the values in the l_x column. This approach is more accurate, but the rounding difference will not be large enough to impact your answer choice on the exam.

Pure Endowment

A *pure endowment* is insurance that pays a benefit at the end of year t if the insured survives until then and nothing otherwise. Because the benefit received with this type of insurance is dependent upon the insured's survival, it is also called a survival benefit.

The actuarial present value of a pure endowment with a benefit of 1 is:

$$\text{APV} = {}_t E_x$$

Endowment Insurance

Endowment insurance is insurance that pays either a death benefit or a survival benefit. If death occurs in the first t years, a death benefit will be paid at the end of the year of death, just like term life insurance. If the insured survives at least t years, a survival benefit will be paid at the end of year t , just like a pure endowment.

The actuarial present value of endowment insurance is:

$$\text{APV} = A_{\cdot\cdot\cdot\overline{t}} = \left(\sum_{k=1}^{t-1} v^{k+1} {}_k p_x q_{x+k} \right) + v^t {}_t p_x$$

$$\begin{aligned} & \left(\sum_{k=0}^{\infty} v^k p_x q_{x+k} \right) + {}_t E_x \\ & = vq_x + v^2 p_x q_{x+1} + \dots + v^{t-1} p_x q_{x+t-1} + {}_t E_x \end{aligned}$$

ENDOWMENT, TERM LIFE, AND PURE ENDOWMENT

Notice the actuarial present value of endowment insurance is the sum of the actuarial present values of term life insurance and a pure endowment.

$$A_{x:\bar{t}} = A_{x:\bar{t}}^1 + {}_t E_x \quad (1.7.2.7)$$

Example 1.7.2.3

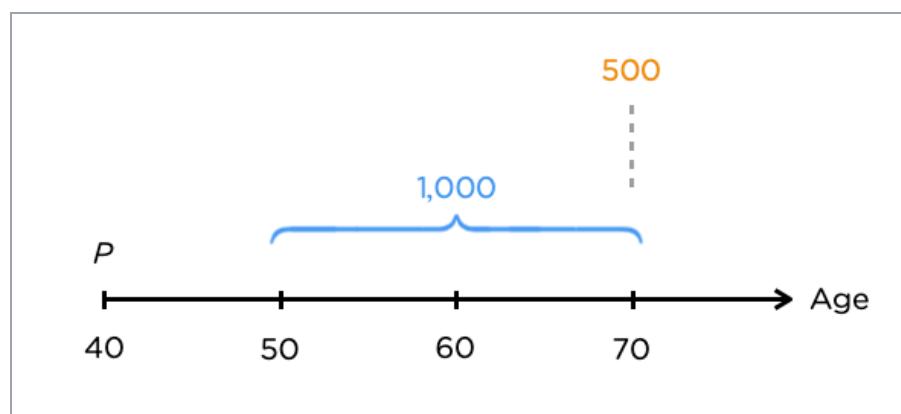
You are given the following information:

- An insurance pays nothing in the first 10 years, 1,000 if the insured dies in the next 20 years, and 500 if the insured is still alive 30 years from now.
- The death benefit is paid at the end of the year of death.
- Mortality follows the Illustrative Life Table.
- $i = 0.06$.

Calculate the actuarial present value of this insurance for a life age 40.

Solution

This insurance is similar to the one in Example 1.7.2.1, except there is an additional survival benefit of 500. Let P be the actuarial present value of the insurance.



The actuarial present value of the survival benefit is:

The actuarial present value of the survival benefit is.

$$\begin{aligned} 500 \cdot {}_{30}E_{40} &= 500 \cdot v^{30} {}_{30}p_{40} \\ &= 500 (v^{20} {}_{20}p_{40} \cdot v^{10} {}_{10}p_{60}) \\ &= 500 ({}_{20}E_{40} \cdot {}_{10}E_{60}) \\ &= 500 (0.27414 \cdot 0.45120) \\ &= 61.8460 \end{aligned}$$

Add the APV of the survival benefit to the APV of the deferred term life insurance (calculated in Example 1.7.2.1) to calculate P .

$$P = 69.9654 + 61.8460 = \mathbf{131.8114}$$



Coach's Remarks

We had to split ${}_{30}E_{40}$ up since the Illustrative Life Table only includes the values of ${}_tE_x$ for $t = 5, 10$, or 20 . Note that there are many possible ways to split ${}_{30}E_{40}$ up using values included in the Illustrative Life Table. For example, in addition to ${}_{20}E_{40} \cdot {}_{10}E_{60}$, we could've used ${}_{10}E_{40} \cdot {}_{20}E_{50}$.

1.7.3 Life Annuities

🕒 20m

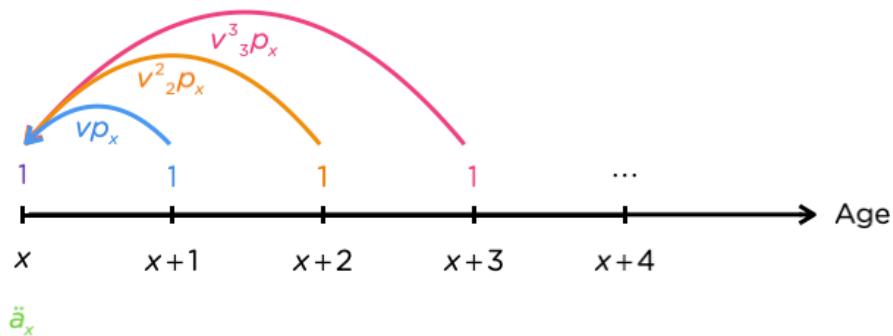
In order to purchase the different types of life insurance covered in the previous subsection, insureds typically pay annual premiums. These annual premiums paid by policyholders have the payment structure of life annuities.

