

# Homework 1

PSTAT 160B: Stochastic Processes

Aarti Garaye

2025-01-17

## Part a

**Exercises:** 6.1-6.8, 6.10a, 6.12

**6.1** Let  $(N_t)_{t \geq 0}$  be a Poisson process with parameter  $\lambda = 1.5$ . Find the following:

- (a)  $P(N_1 = 2, N_4 = 6)$
- (b)  $P(N_4 = 6 | N_1 = 2)$
- (c)  $P(N_1 = 2 | N_4 = 6)$

**Solution:**

**(a)** We want to find the probability that 2 events occur in  $(0, 1]$  and 6 in  $(0, 4]$ . Then, This is essentially just saying that in  $(1, 4]$  there should be  $6 - 2 = 4$  events. Thus,

$$\{N_1 = 2, N_4 = 6\} = \{N_1 = 2, N_{4-1} = 4\}$$

So,

$$\mathbb{P}(\{N_1 = 2, N_4 = 6\}) = \mathbb{P}(\{N_1 = 2, N_{4-1} = 4\})$$

Note that  $N_1$  and  $N_{4-1}$  are independent, we can write the above probability as:

$$\begin{aligned}\mathbb{P}(\{N_1 = 2, N_4 = 6\}) &= \mathbb{P}(N_1 = 2)\mathbb{P}(N_{4-1} = 4) \\ &= \frac{e^{-1.5(1)}(1.5(1))^2}{2!} \cdot \frac{e^{-1.5(3)}(1.5(3))^4}{4!}\end{aligned}$$

We can solve this using the following code:

```
ans1a <- dpois(2, 1.5*1)*dpois(4, 1.5*3)
```

The desired probability is 0.0476458.

- (b)**  $\mathbb{P}(N_4 = 6 | N_1 = 2)$

Note, that if

$$\tilde{N} = N_{t+s} - N_s$$

and,

$$N_s = k$$

Then,

$$N_{t+s} - k \stackrel{d}{=} N_t$$

We know that  $N_1 = 2$ . Thus,  $s = 1$  and  $k = 2$ . Therefore,

$$\begin{aligned}\mathbb{P}(N_4 = 6|N_1 = 2) &= \mathbb{P}(N_{3+1} - 2 = 6 - 2|N_1 = 2) \\ &= \mathbb{P}(N_3 = 4) \\ &= \frac{e^{-1.5(3)}(1.5(3))^4}{4!}\end{aligned}$$

```
ans1b <- dpois(4,1.5*3)
```

The desired probability is 0.1898076.

(c)  $\mathbb{P}(N_1 = 2|N_4 = 6)$

Since the time of event B is after event A, using the Bayes' rule this probability could be written as:

$$\mathbb{P}(N_4 = 6|N_1 = 2) \cdot \frac{\mathbb{P}(N_1 = 2)}{\mathbb{P}(N_4 = 6)}$$

Note that  $\mathbb{P}(N_4 = 6|N_1 = 2)$  is just the probability from part b. Thus,

$$0.1898076 \cdot \frac{e^{-1.5(1)}(1.5(1))^2}{2!} \cdot \frac{6!}{e^{-1.5(4)}(1.5(4))^6}$$

```
ans1c <- ans1b * dpois(2,1.5*1)/dpois(6,1.5*4)
```

The desired probability is 0.2966309.

**6.2** Let  $(N_t)_{t \geq 0}$  be a Poisson process with parameter  $\lambda = 2$ . Find the following:

- (a)  $E(N_3 N_4)$
- (b)  $E(X_3 X_4)$
- (c)  $E(S_3 S_4)$

**Solution:**

(a)  $E(N_3 N_4) = ?$

Note that  $N_4 = N_3 + (N_4 - N_3)$ . Thus,

$$\begin{aligned} E[N_3 N_4] &= E[N_3 \cdot (N_3 + (N_4 - N_3))] \\ &= E[N_3^2 + N_3(N_4 - N_3)] \\ &= E[N_3^2] + E[N_3(N_4 - N_3)] \end{aligned}$$

Let's take this one step at a time. Note that:

$$E[N_t^2] = Var(N_t) + (E[N_t])^2 = \lambda t + (\lambda t)^2$$

Thus,

$$E[N_3^2] = 2 \cdot 3 + (2 \cdot 3)^2 = 6 + 36 = 42$$

And note that:

$$\begin{aligned} E[N_3(N_4 - N_3)] &= E[N_3] \cdot E[N_4 - N_3] \\ &= 6 \cdot 2(1) = 12 \end{aligned}$$

Thus,

$$\begin{aligned} E[N_3 N_4] &= E[N_3 \cdot (N_3 + (N_4 - N_3))] \\ &= 42 + 12 \\ &= 54 \end{aligned}$$

Thus, the desired expectation is 54.

(b)  $E(X_3 X_4) = ?$

The  $X_3$  and  $X_4$  refer to the interarrival times of the events. From previous sections in the textbook and from the lecture we know that  $X_i$  are independent of each other for  $i = 1, 2, \dots, n$ .

Furthermore, we also know that  $X_i$  are iid exponential random variables with parameter  $\lambda$ . Therefore,

$$E[X_3 X_4] = E[X_3] \cdot E[X_4]$$

Since they are exponentially distributed the  $E[X_i] = \frac{1}{\lambda}$ . Thus,

$$E[X_3 X_4] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

The desired expectation is  $\frac{1}{4}$ .

(c)  $E[S_3 S_4] = ?$

We know that  $S_n = X_1 + \dots + X_n$  for  $n = 1, 2, \dots$ . Thus,  $S_k$  is the time of the  $k$ th arrival. We also know from the textbook that  $S_n$  follows the Gamma Distribution where:

$$\mathbb{E}(S_n) = \frac{n}{\lambda} \text{ and } Var(S_n) = \frac{n}{\lambda^2}$$

Similar to part a, note that:

$S_3 = \sum_{i=0}^3 X_i$  and  $S_4 = \sum_{i=0}^4 X_i$ . Also note that:

$$\begin{aligned} S_4 &= \sum_{i=0}^3 X_i + X_4 \\ &= S_3 + X_3 \end{aligned}$$

So,

$$\begin{aligned} \mathbb{E}[S_3 S_4] &= \mathbb{E}[S_3 \cdot (S_3 + X_4)] \\ &= \mathbb{E}[S_3^2 + S_3 X_4] \end{aligned}$$

Note that  $S_3$  and  $X_4$  are independent of each other,

$$\mathbb{E}[S_3 S_4] = \mathbb{E}[S_3^2] + \mathbb{E}[S_3 X_4]$$

Taking it step by step. Note,

$$\begin{aligned} \mathbb{E}[S_t^2] &= Var(S_t) + (\mathbb{E}(S_t))^2 \\ \mathbb{E}[S_3^2] &= Var(S_3) + (\mathbb{E}(S_3))^2 \\ &= \frac{3}{2^2} + \left(\frac{3}{2}\right)^2 \\ &= \frac{3+9}{4} = \frac{12}{4} = 3 \end{aligned}$$

As discussed above  $S_3$  and  $X_4$  are independent, therefore, we can separate the expectation:

$$\begin{aligned} \mathbb{E}[S_3 X_4] &= \mathbb{E}[S_3] \cdot \mathbb{E}[X_4] \\ &= \frac{3}{2} \cdot \frac{1}{2} = \frac{3}{4} \end{aligned}$$

So,

$$\mathbb{E}[S_3 S_4] = 3 + \frac{3}{4} = \frac{12+3}{4} = \frac{15}{4}$$

The desired expectation is 3.75.