A life annuity is similar to a regular annuity, just with an additional mortality component. So, payments are made as long as a person is alive.

An important actuarial symbol for life annuities is \ddot{a}_x , which is referred to as a life-annuity-due. This is the actuarial present value of a life annuity with payments of 1 made at the **beginning** of each year as long as (x) is alive.

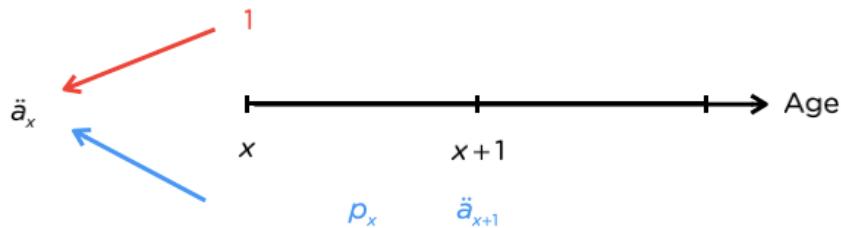
- The first payment of 1 will be made immediately at age x . The actuarial present value of this payment is simply 1.
- The next payment of 1 will only be made if the life survives to age $x + 1$. The actuarial present value of this payment is vp_x , where v is the discount factor.
- Then, another payment will be made if the life survives two years to age $x + 2$, which makes the actuarial present value $v^2 p_x$.

This payment structure continues until death.



$$\begin{aligned}\ddot{a}_x &= 1 + vp_x + v^2 p_x + \dots \\ &= 1 + vp_x(1 + vp_{x+1} + \dots) \\ &= 1 + vp_x \ddot{a}_{x+1}\end{aligned}\tag{1.7.3.1}$$

Notice \ddot{a}_x can be expressed as a recursive formula. \ddot{a}_x is equivalent to a payment of 1 now and a life annuity on a life age $x + 1$ in one year if (x) survives until age $x + 1$.



If the payments are instead made at the end of the year, assuming (x) is alive, then this is referred to as a life-annuity-immediate. In this case, the actuarial present value is denoted as a_x .

Coach's Remarks

There are different types of life annuities that correspond to the different types of life insurance covered in the previous subsection. For example, similar to term life insurance, a term life annuity is an annuity with payments made at the beginning of each year for t years as long as (x) is alive and with no payments made after t years.

Because concepts in Section 1.7.2 can be easily applied to life annuities, we will not cover them in detail here but will instead provide an example to illustrate the concepts (see the alternative solution of Example 1.7.3.1).

Example 1.7.3.1

You are given the following information:

- (50) has a five-year term insurance policy.
- She pays 200 at the beginning of each year that she is alive during the term of the insurance.
- The death benefit, Y , is paid at the end of the year of death.
- Mortality follows the Illustrative Life Table.
- $i = 0.06$.

Calculate the actuarial present value of the premium payments.

Solution

Let APV be the actuarial present value of the premium payments. We can approach this using basic principles.

From Equation 1.7.3.1,

$$\text{APV} = 200 \left(1 + vp_{50} + v^2 p_{50} + v^3 p_{50} + v^4 p_{50} \right)$$

Using the l_x values from the life table, we have:

$$\text{APV} = 200 \left[1 + \left(\frac{1}{1.06} \right) \left(\frac{8,897,913}{8,950,901} \right) + \left(\frac{1}{1.06} \right)^2 \left(\frac{8,840,770}{8,950,901} \right) + \left(\frac{1}{1.06} \right)^3 \left(\frac{8,779,128}{8,950,901} \right) + \left(\frac{1}{1.06} \right)^4 \left(\frac{8,717,585}{8,950,901} \right) \right]$$

$$\begin{aligned} &= 200 \bar{(4.4114)} \\ &= \mathbf{882.2743} \end{aligned}$$



Alternative Solution

We can use a shortcut to solve this example because the interest rate of 0.06 matches the rate used in the Illustrative Life Table.

Using an analogous version of Equation 1.7.2.6 for life annuities, we have:

$$APV = 200 (\ddot{a}_{50} - {}_5E_{50}\ddot{a}_{55})$$

Substitute values from the life table.

$$\begin{aligned} APV &= 200 [13.2668 - 0.72137 (12.2758)] \\ &= \mathbf{882.2812} \end{aligned}$$

The slight difference between the two approaches is due to the rounding of values in the Illustrative Life Table.



Life Annuities and Life Insurance

There is a valuable relationship between the actuarial present values of a life annuity and life insurance that will allow us to convert between the two easily. Before we get to that, let's review a few concepts from interest theory related to an annuity-certain, which is an annuity where the number of payments is fixed (i.e. there is no mortality component).

Consider an annuity-certain which requires an investment of 1 now and pays interest of d at the beginning of every period for n periods. At the end of n periods, you will receive your investment of 1 back. So, the present value (at an interest rate consistent with d) of the interest payments and the original investment will equal 1. From this, we get the formula for an annuity-due.

$$d\ddot{a}_{\overline{n}} + v^n = 1$$

$$\ddot{a}_{\overline{n}} = \frac{1 - v^n}{d}$$

The interest paid at the beginning of each year, d , is called discount and is equal to $\frac{i}{1+i}$. Notice that this is different than d_x used in the life table.

This is analogous to the fact that a bond with a coupon rate equal to the yield to maturity has a price that is equal to its face amount.



We can extend this concept to a life annuity and life insurance. Instead of an annuity-certain over n periods, we have a life annuity that will pay d at the beginning of every year until death. And instead of getting 1 back at the end of n periods, we get a death benefit of 1 at the end of the year of death.

Therefore, investing 1 now is equivalent to receiving d as interest at the beginning of each year when (x) is alive and receiving a payment of 1 at the end of the year of death of (x) .

$$1 = d\ddot{a}_x + A_x \quad (1.7.3.2)$$

$$\ddot{a}_x = \frac{1 - A_x}{d}$$

Example 1.7.3.2

You are given the following information:

- A whole life policy on (59) has a death benefit of 1 paid at the end of the year of death.
- The actuarial present value of the death benefit of this policy is 0.35.
- $i = 0.05$.
- Mortality follows the table below:

x	q_x
58	0.01
59	0.02
60	0.03

Calculate the actuarial present value of a whole life annuity-due with annual payments of 1 on (60).

Solution

From Equation 1.7.3.2, our goal is to calculate:

$$\ddot{a}_{60} = \frac{1 - A_{60}}{d}$$

We are given the APV of a whole life insurance on (59), $A_{59} = 0.35$. We can calculate the APV of a whole life on (60) using the recursive formula, i.e. Equation 1.7.2.2.

$$\begin{aligned} A_{59} &= vq_{59} + vp_{59}A_{60} \\ 0.35 &= \frac{1}{1.05}(0.02) + \frac{1}{1.05}(0.98)A_{60} \\ A_{60} &= 0.3546 \end{aligned}$$

The actuarial present value of a whole life annuity-due on (60) is:

$$\ddot{a}_{60} = \frac{1 - 0.3546}{\frac{0.05}{1.05}} = \mathbf{13.5536}$$



Joint Lives

We can also have annuities that are based on the survival of multiple lives. The most common type are for two lives, (x) and (y) , which are usually a married couple.

We will briefly introduce two types of annuities for joint lives. The first is called a *joint life annuity*, and its APV is denoted as \ddot{a}_{xy} . This annuity makes payments until the **first** death of (x) and (y) . The second is called a *last survivor annuity*, which makes payments until the **second** death of (x) and (y) . The APV of this type of annuity is denoted as $\ddot{a}_{\overline{xy}}$. We can relate the annuities on joint lives to the annuities on individual lives:

$$\ddot{a}_x + \ddot{a}_y = \ddot{a}_{xy} + \ddot{a}_{\overline{xy}}$$

If (x) dies first, then \ddot{a}_{xy} will be equivalent to \ddot{a}_x , and $\ddot{a}_{\overline{xy}}$ will be equivalent to \ddot{a}_y . If (y) dies first, then \ddot{a}_{xy} will be equivalent to \ddot{a}_y , and $\ddot{a}_{\overline{xy}}$ will be equivalent to \ddot{a}_x . So, regardless of the order of death, the value of \ddot{a}_{xy} will match one of the values of \ddot{a}_x and \ddot{a}_y , and the value of $\ddot{a}_{\overline{xy}}$ will match the other.

1.7.4 Equivalence Principle

🕒 10m

The **equivalence principle** is one of the most important concepts in life contingencies. It equates the actuarial present value of the benefits from an insurance policy with the actuarial present value of the premiums so that insureds receive a "fair" coverage. From an insurance company's standpoint, expenses and profit need to be included when pricing products. This means the actuarial present value of the premiums received by the company need to cover benefits as well as expenses and profit.

For this exam, we are only going to focus on benefits and net premium. **Net premium** is the amount of premium needed to cover the cost of the benefits paid by the insurance company, **ignoring expenses and profit**.

$$\text{APV}_{\text{Premium}} = \text{APV}_{\text{Benefit}} \quad (1.7.4.1)$$

Example 1.7.4.1

You are given the following information:

- A three-year term life insurance policy on (55) provides a level death benefit of 3,000.
- The premium P is paid at the beginning of each year during the term as long as the insured is alive.
- The death benefit is paid at the end of the year of death.
- $i = 0.06$.
- Mortality is as follows:

x	q_x
55	0.01
56	0.02
57	0.04

Calculate P using the equivalence principle.

Solution

Let's calculate the actuarial present value of the death benefit, $\text{APV}_{\text{Death Benefit}}$.

$$\begin{aligned}\text{APV}_{\text{Death Benefit}} &= 3,000 (vq_{55} + v^2 p_{55} q_{56} + v^3 {}_2p_{55} q_{57}) \\ &= 3,000 \left[\frac{1}{1.06} (0.01) + \frac{1}{1.06^2} (0.99) (0.02) + \frac{1}{1.06^3} (0.99) (0.98) (0.04) \right] \\ &= 178.9195\end{aligned}$$

Next, calculate the actuarial present value of the premium payments, $\text{APV}_{\text{Premium}}$.

$$\begin{aligned}\text{APV}_{\text{Premium}} &= P (1 + vp_{55} + v^2 {}_2p_{55}) \\ &= P \left(1 + \frac{1}{1.06} (0.99) + \frac{1}{1.06^2} (0.99) (0.98) \right) \\ &= 2.7974P\end{aligned}$$

By the equivalence principle, the actuarial present value of these two should be the same.

$$\begin{aligned}2.7974P &= 178.9195 \\ P &= \mathbf{63.9584}\end{aligned}$$



Example 17.1.2 (Adapted from CAS C Exam 2015 121)

Example 1.7 - The Insurance from CAS 3.1 2019-19,

You are given the following information:

- A life insurance company issues a special three-year term life insurance policy on (x) .
- If the insured dies, there will be a random draw at the end of the year of death. The death benefit will either be 100,000, with a probability of 0.75, or it will be 0, with a probability of 0.25.
- At the beginning of each year the policy is in effect, if (x) is alive, there will be a random draw. The insured will either pay P , with a probability of 0.2, or nothing, with a probability of 0.8.
- The random draws are independent.
- The probability of dying between age $x + t$ and age $x + t + 1$ is $q_{x+t} = 1 - 0.98^{1+t}$ for $t = 0, 1, 2, \dots$
- $i = 0.05$.

Calculate P using the equivalence principle.