**6.3** Calls are received at a company call center according to a Poisson process at the rate of five calls per minute.

- (a) Find the probability that no call occurs over a 30-second period.
- (b) Find the probability that exactly four calls occur in the first minute, and six calls occur in the second minute.
- (c) Find the probability that 25 calls are received in the first 5 minutes and six of those calls occur in the first minute.

**Solution:**

(a)  $\mathbb{P}(\text{No calls over 30 seconds period}) = ?$

$$\mathbb{P}(N_{0.5} = 0)$$

$$\frac{e^{-5(0.5)}(5(0.5))^0}{0!}$$

ans3a <- dpois(0, 5\*0.5)

The probability that no calls are received over a 30 second period is 0.082085.

(b)  $\mathbb{P}(N_1 = 4 \cap N_2 = 6)$

This is just saying  $\mathbb{P}(N_1 = 4, N_2 = 6)$  which is equivalent to

$$\mathbb{P}(N_1 = 4, N_2 - N_1 = 6)$$

because the six calls that are occurring in the second minute are actually just six calls occurring in one total minute (60 seconds). Therefore,  $t = \text{length of the time}$  is just 1.

As we know  $N_1 \perp\!\!\!\perp N_2 - N_1$ . Thus,

$$\mathbb{P}(N_1 = 4, N_2 - N_1 = 6) = \mathbb{P}(N_1 = 4) \cdot \mathbb{P}(N_2 - N_1 = 6)$$

$$\frac{e^{-5(1)}(5(1))^4}{4!} \cdot \frac{e^{-5(1)}(5(1))^6}{6!}$$

ans3b <- dpois(4, 5\*1) \* dpois(6, 5\*1)

The probability that exactly four calls occur in the first minute, and six calls occur in the second minute is 0.0256573.

(c)  $\mathbb{P}(N_5 = 25 | N_1 = 6)$

Note that this is the same as the probability of 6 calls being received in the first minute and the remaining calls received in the next 4 minutes. Writing it mathematically,

$$\mathbb{P}(N_{4+1} - 6 = 25 - 6, N_1 = 6)$$

$$\mathbb{P}(N_4 = 19) \cdot \mathbb{P}(N_1 = 6)$$

$$\frac{e^{-5(4)}(5(4))^{19}}{19!} \cdot \frac{e^{-5(1)}(5(1))^6}{6!}$$

ans3c <- dpois(19, 5\*4) \* dpois(6, 5\*1)

The probability that 25 calls are received in the first 5 minutes and six of those calls occur in the first minute is 0.0129897.

**6.4** Starting 9 a.m., patients arrive at a doctor's office according to a Poisson process. On average, three patients arrive every hour.

- (a) Find the probability that at least two patients arrive by 9:30 a.m.
- (b) Find the probability that 10 patients arrive by noon and eight of them come to the office before 11 a.m.
- (c) If six patients arrive by 10 a.m., find the probability that only one patient arrives by 9:15 a.m.

**Solution:**

(a)  $\mathbb{P}(\text{at least 2 arrive by 9:30})$

Since the doctor's office starts at 9:00 am, 9:30 am is  $t = 0.5$  So we want,

$$\mathbb{P}(N_{0.5} \geq 2)$$

$$= 1 - \sum_{i=0}^2 \mathbb{P}(N_{0.5} = i)$$

```
ans4a <- 1 - ppois(2,3*0.5)
```

The probability that at least two patients arrive by 9:30 a.m. is 0.1911532.

(b)  $\mathbb{P}(10 \text{ patients by noon AND 8 of them come before 11 am})$

$$\begin{aligned} & \mathbb{P}(N_3 = 10, N_2 = 8) \\ &= \mathbb{P}(N_{1+2} - 8 = 10 - 8, N_2 = 8) \\ &= \mathbb{P}(N_1 = 2) \cdot \mathbb{P}(N_2 = 8) \\ &= \frac{e^{-3(1)}(3(1))^2}{2!} \cdot \frac{e^{-3(2)}(3(2))^8}{8!} \end{aligned}$$

```
ans4b <- dpois(2,3*1)*dpois(8,3*2)
```

The probability that 10 patients arrive by noon and eight of them come to the office before 11 a.m. is 0.023134.

(c)  $\mathbb{P}(N_{0.25} = 1 | N_1 = 6)$

Using the Bayes rule of conditional probability we can expand this as:

$$\begin{aligned} \mathbb{P}(N_{0.25} = 1 | N_1 = 6) &= \frac{\mathbb{P}(N_1 = 6 | N_{0.25} = 1) \cdot \mathbb{P}(N_{0.25} = 1)}{\mathbb{P}(N_1 = 6)} \\ &= \frac{\mathbb{P}(N_{0.75+0.25} - 1 = 6 - 1 | N_{0.25} = 1) \cdot \mathbb{P}(N_{0.25} = 1)}{\mathbb{P}(N_1 = 6)} \\ &= \frac{\mathbb{P}(N_{0.75} = 5) \cdot \mathbb{P}(N_{0.25} = 1)}{\mathbb{P}(N_1 = 6)} \\ &= \frac{\frac{e^{-3(0.75)}(3(0.75))^5}{5!} \cdot \frac{e^{-3(0.25)}(3(0.25))^1}{1!}}{\frac{e^{-3(1)}(3(1))^6}{6!}} \end{aligned}$$

```
ans4c <- (dpois(5,3*0.75)*dpois(1,3*0.25))/dpois(6,3*1)
```

The probability that if six patients arrive by 10 a.m., and that only one patient arrives by 9:15 a.m is 0.355957.

**6.5** Let  $(N_t)_{t \geq 0}$  be a Poisson process. Explain what is wrong with the following proof that  $N_3$  is a constant.

$$\begin{aligned} E((N_3)^2) &= E(N_3 N_3) = E(N_3(N_6 - N_3)) \\ &= E(N_3)E(N_6 - N_3) = E(N_3)E(N_3) \\ &= E(N_3)^2 \end{aligned}$$

Thus,  $\text{Var}(N_3) = E((N_3)^2) - E(N_3)^2 = 0$ , which gives that  $N_3$  is a constant with probability 1.

**Solution:**

The mistake in this proof is in the second line,  $N_3 \neq N_6 - N_3$ . Remember that there are not independent. It is however true that  $N_3 \stackrel{d}{=} N_6 - N_3$ . In a Poisson Process the interarrival times are independent not the number of events. The number of arrivals or occurrences of events in  $(3, 6]$  is not necessarily equal to the number of arrivals in  $(0, 3]$ .

The counterexample of  $X \stackrel{d}{=} Y \implies X^2 \stackrel{d}{=} XY$  is as follows:

Consider

$$X = \begin{cases} 1 & p = \frac{1}{2} \\ -1 & q = \frac{1}{2} \end{cases}$$

and that X and Y are iid random variables. Then,  $X^2 = 1$ , however,

$$XY = \begin{cases} 1 & p = \frac{1}{2} \\ -1 & q = \frac{1}{2} \end{cases}$$

Thus,  $X^2 \stackrel{d}{\neq} XY$ . To reiterate,  $N_3 \neq N_6 - N_3$ .

**6.6** Occurrences of landfalling hurricanes during an El Niño event are modeled as a Poisson process in Bove et al. (1998). The authors assert that “During an El Niño year, the probability of two or more hurricanes making landfall in the United States is 28%.” Find the rate of the Poisson process.

**Solution:**

From the information given we can say:

$$\begin{aligned} 0.28 &= \mathbb{P}(N_1 \geq 2) \\ &= 1 - \mathbb{P}(N_1 < 2) \end{aligned}$$

Note that:

$$\mathbb{P}(N_1 < 2) = \mathbb{P}(N_1 = 0) + \mathbb{P}(N_1 = 1)$$

Thus,

$$\begin{aligned} 0.28 &= 1 - (\mathbb{P}(N_1 = 0) + \mathbb{P}(N_1 = 1)) \\ &= 1 - \left( \frac{e^{-\lambda(1)}(\lambda(1))^0}{0!} + \frac{e^{-\lambda(1)}(\lambda(1))^1}{1!} \right) \\ &= 1 - e^{-\lambda} - e^{-\lambda}\lambda \end{aligned}$$

This gives,

$$e^{-\lambda}(1 + \lambda) = 0.72$$

To solve this, let's try graphing it and see where the line  $f(\lambda) = 0.72$  intersects the functions  $e^{-\lambda}(1 + \lambda)$  so that  $\lambda$  is a positive value.

```
f <- function(lambda) exp(-lambda) * (1 + lambda)

lambda_values <- seq(0, 2, by = 0.001)

y_values <- f(lambda_values)

plot(lambda_values, y_values, type = "l", col = "blue", lwd = 2,
      xlab = expression(lambda), ylab = expression(e^{-(lambda)} * (1 + lambda)),
      main = expression(e^{-(lambda)} * (1 + lambda) - 0.72))
abline(h = 0.72, col = "red", lty = 2)

f2 <- function(lambda) exp(-lambda) * (1 + lambda) - 0.72
root <- uniroot(f2, interval = c(0, 5))

points(root$root, 0.72, col = "black", pch = 19, cex = 1.5)
text(root$root, 0.75, labels = "1.04285", col = "black", pos = 4)

grid()
```

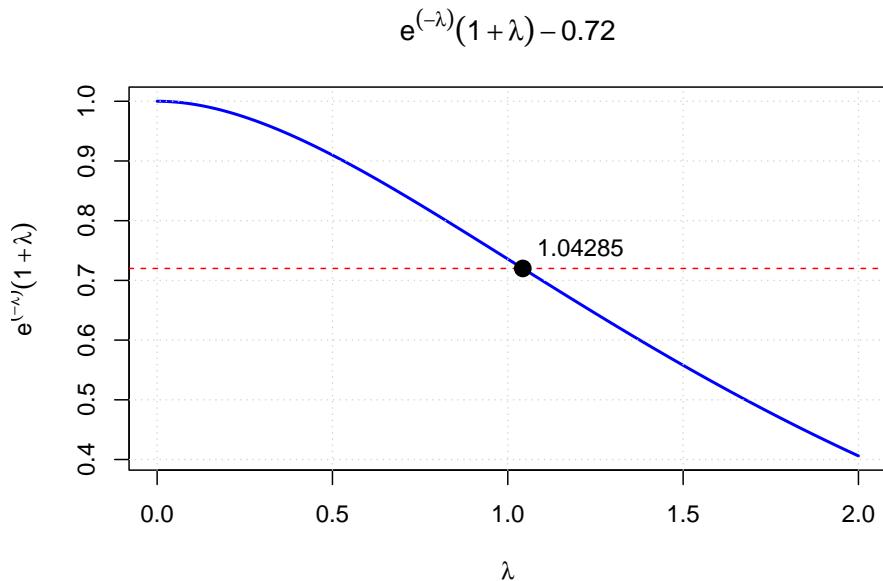


Figure 1: The above graph shows the rate of the Poisson process. The blue curve is the curve for the equation. The red dotted line is the probability line given from the question. The point where these two intersect, 1.04285, is the desired rate for the Poisson process.

This gives the solution that the rate of the Poisson process that models the occurrences of landfalling hurricanes during an El Niño event is 1.0428513.

**6.7** Ben, Max, Yolanda are at the front of three separate lines in the cafeteria waiting to be served. The serving times for the three lines follow independent Poisson processes with respective parameters 1, 2, and 3.

- (a) Find the probability that Yolanda is served first.
- (b) Find the probability that Ben is served before Yolanda.
- (c) Find the expected waiting time for the first person served.

**Solution:**

**(a)** Let  $X_B$ ,  $X_M$ , and  $X_Y$  be the waiting time of Ben, Max, and Yolanda respectively before getting first served.

For Yolanda to be the first one to be served, she must have the minimum of the wait time among all three of them. Let  $M = \min\{X_B, X_M, X_Y\}$

Furthermore, we know that the wait times for Ben, Max, and Yolanda are independent of each other as the serving times for the three lines follow **independent** Poisson processes. Thus, by the minimum of independent exponential random variable we know that:

$$\mathbb{P}(M = X_k) = \frac{\lambda_k}{\lambda_1 + \dots + \lambda_n}$$

where  $M$  is the minimum of  $X_1, \dots, X_n$ . So,

$$\mathbb{P}(M = X_Y) = \frac{3}{1+2+3} = \frac{3}{6} = \frac{1}{2}$$

The probability that Yolanda is served first is 0.50.

**(b)** For Ben to be served before Yolanda, Ben's wait time must be less than Yolanda's. Let  $M$  now be the minimum of  $X_B$  and  $X_Y$ . Therefore,

$$\mathbb{P}(\min(X_B, X_Y) = X_B) = \frac{1}{1+3} = \frac{1}{4}$$

The probability that Ben gets served before Yolanda is 0.25.

**(c)** The time of the first serve is the minimum of  $X_B$ ,  $X_M$ , and  $X_Y$  which is just a sum of independent random exponential variables. Thus, the minimum of  $X_B$ ,  $X_M$ , and  $X_Y$  has an exponential distribution with parameter  $1+2+3=6$ . So the expected value is:

$$\mathbb{E}(\min(X_B, X_M, X_Y)) = \frac{1}{1+2+3} = \frac{1}{6}$$

The expected wait time for the first person to be served is 0.1666667.

**6.8** Starting at 6 a.m., cars, buses, and motorcycles arrive at a highway toll booth according to independent Poisson processes. Cars arrive about once every 5 minutes. Buses arrive about once every 10 minutes. Motorcycles arrive about once every 30 minutes.

- (a) Find the probability that in the first 20 minutes, exactly three vehicles—two cars and one motorcycle—arrive at the booth.
- (b) At the toll booth, the chance that a driver has exact change is  $\frac{1}{4}$ , independent of vehicle. Find the probability that no vehicle has exact change in the first 10 minutes.
- (c) Find the probability that the seventh motorcycle arrives within 45 minutes of the third motorcycle.
- (d) Find the probability that at least one other vehicle arrives at the toll booth between the third and fourth car arrival.

**Solution:**

(a) Let  $(N_t)_{t \geq 0}, (C_t)_{t \geq 0}, (B_t)_{t \geq 0}, (M_t)_{t \geq 0}$  denote the overall vehicles, cars, buses, and motorcycle arrival at the toll booth respectively.

The desired probability is:

$$\mathbb{P}(C_{20} = 2, B_{20} = 0, M_{20} = 1)$$

Note that it says there is one car every five minute which means the rate of cars per minute is  $\frac{1}{5}$ th. Similarly, it's  $\frac{1}{10}$ th for buses and  $\frac{1}{30}$ th for motorcycles.