Solution

The probabilities of death and survival are calculated and tabulated below:

t	q_{x+t}	p_{x+t}
0	0.02	0.98
1	0.0396	0.9604
2	0.058808	0.941192

The actuarial present value of the benefit is:

$$\begin{aligned} APV &= 0.75(100,000) (vq_x + v^2 p_x q_{x+1} + v^3 p_x q_{x+2}) + 0.25(0) \\ &= 7,654.5526 \end{aligned}$$

The actuarial present value of the premium is:

$$\begin{aligned} \text{APV} &= 0.2P(1 + vp_x + v^2 2p_x) + 0.8(0) \\ &= 0.5574P \end{aligned}$$

By the equivalence principle, the actuarial present value of these two should be the same.

$$\begin{aligned} 0.5574P &= 7,654.5526 \\ P &= \mathbf{13,732.4929} \end{aligned}$$



Level Premium Products

Assume that an insurance company sells a level premium product. Then, if the mortality is non-decreasing over the life of the policy, insureds overpay in the early years of the contract and underpay in the later years. So, the "extra" premium received at the beginning of the policy helps cover the "insufficient" premium in later years when the mortality is higher.

If the mortality is decreasing over the life of the policy, insureds underpay in the early years of the contract and overpay in the later years. So, the "insufficient" premium at the beginning of the policy is made up in later years when the mortality is lower.

However, this works out only if the policyholder continues to pay for insurance coverage each year. If the policyholder *lapses*, or does not renew the policy, then the benefits paid will not be equal to the premiums received. In the first case, when mortality is non-decreasing, the insurance company has additional premium, although state laws typically require insurance companies to share some of this additional premium with the lapsed insured. But in the second case, when mortality is decreasing, the insurance company will not have enough premium to cover the benefits if the policyholder lapses. In fact, it would be cheaper for the policyholder to lapse after they have underpaid for the coverage and then purchase more insurance to get equivalent coverage at a lower cost.

~~Insurance to get equivalent coverage at a lower cost.~~

Thus, insurance companies typically don't sell level premium products when mortality is not constant. This is especially important in the property and casualty industry, as P&C risks often have decreasing mortality. This is due to the bigger risks typically having accidents earlier on, leaving the group to mainly have better risks over time.

1.7 Summary

🕒 5m

	Important Formulas
Number of Deaths	$d_x = l_x - l_{x+1}$
Probability of Survival	$t p_x = \frac{l_{x+t}}{l_x}$
Probability of Death	$t q_x = \frac{l_x - l_{x+t}}{l_x}$
Curtate Life Expectancy	$e_x = \sum_{k=1}^{\infty} k p_x$
Complete Expectation of Life	$0.5 + \sum_{k=1}^{\infty} k p_x$
Whole Life Insurance	$\begin{aligned} A_x &= \sum_{k=0}^{\infty} v^{k+1} k p_x q_{x+k} \\ &= v q_x + v p_x A_{x+1} \end{aligned}$
Whole Life Annuity	$\begin{aligned} \ddot{a}_x &= \sum_{k=0}^{\infty} v^k k p_x \\ &= 1 + v p_x \ddot{a}_{x+1} \end{aligned}$
Mortality Discount Factor	$t E_x = v^t t p_x$

- Let D represent the random variable for the number of deaths in the next t years from a group of n insureds that are all age x . The mean, variance, and $100k\%$ confidence interval of D are

$$\mathbb{E}[D] = n \cdot {}_t q_x$$

$$\text{Var}[D] = n \cdot {}_t q_x (1 - {}_t q_x)$$

$$\mathbb{E}[D] \pm z_{(1+k)/2} \sqrt{\text{Var}[D]}$$

where z_q is the $100q^{\text{th}}$ percentile of the standard normal distribution.

- The actuarial present value of whole life insurance is the sum of the actuarial present values of term life and deferred whole life.
- The actuarial present value of endowment insurance is the sum of the actuarial present values of term life and a pure endowment.
- The actuarial present value of life insurance is related to the actuarial present value of a life annuity as follows:

$$A_x = 1 - d\ddot{a}_x$$

- Annuities on joint lives are related to annuities on individual lives as follows:

$$\ddot{a}_x + \ddot{a}_y = \ddot{a}_{xy} + \ddot{a}_{\bar{xy}}$$

- Under the equivalence principle, the actuarial present value of benefits paid must be equal to the actuarial present value of premiums received.

1.8.0 Overview

 5m

Simulation is important in many fields, especially in statistics and actuarial science. Actuarial work often relies on simulation as a way to model events, e.g. how a premium increase would impact policyholder behavior and the company's future cash flows. Simulation allows actuaries and statisticians to gain important insights into the real world.

One of the many well-known simulation approaches is the Monte Carlo simulation method. In this subsection, we will cover uniform number generation, two methods for simulating continuous random variables, and the basics of Monte Carlo simulation. For this subsection, we will refer to independent random variables simulated from a given distribution as *random numbers*.

1.8.1 Uniform Number Generation

 5m

Many simulation methods make use of uniform random numbers. So, before we get into other simulation methods, we will first learn how uniform random numbers are generated.

It is unlikely that you will need to generate uniform random numbers during the exam. Instead, you will likely be given these numbers to generate random numbers from a different distribution specified in the problem. However, for the exam, the learning objectives state that candidates should understand the mechanics of generating uniform random numbers.

Conventional Method

One way to simulate uniform random numbers on $[0, 1)$ is to number ten identical pieces of paper from 0 to 9, put them in a bowl, and randomly draw one at a time with replacement.

For example, if the first five numbers drawn are 9, 6, 4, 0, and 4, then the generated uniform number will be 0.96404, rounded to five decimal places.