By the thinned Poisson Process, we can expand this as:

$$\begin{aligned}\mathbb{P}(C_{20} = 2, B_{20} = 0, M_{20} = 1) &= \mathbb{P}(C_{20} = 2) \cdot \mathbb{P}(B_{20} = 0) \cdot \mathbb{P}(M_{20} = 1) \\ &= \frac{e^{-\frac{1}{5}(20)}(\frac{1}{5}(20))^2}{2!} \cdot \frac{e^{-\frac{1}{10}(20)}(\frac{1}{10}(20))^0}{0!} \cdot \frac{e^{-\frac{1}{30}(20)}(\frac{1}{30}(20))^1}{1!}\end{aligned}$$

```
ans8a <- dpois(2,(1/5)*20) * dpois(0,(1/10)*20) * dpois(1,(1/30)*20)
```

The probability that in the first 20 minutes, exactly three vehicles—two cars and one motorcycle—arrive at the booth is 0.0067874.

(b) We want to find the probability that none of the vehicles have the change.

By the superposition theorem we know that the parameter for the overall vehicles would be  $\lambda = \lambda_c + \lambda_b + \lambda_m$ . Thus,

$$\lambda = \frac{1}{5} + \frac{1}{10} + \frac{1}{30} = \frac{6+3+1}{30} = \frac{10}{30} = \frac{1}{3}$$

We know that regardless of vehicles, the probability of a driver carrying change is  $\frac{1}{4}$ th. Therefore, the  $\lambda$  would be

$$\lambda p = \frac{1}{3} \cdot \frac{1}{4} = \frac{1}{12}$$

So, the probability of no one carrying exact change would be:

$$\mathbb{P}(N_{10} = 0) = \frac{e^{-\frac{1}{12}(10)}(\frac{1}{12}(10))^0}{0!}$$

```
ans8b <- dpois(0,(1/12)*10)
```

the probability that no vehicle has exact change in the first 10 minutes is 0.4345982.

**(c)** The probability of 7th motorcycle within 45 minutes of the 3rd motorcycle should be modeled with probability of arrival times denoted by  $S_k$  which is the arrival time of the  $k$ th event. Thus, the desired probability is:

$$\mathbb{P}(S_7 - S_3 < 45)$$

Note that

$$S_7 = S_3 + X_4 + X_5 + X_7 + X_7$$

where  $X_i$  is the interarrival times i.e. the time passed between the arrival of two events. Thus,

$$\begin{aligned} \mathbb{P}(S_7 - S_3 < 45) &= \mathbb{P}(S_3 + X_4 + X_5 + X_7 + X_7 - S_3 < 45) \\ &= \mathbb{P}(X_4 + X_5 + X_7 + X_7 < 45) \end{aligned}$$

Since the  $X_i$ 's are the interarrival times and are iid exponential distribution with parameter  $\lambda$ . Thus, as proved in class and in the textbook problem 6.10, the sum of iid exponential distribution follows a Gamma distribution with parameters  $n$  and  $\lambda$ . So,

$$\begin{aligned} \mathbb{P}(X_4 + X_5 + X_7 + X_7 < 45) &\sim \text{Gamma}(n, \lambda) \\ &= \mathbb{P}(X_4 + X_5 + X_7 + X_7 < 45) \sim \text{Gamma}(45, \frac{1}{30}) \end{aligned}$$

`ans8c <- pgamma(45, 4, (1/30))`

Therefore, the probability that the seventh motorcycle arrives within 45 minutes of the third motorcycle is 0.0656425.

**(d)** When they say other vehicle we want to find the probability that either a bus or a motorcycle arrives between the third and the fourth car arrival.

Let  $N_t^{oth}$  denote the number of buses plus the number of motorcycles. So the interarrival times of other vehicles,  $X_i^{oth}$ , would follow an exponential distribution.

$$\lambda^{oth} = \frac{1}{10} + \frac{1}{30} = \frac{3+1}{30} = \frac{2}{15}$$

So,  $X_i^{oth} \sim \text{Exp}(\lambda = \frac{2}{15})$

We know that the interarrival time of cars,  $X_i^c$  would also be an exponential distribution  $X_i^c \sim \text{Exp}(\lambda = \frac{1}{5})$ . So, the desired probability is:

$$\mathbb{P}(X_i^{oth} < N_1^c)$$

As discussed in the section, the formula for the above probability is:

$$\begin{aligned} \mathbb{P}(X_i^{oth} < N_1^c) &= \frac{\lambda^{oth}}{\lambda^{oth} + \lambda^c} \\ \mathbb{P}(X_i^{oth} < N_1^c) &= \frac{\frac{2}{15}}{\frac{2}{15} + \frac{1}{5}} = \frac{\frac{2}{15}}{\frac{5}{15}} = \frac{2}{5} \end{aligned}$$

Therefore, the probability that at least one other vehicle arrives at the toll booth between the third and fourth car arrival is 0.4.

**6.10** Assume that  $X_1, X_2, \dots$  is an i.i.d. sequence of exponential random variables with parameter  $\lambda$ . Let  $S_n = X_1 + \dots + X_n$ . Show that  $S_n$  has a gamma distribution with parameters  $n$  and  $\lambda$

(a) by moment-generating functions

**Solution:**

*Proof:* We want to show that  $S_n$ , which is the sum of the  $X_1, X_2, \dots$ , an i.i.d. sequence of exponential random variables with parameter  $\lambda$ , follows a gamma distribution with parameters  $n$  and  $\lambda$ . Remember that the mgf of a Gamma function looks like:

$$M_x(t) = \frac{1}{(1-\theta t)^\alpha}, t < \frac{1}{\theta}$$

Let's start by seeing what the mgf of  $X_i$  looks like. Remember that  $M_x(t) = \mathbb{E}[e^{tX}]$ . So,

$$M_x(X_1) = \mathbb{E}[e^{tX_1}]$$

Note that since  $X_1, X_2, \dots$  is an i.i.d. sequence of exponential random variables with parameter  $\lambda$ , the mgf for all  $X_i$ s is the same.

$$M_x(t) = \int_{-\infty}^{\infty} e^{tX_1} \cdot f_x(X_1) dx$$

Since it's an exponential random variable, the lower bound for the integral would be just zero instead of negative infinity.

$$\begin{aligned} M_x(t) &= \int_0^{\infty} e^{tx} \cdot \lambda e^{-\lambda x} dx \\ &= \lambda \int_0^{\infty} e^{(t-\lambda)x} dx \\ &= \lambda \left[ \frac{1}{t-\lambda} \cdot -e^{(t-\lambda)x} \right]_0^{\infty} \end{aligned}$$

Note that this is finite if and only if  $t < \lambda$

$$\begin{aligned} &= \lambda \left[ 0 + \frac{1}{\lambda-t} \right] \\ &= \frac{\lambda}{\lambda-t} \end{aligned}$$

Now that we have the mgf of  $X_i$ , let's try proving that  $S_n$  follows the gamma distribution. Let's start by taking the mgf of  $S_n$

$$M_x(S_n) = \mathbb{E}[e^{t(X_1+\dots+X_n)}]$$

Since  $X_i$ s are independent of each other we can write the expectation as:

$$\begin{aligned} M_x(S_n) &= \mathbb{E}[e^{t(X_1)}] + \mathbb{E}[e^{t(X_2)}] + \dots + \mathbb{E}[e^{t(X_n)}] \\ &= M_x(X_1) + M_x(X_2) + \dots + M_x(X_n) \\ &= \frac{\lambda}{\lambda-t} + \frac{\lambda}{\lambda-t} + \dots + \frac{\lambda}{\lambda-t} \\ &= \left( \frac{\lambda}{\lambda-t} \right)^n \end{aligned}$$

This is the mgf of the gamma distribution. Thus  $S_n \sim \text{Gamma}(n, \lambda)$  ■

**6.12** Starting at noon, diners arrive at a restaurant according to a Poisson process at the rate of five customers per minute. The time each customer spends eating at the restaurant has an exponential distribution with mean 40 minutes, independent of other customers and independent of arrival times. Find the distribution, as well as the mean and variance, of the number of diners in the restaurant at 2 p.m.

**Solution:**

Let  $D_t$  denote the number of diners at the restaurant at time  $t$ . Conditioning on  $N_t$ ,

$$\begin{aligned}\mathbb{P}(D_t = k) &= \sum_{n=k}^{\infty} \mathbb{P}(D_t = k | N_t = n) \mathbb{P}(N_t = n) \\ &= \sum_{n=k}^{\infty} \mathbb{P}(D_t = k | N_t = n) \frac{e^{-\lambda t} (\lambda t)^n}{n!}\end{aligned}$$

Assume that  $n$  diners enter the building by the time  $t$ , with arrival times  $S_1, \dots, S_n$ . Let  $Z_k$  be the length of time spent in the building by the  $k$ th diner, for  $1 \leq k \leq n$ . Then,  $Z_1, \dots, Z_n$  are iid random variables with cdf  $F$ , which we know from the question is exponentially distributed with mean 40. Diners leaving the restaurant then at times  $S_1 + Z_1, \dots, S_n + Z_n$ . There are  $k$  diners in the restaurant at time  $t$  if and only if  $k$  of the departure times  $S_1 + Z_1, \dots, S_n + Z_n$  exceed time  $t$ . This gives,

$$\begin{aligned}\mathbb{P}(D_t = k | N_t = n) &= \mathbb{P}(k \text{ of the } S_1 + Z_1, \dots, S_n + Z_n \text{ exceed } t | N_t = n) \\ &= \mathbb{P}(k \text{ of the } U_{(1)} + Z_1, \dots, U_{(n)} + Z_n \text{ exceed } t) \\ &= \mathbb{P}(k \text{ of the } U_1 + Z_1, \dots, U_n + Z_n \text{ exceed } t) \\ &= \binom{n}{k} p^k (1-p)^{n-k}\end{aligned}$$

where,

$$p_t = \mathbb{P}(U_1 + Z_1 > t) = \frac{1}{t} \int_0^t \mathbb{P}(Z_1 > t - x) dx = \frac{1}{t} \int_0^t [1 - F(x)] dx$$

We know  $F(x)$  is the cdf of the exponential function with mean 40. Therefore,

$$\begin{aligned}p_t &= \int_0^t e^{-\frac{x}{40}} dx \\ &= -40 \int_0^t \frac{-1}{40} e^{-\frac{x}{40}} dx \\ &= -40 e^{-\frac{x}{40}} \Big|_0^t \\ &= -40 e^{-\frac{t}{40}} + 40 \\ &= 40(1 - e^{-\frac{t}{40}})\end{aligned}$$

Thus, at 2 p.m. the distribution of number of diners present at the restaurant follow a Poisson Distribution with parameter  $\lambda = 5$ . Therefore, the mean is  $\lambda p$ . Note that the restaurant opens at noon. Which means that 2 p.m. is 2 hours which is equivalent to 120 minutes. Thus,

$$p_{120} = 40(1 - e^{-\frac{120}{40}})$$

```
p120 <- 40 * (1 - exp(-120 / 40))
```

The associated probability is 38.0085173. Thus mean would be  $\lambda(p)$  which is  $5 * 38.0085$ .

```
mean12 <- 5 * p120
```

The mean is 190.0425863 With a variance of 190.0425863 as well. This means that the standard deviation is 13.7855934. Therefore, at 2 p.m. there are on an average 190 diners at the restaurant give of take 14 people.

## Part b

**Exercises:** 6.13, 6.15, 6.18, 6.20, 6.21, 6.28, 6.33, 6.35, 6.41, 6.43

**6.13** Assume that  $(N_t)_{t \geq 0}$  is a Poisson process with parameter  $\lambda$ . Find the conditional distribution of  $N_s$  given  $N_t = n$ , for

- (a)  $s < t$ ,
- (b)  $s > t$ .

**Solution:**

**(a)** The desired probability statement is where  $s < t$

$$\mathbb{P}(N_s = k | N_t = n) = \frac{\mathbb{P}(N_s = k, N_t = n)}{\mathbb{P}(N_t = n)}$$

In this notation we cannot separate the conditional probability. However, note that we can write the above statement as:

$$\mathbb{P}(N_s = k | N_t = n) = \frac{\mathbb{P}(N_s = k) \cdot \mathbb{P}(N_t - N_s = n - k)}{\mathbb{P}(N_t = n)}$$

because  $N_s$  and  $N_t - N_s$  are independent. Expanding it out,

$$\begin{aligned} \mathbb{P}(N_s = k | N_t = n) &= \frac{\frac{e^{-\lambda s} (\lambda s)^k}{k!} \cdot \frac{e^{-\lambda(t-s)} (\lambda(t-s))^{n-k}}{(n-k)!}}{\frac{e^{-\lambda t} (\lambda t)^n}{n!}} \\ &= \frac{e^{-\lambda s} (\lambda s)^k}{k!} \cdot \frac{e^{-\lambda(t-s)} (\lambda(t-s))^{n-k}}{(n-k)!} \cdot \frac{n!}{e^{-\lambda t} (\lambda t)^n} \\ &= \frac{n!}{k!(n-k)!} \cdot e^{-\lambda s - \lambda t + \lambda s + \lambda t} \cdot \frac{(\lambda s)^k (\lambda(t-s))^{n-k}}{(\lambda t)^n} \end{aligned}$$

Note that  $e^{-\lambda s - \lambda t + \lambda s + \lambda t} = e^0 = 1$  and that  $\frac{n!}{k!(n-k)!}$  is just  $\binom{n}{k}$ . Simplifying further,

$$\begin{aligned} \mathbb{P}(N_s = k | N_t = n) &= \binom{n}{k} \cdot \frac{\lambda^k s^k \cdot \lambda^{n-k} (t-s)^{n-k}}{\lambda^n t^n} \\ &= \binom{n}{k} \cdot \frac{\lambda^{k+n-k-n} \cdot s^k (t-s)^{n-k}}{t^n} \end{aligned}$$

Note that  $\lambda^{k+n-k-n} = \lambda^0 = 1$ . Also note that  $k \leq n$ . Further factoring,

$$\begin{aligned} \mathbb{P}(N_s = k | N_t = n) &= \binom{n}{k} \cdot \left(\frac{s}{t}\right)^k \cdot \frac{(t-s)^{n-k}}{t^{n-k}} \\ &= \binom{n}{k} \cdot \left(\frac{s}{t}\right)^k \cdot \left(\frac{t-s}{t}\right)^{n-k} \\ &= \binom{n}{k} \cdot \left(\frac{s}{t}\right)^k \cdot \left(1 - \frac{s}{t}\right)^{n-k} \end{aligned}$$

Thus, notice that  $\mathbb{P}(N_s = k | N_t = n) \sim \text{Binom}(n, p = \frac{s}{t})$  when  $s < t$ .