Computational Method

Digital computers do not generate uniform random numbers using the approach described above. Instead, they simulate uniform random numbers using pseudo random numbers. The steps are as follows:

1. Choose an initial value X_0 . This is called the seed.
2. Generate pseudo random numbers using the following equation:

$$X_{n+1} = (aX_n + c) \text{ mod } m, \quad n \geq 0$$

a , c , and m are specified in advance. These values need to be chosen carefully for the numbers to mimic true randomness.

3. Divide the resulting value X_{n+1} by m to simulate a uniform random number U on $[0, 1)$.

Coach's Remarks

Given two positive numbers x and y , the modulo operation, or $x \bmod y$, returns the remainder of the division $\frac{x}{y}$. For example, $33 \bmod 7$ returns 5.

Since the initial value X_0 is chosen, it should not be used as a random number. However, the subsequent simulated values can be used as random numbers, and the process is repeated until enough random numbers are generated.

Consider the following example.

Let $a = 61$, $c = 101$, $m = 100$, and $X_0 = 50$.

Using pseudo random numbers, simulate two uniform random numbers between 0 and 1.

The first iteration of the equation above is:

$$\begin{aligned}X_1 &= (61 \cdot 50 + 101) \bmod 100 \\&= 3,151 \bmod 100 \\&= 51\end{aligned}$$

Therefore, the first simulated uniform random number is:

$$U_1 = \frac{X_1}{m} = \frac{51}{100} = \mathbf{0.51}$$

The second iteration is:

$$\begin{aligned} X_2 &= (61 \cdot 51 + 101) \bmod 100 \\ &= 3,212 \bmod 100 \\ &= 12 \end{aligned}$$

Therefore, the second simulated uniform random number is:

$$U_2 = \frac{X_2}{m} = \frac{12}{100} = \mathbf{0.12}$$

1.8.2 Inversion Method

 5m

The inversion method and the acceptance-rejection method are two ways to generate random numbers from a specified distribution using uniform random numbers. We will discuss the inversion method first.

Suppose we want to simulate values from a distribution with density function f using uniform random numbers. Here is one way to do it. For this, let U be a uniform $[0, 1)$ random number, and let X be a random number with cumulative distribution function F .

We can simulate a random number X with cumulative distribution function F using these steps:

1. Simulate a uniform random number U .
2. Set $F(X) = U$ and solve for X . This is equivalent to deriving the inverse function of F , which is denoted as F^{-1} , and evaluating it at U . Then, the random number X has the cumulative distribution function F .

Repeat these steps until you have a large enough pool of simulated numbers. This approach is called the *inverse transformation method*, or the *inversion method*.

Example 1.8.2.1

You are given the following density function:

$$f(x) = \frac{1}{5}e^{-x/5}, \quad x > 0$$

You want to simulate random numbers from this distribution. You have generated two uniform $[0, 1)$ random numbers, 0.18631 and 0.37629.

Using the inversion method, calculate the first two simulated values from this distribution.

Solution

Notice that the given density function is for an exponential distribution with mean 5. So, the CDF is:

$$F(x) = 1 - e^{-x/5}, \quad x > 0$$

The inverse of this function is:

$$F^{-1}(u) = -5 \ln(1 - u)$$

So, the first generated exponential random number is:

$$\begin{aligned} X_1 &= F^{-1}(U_1) \\ &= -5 \ln(1 - 0.18631) \\ &= \mathbf{1.0309} \end{aligned}$$

The second generated exponential random number is:

$$\begin{aligned} X_2 &= F^{-1}(U_2) \\ &= -5 \ln(1 - 0.37629) \\ &= \mathbf{2.3603} \end{aligned}$$



Example 1.8.2.2

You are given:

- Losses follow a lognormal distribution with $\mu = 3.5$ and $\sigma = 1.96$.
- Losses are simulated using the inversion method.
- The following random numbers are generated from a uniform distribution on the interval $[0, 1]$: 0.8461, 0.4013, 0.7123, 0.5832.

Calculate the average simulated loss.

Solution

Using the inversion method, calculate the simulated losses. The CDF of a lognormal distribution is:

$$F(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right)$$

We want the values of x that satisfy this equation:

$$\begin{aligned} F(x) &= u \\ \Phi\left(\frac{\ln x - \mu}{\sigma}\right) &= u \\ \frac{\ln x - \mu}{\sigma} &= \Phi^{-1}(u) \\ x &= \exp[\sigma \cdot \Phi^{-1}(u) + \mu] \\ &= \exp[1.96 \cdot \Phi^{-1}(u) + 3.5] \end{aligned}$$

Thus, the simulated losses are:

i	U_i	$\Phi^{-1}(U_i)$	X_i
1	0.8461	1.02	244.4963
2	0.4013	-0.25	20.2874
3	0.7123	0.56	99.2458
4	0.5832	0.21	49.9789