**(b)** The desired conditional probability for  $s > t$  and for  $k \geq n$  is:

$$\mathbb{P}(N_s = k | N_t = n) = \mathbb{P}(N_{s-t} = k - n)$$

We can just solve this using the Poisson probability formula

$$\mathbb{P}(N_s = k | N_t = n) = \frac{e^{-\lambda(s-t)} (\lambda(s-t))^{k-n}}{(k-n)!}$$

Thus,  $\mathbb{P}(N_s = k | N_t = n) \sim \text{Pois}(\lambda(s-t))$  it's just shifted by  $(s-t)$ .

**6.15** Failures occur for a mechanical process according to a Poisson process. Failures are classified as either major or minor. Major failures occur at the rate of 1.5 failures per hour. Minor failures occur at the rate of 3.0 failures per hour.

- (a) Find the probability that two failures occur in 1 hour.
- (b) Find the probability that in half an hour, no major failures occur.
- (c) Find the probability that in 2 hours, at least two major failures occur or at least two minor failures occur.

**Solution:**

(a) Let  $N_t^M$  be the number of major failures with parameter  $\lambda^M = 1.5$  and let  $N_t^N$  be the minor failures with parameter  $\lambda^N = 3$  and let  $N_t$  be the overall failures.

We want to find the probability that two failures occur in 1 hour.

$$\mathbb{P}(N_1 = 2)$$

By the thinning Poisson process we know that the overall parameter is just the sum of  $k$  types of bifurcations, which in this case is two. So,

$$\lambda = \lambda^M + \lambda^N = 1.5 + 3 = 4.5$$

Thus,

$$\mathbb{P}(N_1 = 2) = \frac{e^{-4.5(1)}(4.5(1))^2}{2!}$$

```
ans15a <- dpois(2, 4.5*1)
```

The probability that two failures occur in 1 hour is 0.1124786.

- (b) The desired probability is defined as:

$$\mathbb{P}(N_{0.5}^M = 0) = \frac{e^{-1.5(0.5)}(1.5(0.5))^0}{0!}$$

```
ans15b <- dpois(0, 1.5*0.5)
```

The probability that no major failures occur in half an hour is 0.4723666.

- (c) The desired probability can be mathematically be represented as:

$$\mathbb{P}(N_2^M \geq 2) \cup \mathbb{P}(N_2^N \geq 2) = \mathbb{P}(N_2^M \geq 2) + \mathbb{P}(N_2^N \geq 2) - \mathbb{P}(N_2^M \geq 2 \cap N_2^N \geq 2)$$

Since failures can either be major or minor, they can never be both. Thus, major and minor failures are disjoint events. Therefore,  $\mathbb{P}(N_2^M \geq 2 \cap N_2^N \geq 2) = 0$  So,

$$\begin{aligned} \mathbb{P}(N_2^M \geq 2) \cup \mathbb{P}(N_2^N \geq 2) &= \mathbb{P}(N_2^M \geq 2) + \mathbb{P}(N_2^N \geq 2) \\ &= 1 - \mathbb{P}(N_2^M \leq 1) \cdot \mathbb{P}(N_2^N \leq 1) \\ &= 1 - \frac{e^{-1.5(2)}(1.5(2))^1}{1!} \cdot \frac{e^{-3(2)}(3(2))^1}{1!} \end{aligned}$$

```
ans15c <- 1 - (dpois(1, 1.5*2)* dpois(1, 3*2))
```

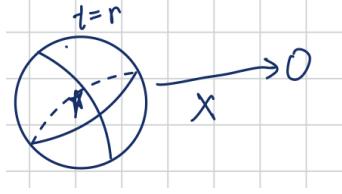
The probability that in 2 hours, at least two major failures occur or at least two minor failures occur is 0.9977786.

**6.18** The planets of the Galactic Empire are distributed in space according to a spatial Poisson process at an approximate density of one planet per cubic parsec. From the Death Star, let  $X$  be the distance to the nearest planet.

- (a) Find the probability density function of  $X$ .
- (b) Find the mean distance from the Death Star to the nearest planet.

**Solution:**

(a) Consider the Poisson process in  $\mathbb{R}^3$ . Let  $\star$  denote the Death Star, that is fixed. Let  $X$  be the distance from  $\star$  to its nearest planet. The event  $\{X > t\}$  occurs if and only if there are no points in the sphere centered at  $\star$  of radius  $t$ . Let  $C_x$  denote such a sphere. Refer to the picture below for reference.



Then:

$$\mathbb{P}(X > t) = \mathbb{P}(N_{C_x} = 0) = e^{\lambda|C_x|} = e^{-\lambda \frac{4}{3}\pi t^3}, \text{ for } t > 0$$

Since  $\lambda = 1$ ,

$$\mathbb{P}(X > t) = e^{-\frac{4}{3}\pi t^3}$$

To find the distribution of  $X$  we would have to integrate:

$$\begin{aligned} & \int_0^\infty \mathbb{P}(X > t) \\ &= \int_0^\infty e^{-\frac{4}{3}\pi t^3} dt \end{aligned}$$

Thus,

$$f_x(t) = 4\pi t^2 \lambda e^{-\lambda \frac{4}{3}\pi t^3}$$

$$f_x(t) = 4\pi t^2 e^{-\frac{4}{3}\pi t^3}, \text{ for } t > 0$$

(b) To find the mean distance from the  $\star$ , we would have to take the expected value:

$$\begin{aligned} \mathbb{E}[X] &= \int_0^\infty t f_x(t) dt \\ &= \int_0^\infty t(4\pi t^2 e^{-\frac{4}{3}\pi t^3}) dt \end{aligned}$$

Using an online calculator (Wolfram Alpha), the integral solves to 0.5539603. Thus, the mean distance of the nearest planet from then  $\star$  is 0.5539603.

**(6.20)** Consider a spatial point process in  $\mathbb{R}^2$  with parameter  $\lambda$ . Assume that  $A$  is a bounded set in  $\mathbb{R}^2$  which contains exactly one point of the process. Given  $B \subseteq A$ , find the probability that  $B$  contains one point.

**Solution:**

We want to find the probability of set  $B$  containing one point, given that set  $A$  has one point. It can be written as:

$$\begin{aligned}\mathbb{P}(N_B = 1 | N_A = 1) &= ? \\ &= \frac{\mathbb{P}(N_B = 1, N_A = 1)}{\mathbb{P}(N_A = 1)}\end{aligned}$$

Note that saying  $N_A = 1$  is the same as  $N_{ABC} = 0$ . Substituting:

$$\begin{aligned}\mathbb{P}(N_B = 1 | N_A = 1) &= \frac{\mathbb{P}(N_B = 1, N_{ABC} = 0)}{\mathbb{P}(N_A = 1)} \\ &= \frac{\frac{e^{-\lambda|B|}(\lambda|B|)^1}{1!} \cdot \frac{e^{-\lambda|AB^C|}(\lambda|AB^C|)^0}{0!}}{\frac{e^{-\lambda|A|}(\lambda|A|)^1}{1!}}\end{aligned}$$

Note that  $AB^C = |A| - |B|$ ,

$$\begin{aligned}\mathbb{P}(N_B = 1 | N_A = 1) &= \frac{\frac{e^{-\lambda|B|}(\lambda|B|)^1}{1!} \cdot \frac{e^{-\lambda(|A|-|B|)}(\lambda|AB^C|)^0}{0!}}{\frac{e^{-\lambda|A|}(\lambda|A|)^1}{1!}} \\ &= \frac{e^{-\lambda|B|-\lambda|A|+\lambda|B|+\lambda|A|}(\lambda|B|)}{\lambda|A|} \\ &= \frac{e^0 \lambda|B|}{\lambda|A|} \\ &= \frac{|B|}{|A|}\end{aligned}$$