The average simulated loss is

$$\frac{244.4963 + 20.2874 + 99.2458 + 49.9789}{4} = \mathbf{103.5021}$$

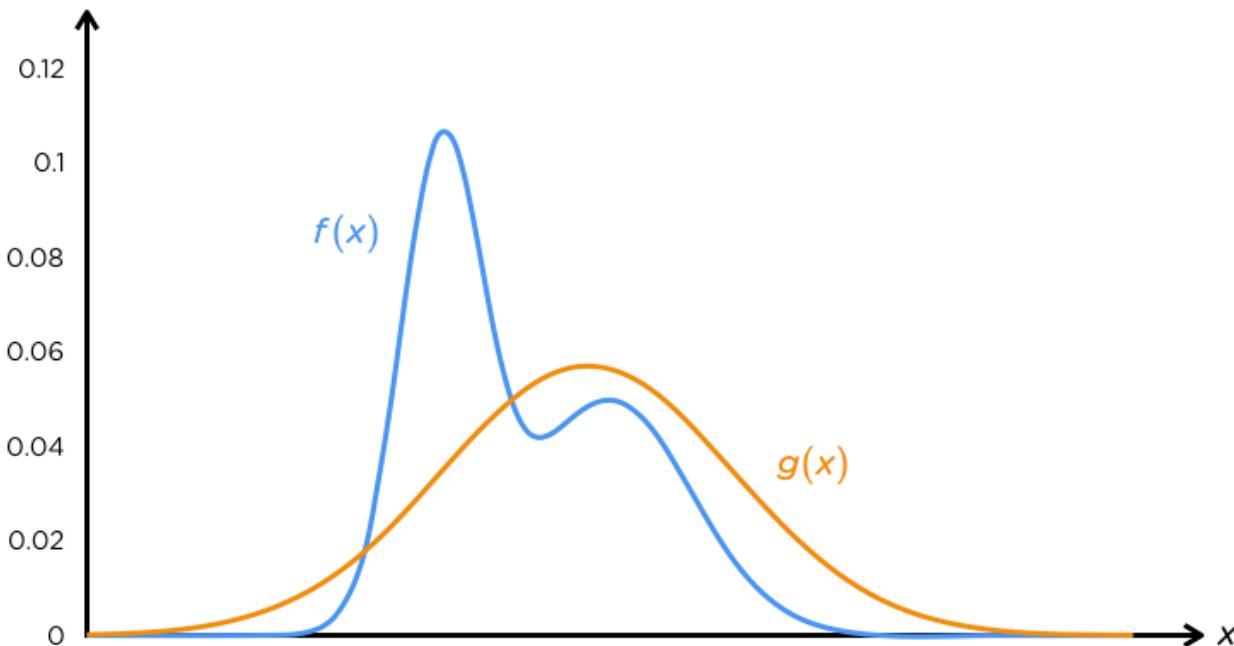


1.8.3 Acceptance-Rejection Method

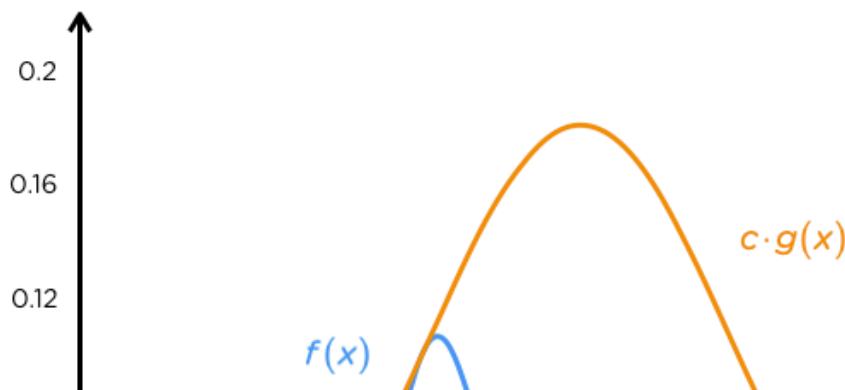
🕒 15m

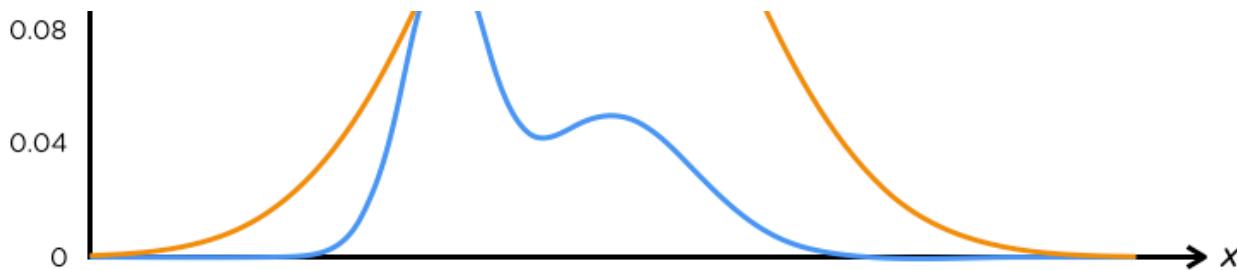
Sometimes, it is difficult or impossible to derive the inverse of a function. In these cases, we cannot rely on the inversion method to simulate random values. One method we can use instead is the *acceptance-rejection method*, or the *rejection method*.

Suppose we can simulate values from a distribution with density function g , and we want to simulate values from another distribution with density function f . For example, we may have the following density functions:



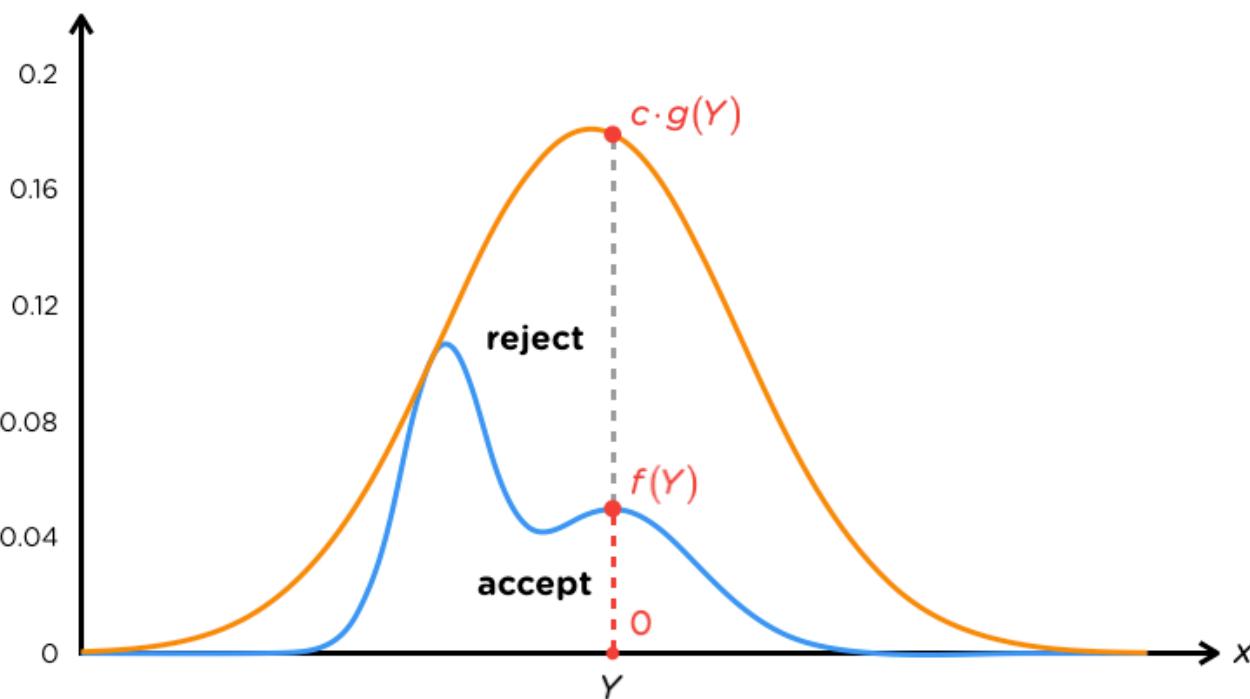
We first find a constant c that guarantees that cg will always be at least as high as f .





Then, let Y be a simulated random number from the known density g . To determine if Y qualifies as a simulated value from density f , we compare the density heights of $f(Y)$ and $cg(Y)$. Since the density f is completely under the density cg , no portion of f will be left out when we compare the heights.

The idea is that Y will be rejected more often if the height of $f(Y)$ is low (i.e. an unlikely sample) relative to $cg(Y)$. To be specific, we use a uniform random number between 0 and $cg(Y)$. If this uniform random number is also between 0 and $f(Y)$, we accept the simulated value Y . Otherwise, we reject it.



Recall that uniform $[0, 1]$ random variables are typically used in simulation. So, we can equivalently use a uniform random number between 0 and $\frac{cg(Y)}{cg(Y)} = 1$. Then, if the random number is also between 0 and $\frac{f(Y)}{cg(Y)}$, we accept Y as being from the distribution

with density function f . Otherwise, we reject it.

We can summarize the steps for the rejection method as follows. For this, let U be a uniform $[0, 1)$ random number. Then, we can simulate a random number with density function f by following these steps:

1. Find a constant c that satisfies the inequality below.

$$\frac{f(x)}{g(x)} \leq c, \quad \text{for all } x$$

The smallest c that satisfies the inequality is the maximum value of the function $\frac{f(x)}{g(x)}$, for all x .

2. Simulate a random number Y with density function g , and simulate a random number U .
3. Calculate $\frac{f(Y)}{cg(Y)}$ for the simulated value Y .
4. Compare the simulated value U to $\frac{f(Y)}{cg(Y)}$. If $U \leq \frac{f(Y)}{cg(Y)}$, accept the value Y . Otherwise, reject it and return to step 2.

Repeat these steps until you have generated enough random numbers. The accepted numbers follow the distribution of the random variable with density function f . Note that for this method, the number of iterations needed to get a simulated random number follows a geometric distribution with mean c .

Coach's Remarks

In order to align with the parameterization provided in the MAS-I exam tables for the geometric distribution, we can rephrase the distribution representing the number of iterations needed to get a simulated random number.

The number of failed iterations before obtaining a simulated random number has a geometric distribution with mean $\beta = c - 1$.

Coach's Remarks

The smallest desired constant c is the global maximum of the function $\frac{f(x)}{g(x)}$.

Recall from calculus that setting the first derivative to zero will yield the critical points. These critical points correspond to the local maxima and local minima. To identify if a critical point corresponds to a local maximum or a local minimum, use the second derivative.

In some cases, the global maximum may not be one of the local maxima. It is then crucial to evaluate $\frac{f(x)}{g(x)}$ at the endpoints and compare with the local maxima. The global maximum is the largest $\frac{f(x)}{g(x)}$ value.

Example 1.8.3.1

You are given:

- X follows a beta distribution with $a = 2$, $b = 2$, and $\theta = 2$.
- Values of X are simulated using the rejection method (assuming the minimum possible value for c) with the following density function:

$$g(x) = \frac{1}{2}, \quad 0 < x < 2$$

- The following random numbers are generated from the given distribution above: 1.3667, 0.3054, 1.8325, 0.8442.

- The following random numbers are generated from a uniform distribution on the interval $[0, 1]$: 0.4013, 0.8461, 0.2097, 0.5832.

Calculate the first two simulated values of X .

Solution

Note that in this example, X refers to a random variable, rather than a simulated random number. Exam questions will often use uppercase letters, like X , to refer to both random variables and simulated random numbers. In this example, we will use X exclusively for the random variable and Y and U to represent simulated random numbers.

The density function of the beta distribution is:

$$f(x) = \frac{3}{2}x \left(1 - \frac{x}{2}\right)$$

Calculate the constant c .

$$\begin{aligned} \frac{f(x)}{g(x)} &= 3x \left(1 - \frac{x}{2}\right) \\ &= 3x - \frac{3x^2}{2} \\ \frac{d}{dx} \left[\frac{f(x)}{g(x)} \right] &= 3 - 3x \\ 3 - 3x &= 0 \\ x &= 1 \\ c &= \frac{f(1)}{g(1)} \\ &= 3(1) - \frac{3(1^2)}{2} \\ &= 1.5 \end{aligned}$$

Check the first pair of random numbers, i.e. $Y_1 = 1.3667$ and $U_1 = 0.4013$.

$$\begin{aligned}\frac{f(Y_1)}{cg(Y_1)} &= 2Y_1 - Y_1^2 \\ &= 2(1.3667) - 1.3667^2 \\ &= 0.8655 \\ &\geq 0.4013\end{aligned}$$

Since the value is greater than 0.4013, **1.3667** is the first accepted simulated value.