**6.21** For a Poisson process with parameter  $\lambda$  show that for  $s < t$ , the correlation between  $N_s$  and  $N_t$  is

$$\text{Corr}(N_s, N_t) = \sqrt{\frac{s}{t}}.$$

**Solution:**

*Proof* Starting with the proven Correlation formula in PSTAT 120B,

$$\begin{aligned}\text{Corr}(N_s, N_t) &= \frac{\text{Cov}(N_s, N_t)}{\sqrt{\text{Var}(N_s)\text{Var}(N_t)}} \\ &= \frac{\mathbb{E}(N_s N_t) - \mathbb{E}(N_s)\mathbb{E}(N_t)}{\sqrt{\lambda s \lambda t}} \\ &= \frac{\mathbb{E}(N_s N_t) - \lambda s \lambda t}{\sqrt{\lambda s \lambda t}}\end{aligned}$$

Note that  $s < t$ . We can, therefore, write  $N_t = N_s + (N_t - N_s)$

$$\text{Corr}(N_s, N_t) = \frac{\mathbb{E}[N_s(N_s + (N_t - N_s))] - \lambda s \lambda t}{\sqrt{\lambda s \lambda t}}$$

Also note that  $\mathbb{E}[N_s^2] = \text{Var}(N_s) + (\mathbb{E}(N_s))^2 = \lambda s + (\lambda s)^2$

$$\begin{aligned}\text{Corr}(N_s, N_t) &= \frac{\lambda s + (\lambda s)^2 + \lambda s(\lambda(t-s)) - \lambda s \lambda t}{\sqrt{\lambda s \lambda t}} \\ &= \frac{\lambda s + (\lambda s)^2 + \lambda s \lambda t - (\lambda s)^2 - \lambda s \lambda t}{\sqrt{\lambda s} \sqrt{\lambda t}} \\ &= \frac{\lambda s}{\sqrt{\lambda s} \sqrt{\lambda t}} \\ &= \frac{(\lambda s)^{1-\frac{1}{2}}}{\sqrt{\lambda t}} \\ &= \frac{\sqrt{\lambda s}}{\sqrt{\lambda t}} \\ &= \frac{\sqrt{\lambda} \sqrt{s}}{\sqrt{\lambda} \sqrt{t}} \\ &= \sqrt{\frac{s}{t}}\end{aligned}$$

Thus, the above proof shows  $\text{Corr}(N_s, N_t) = \sqrt{\frac{s}{t}}$  ■

**6.28** Tornadoes hit a region according to a Poisson process with  $\lambda = 2$ . The number of insurance claims filed after any tornado has a Poisson distribution with mean 30. The number of tornadoes is independent of the number of insurance claims. Find the expectation and standard deviation of the total number of claims filed by time  $t$ .

**Solution:**

We are given that Tornadoes hit a region according to a Poisson process with  $\lambda = 2$  and that the number of insurance claims filed after any tornado has a Poisson distribution with mean 30. Furthermore we also know that the number of tornadoes are independent from the number of insurance claims. We want to find the  $\mathbb{E}[\text{number of claims at time } t] = ?$  and the  $SD[\text{number of claims at time } t] = ?$ .

Let  $X_1, X_2, \dots, X_n$  be the i.i.d random variable where  $X_i$  is the number of claims after the  $i^{th}$  tornado hit with mean 30 at time  $t$ . Let  $T_t$  be the total number of claims at time  $t$ . Then,

$$T_t = \sum_{i=1}^{N_t} X_i$$

We want to find  $\mathbb{E}[T_t]$  and  $SD[T_t]$ .

$$\mathbb{E}[T_t] = \mathbb{E}\left[\sum_{i=1}^{N_t} X_i\right]$$

We know that  $X_i$ s are i.i.d random variables. Thus,  $\mathbb{E}[T_t] = \mathbb{E}[N_t] \cdot \mathbb{E}[X_i]$  from conditioning on  $N_t$  since we don't know the number of tornadoes. We are using the fact the sums of random variables is a random variable. Thus,

$$\mathbb{E}[T_t] = 2t \cdot 30 = 60t$$

Moving on to the standard deviation

$$SD[T_t] = \sqrt{Var(T_t)}$$

So let's start by finding the variance first:

$$\begin{aligned} Var(T_t) &= Var\left(\sum_{i=1}^{N_t} X_i\right) \\ &= Var(X_1)\mathbb{E}[N_t] + (\mathbb{E}[X_1])^2 Var(N_t) \\ &= 30(2t) + 30^2(2t) = 60t + 1800t = 1860t \end{aligned}$$

Thus, the standard deviation is:

$$SD(T_t) = \sqrt{1860t}$$

The expected value of the total number of claims filed after the tornado is  $60t$  with the standard deviation of  $43.1277173\sqrt{t}$ .

**6.33** Let  $S_1, S_2, \dots$  be the arrival times of a Poisson process with parameter  $\lambda$ . Given the time of the  $n$ th arrival, find the expected time  $\mathbb{E}[S_1|S_n]$  of the first arrival.

**Solution:**

Intuition: Just by looking at the question if I think that I receive 10 calls in 10 minutes then the expected time of arrival of the first call should be in the first minute. Same logic should be applied here. Let's try to solve. We know:

$$S_n = X_1 + X_2 + \dots + X_n$$

where  $X_i$ s are the interarrival times i.e. the time passed in between for the next arrival. We can, therefore, say that the time of first arrival,  $S_1$ , is the time that passed in between,  $X_1$ .

Furthermore, from our intuition, the expected time  $\mathbb{E}[S_1|S_n]$  of the first arrival must be a function of  $S_n$ .

For a random variable  $X$  and  $Y$ , consider the conditional pdf  $p(x|y)$  and the conditional expectation  $\mathbb{E}[x|y]$ . Then,

$$\mathbb{E}[x|y] = \int_{-\infty}^{\infty} x \cdot p(x|y) dx$$

is a function of  $y$  since  $x$  is being integrated. Hence, when taking the expectation again on  $\mathbb{E}[x|y]$ , we write:

$$\mathbb{E}[\mathbb{E}[x|y]] = \int_{y=-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} x \cdot p(x|y) dx \right] p(y) dy$$

Therefore, consider  $\mathbb{E}[S_1|S_n] = g(S_n)$  which by the above statements must be an expression of  $S_n$ , since that's what we are conditioning our expectation on. Note that  $S_1 = X_1$ . So we can write,

$$\mathbb{E}[S_1|S_n] = \mathbb{E}[X_1|S_n]$$

We know that  $X_i$  are interarrival times that are iid exponential random variable. Thus,

$$\mathbb{E}[X_1|S_n] = \mathbb{E}[X_2|S_n] = \dots = \mathbb{E}[X_n|S_n] = g(S_n)$$

Furthermore, note that:

$$\mathbb{E}[X_1|S_n] + \mathbb{E}[X_2|S_n] + \dots + \mathbb{E}[X_n|S_n] = \mathbb{E}[X_1 + X_2 + \dots + X_n|S_n] = \mathbb{E}[S_n|S_n]$$