Check the second pair of random numbers, i.e. $Y_2 = 0.3054$ and $U_2 = 0.8461$.

$$\begin{aligned}\frac{f(Y_2)}{cg(Y_2)} &= 2Y_2 - Y_2^2 \\ &= 2(0.3054) - 0.3054^2 \\ &= 0.5175 \\ &\not\geq 0.8461\end{aligned}$$

The value is rejected, so we repeat the steps with the third pair of random numbers, i.e. $Y_3 = 1.8325$ and $U_3 = 0.2097$.

$$\begin{aligned}\frac{f(Y_3)}{cg(Y_3)} &= 2Y_3 - Y_3^2 \\ &= 2(1.8325) - 1.8325^2 \\ &= 0.3069 \\ &\geq 0.2097\end{aligned}$$

Therefore, the second simulated value is **1.8325**.

1.8.4 Monte Carlo Simulation

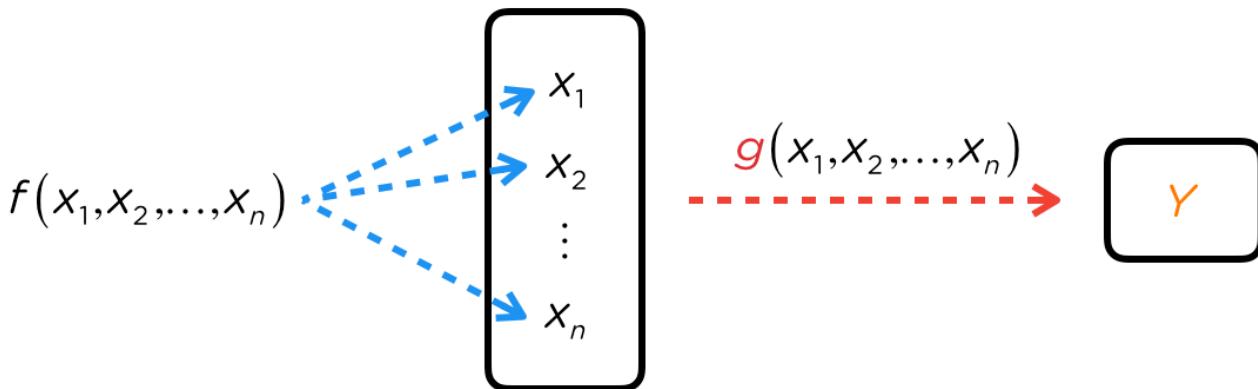
🕒 5m

Monte Carlo simulation is a simulation approach that involves repeated random sampling from a distribution to obtain a numerical estimate. The general methodology for a Monte Carlo simulation is as follows:

1. Generate a vector of random numbers, (X_1, X_2, \dots, X_n) , from a specified distribution, $f(x_1, x_2, \dots, x_n)$.
2. Apply a function g to the random numbers to compute Y , where $Y = g(x_1, x_2, \dots, x_n)$.
3. Repeat the first two steps r times. Make sure r is relatively large.
4. Compute any desired quantity (probability, mean, variance, etc.) using the generated Y 's.

Note that Monte Carlo simulation is a more generalized approach to simulation, as it does not specify how to generate random numbers from the given distribution. However, the methods covered previously are some of the possible ways to complete step 1 of Monte Carlo simulation.

The diagram below illustrates one iteration of the Monte Carlo simulation.



For instance, suppose we are interested in the probability of drawing two cards with a sum of at least 20 from a standard 52-card deck (where Ace = 11, King/Queen/Jack = 10). A simple example of a Monte Carlo simulation would be:

1. Generate two distinct random numbers from 1 to 52, where each number represents each card.

2. Calculate the sum of the card values. Define g such that if the sum is at least 20, then $Y = 1$, and if the sum is less than 20, then $Y = 0$.
3. Repeat the first two steps 1,000 times.
4. The estimated probability is the sum of the Y_i 's divided by 1,000.

1.8 Summary

 5m

Uniform Number Generation

To simulate uniform random numbers using pseudo random numbers,

1. Choose an initial value X_0 , as well as values for a , c , and m .
2. Generate pseudo random numbers using the following equation:

$$X_{n+1} = (aX_n + c) \bmod m, \quad n \geq 0$$

3. Divide the resulting value X_{n+1} by m to simulate a uniform random number on $[0, 1)$.

Inversion Method

Let U be a uniform $[0, 1)$ random number and F be the distribution function for a random number X . To simulate a random number X with distribution function F ,

1. Simulate a uniform random number U .
2. Solve for X as $X = F^{-1}(U)$.

Acceptance-Rejection Method

Let U be a uniform $[0, 1)$ random number, and let f and g be the density functions for two random variables. To simulate a random number Y with density function f ,

1. Find a constant c that satisfies:

$$\frac{f(x)}{g(x)} \leq c, \quad \text{for all } x$$

2. Simulate a random number Y with density function g , and simulate a random number U .
3. Calculate $\frac{f(Y)}{cg(Y)}$ for the simulated value Y .
4. Compare the simulated value U to $\frac{f(Y)}{cg(Y)}$. If $U \leq \frac{f(Y)}{cg(Y)}$, accept the value Y . Otherwise, reject it and return to step 2.

Monte Carlo Simulation

Monte Carlo simulation involves repeated random sampling from a distribution to obtain a numerical estimate.