Conditional expectation of something on itself is just itself. Thus,

$$\mathbb{E}[S_n|S_n] = S_n$$

Let's call

$$n \cdot g(S_n) = S_n$$

Thus,

$$g(S_n) = \frac{S_n}{n}$$

So,

$$\mathbb{E}[X_1|S_n] = \frac{S_n}{n}$$

$$\mathbb{E}[S_1|S_n] = \frac{S_n}{n}$$

Another way of showing the same thing would be:

$$\mathbb{E}[X_1 + X_2 + \dots + X_n|S_n] = \sum_{i=1}^n \mathbb{E}[X_i|S_n] = \sum_{i=1}^n \mathbb{E}[X_1|S_n] = n\mathbb{E}[X_1|S_n]$$

Since we know  $S_1 = X_1$ , it gives

$$\mathbb{E}[S_1|S_n] = \mathbb{E}[X_1|S_n] = \frac{S_n}{n}$$

**6.35** See the definitions for the spatial and nonhomogeneous Poisson processes. Define a nonhomogeneous, spatial Poisson process in  $\mathbb{R}^2$ . Consider such a process  $(N_A)_{A \subseteq \mathbb{R}^2}$  with intensity function

$$\lambda(x, y) = e^{-(x^2+y^2)}, \text{ for } -\infty < x, y < \infty.$$

Let  $C$  denote the unit circle, that is, the circle of radius 1 centered at the origin. Find  $P(N_C = 0)$

**Solution:**

$(N_t)_{t \geq 0}$  is a nonhomogeneous Poisson process with intensity  $\lambda(x, y)$  in  $\mathbb{R}^n$  and a spatial Poisson process, then:

- (i)  $N_0 = 0$
- (ii) For all  $(N_A)_{A \subseteq \mathbb{R}^2} A > 0, N_t$

$$\mathbb{E}[N_A] = \int \int_0^t \lambda(x, y) dx dy$$

Combining the two definitions of spatial Poisson process and a nonhomogeneous Poisson process is:

A collection of random variables  $(N_A)_{A \subseteq \mathbb{R}^2}$  is a nonhomogeneous spatial Poisson process in  $\mathbb{R}^2$  with the intensity function  $\lambda(x, y)$  if:

- (i)  $N_0 = 0$
- (ii) For each bounded set  $A \subseteq \mathbb{R}^2$ ,  $N_A$  has a Poisson distribution with parameter  $\lambda|A|$ . Note that whenever  $A$  and  $B$  are disjoint sets,  $N_A$  and  $N_B$  are independent random variables.
- (iii) For all  $t > 0$ ,  $N_t$ , has Poisson Distribution with mean:

$$\mathbb{E}[N_t] = \int \int_0^t \lambda(x, y) dx dy$$

For  $0 \leq q < r \leq s < t$ ,  $N_r - N_q \perp\!\!\!\perp N_t - N_s$  random variables.

Let  $C$  be a unit circle in  $\mathbb{R}^2$ . Since  $A \subseteq \mathbb{R}^2$  by extension  $C \subset A$ . Thus,  $C \subset \mathbb{R}^2$  since  $C$  is a unit circle in  $\mathbb{R}^2$ . By the definition of spatial Poisson process, we know for each bounded set  $C \subseteq \mathbb{R}^2$ ,  $N_C$  has a Poisson distribution with parameter  $\lambda|C|$ , where  $|C|$  is the area of the unit circle. Thus,

$$\lambda|C| = \lambda(\pi r^2) = \lambda(\pi(1)^2) = \lambda\pi$$

Then,

$$\mathbb{P}(N_C = 0) = e^{- \int \int_C e^{-(x^2+y^2)} dx dy}$$

Since  $C$  is a unit circle, we know  $x^2 + y^2 = 1$ , therefore, the bounds of integration are as follows:

$$\mathbb{P}(N_C = 0) = e^{- \int_{-1}^1 \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} e^{-(x^2+y^2)} dx dy}$$

From the integration calculator, we get:

$$e^{- \int_{-1}^1 \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} e^{-(x^2+y^2)} dx dy} = e^{-1.98587} = 0.137261145$$

Thus, the probability that the unit circle  $C$  contains zero points in  $\mathbb{R}^2$  is 0.137261145.

**6.41 R :** Goals occur in a soccer game according to a Poisson process. The average total number of goals scored for a 90-minute match is 2.68. Assume that two teams are evenly matched. Use simulation to estimate the probability both teams will score the same number of goals. Compare with the theoretical results.

**Solution:**

For this question we will compute the theoretical result first and then move on to the simulation parts.

Let there be team A and team B that are playing this match. We are given that both A and B are equally matched in their skill set, so the total average of goals A and B score as individual teams should be the same. Therefore,

$$\lambda_A = \lambda_B = \frac{2.68}{2} \text{ per 90 minutes} = 1.34 \text{ goals per 90 minutes.}$$

Desired probability is that number of goals for team A is the same as number of goals scored by team B.

$$\mathbb{P}(\text{same goals}) = \mathbb{P}(N_A = k \cap N_B = k)$$

where  $k$  is the number of goals scored in a 90-minute match.

Note that  $N_A$  and  $N_B$  are independent of each other. The two teams' goals are measured as independent Poisson random variable. Therefore,

$$\mathbb{P}(N_A = k \cap N_B = k) = \mathbb{P}(N_A = k) \cdot \mathbb{P}(N_B = k)$$

Consider the first game i.e. the first 90-minute period. Then,

$$\mathbb{P}(N_A = k) \cdot \mathbb{P}(N_B = k) = \frac{e^{-\lambda}(\lambda)^k}{k!} \cdot \frac{e^{-\lambda}(\lambda)^k}{k!}$$

Note that the  $\lambda$  here for both are the same since A and B are evenly matched.

$$\frac{e^{-\lambda}(\lambda)^k}{k!} \cdot \frac{e^{-\lambda}(\lambda)^k}{k!} = \frac{e^{-2\lambda}(\lambda)^{2k}}{k!^2} = \left( \frac{e^{-\lambda}(\lambda)^k}{k!} \right)^2$$

Since we are simulating this for tens and thousands of matches, the probability is just the sum of all the probabilities as  $k = 0, 1, 2, \dots$

$$\mathbb{P}(N_A = k) \cdot \mathbb{P}(N_B = k) = \sum_{k=0}^{\infty} \left( \frac{e^{-\lambda}(\lambda)^k}{k!} \right)^2$$

We know  $\lambda = 1.34$ , thus,

$$\mathbb{P}(N_A = k) \cdot \mathbb{P}(N_B = k) = \sum_{k=0}^{\infty} \left( \frac{e^{-1.34}(1.34)^k}{k!} \right)^2$$

Unfortunately, this sum doesn't expand/converge to anything. If the summation wasn't to the power of 2, we would've used the Taylor series expansion to solve the series. Thus, we move on to simulation:

```
set.seed(123)
num_trails <- 100000000
ans41 <- mean(rpois(num_trails, 1.34)==rpois(num_trails, 1.34))
```

The estimated probability both teams will score the same number of goals is 0.2592821.

**6.43 R :** Simulate a spatial Poisson process with  $\lambda = 10$  on the box of volume 8 with vertices at the 8 points  $(\pm 1, \pm 1, \pm 1)$ . Estimate the mean and variance of the number of points in the ball centered at the origin of radius 1. Compare with the exact values.

**Solution:**

Since we are dealing in a  $\mathbb{R}^3$  space, we would have to consider the volume of the ball which is essentially a sphere with radius 1.

The box with volume 8 with vertices at points  $(\pm 1, \pm 1, \pm 1)$  is a bounded object on  $\mathbb{R}^3$ . Let  $C_x$  denote the ball (sphere) at the origin of radius 1. Then,

$$\lambda = \lambda |C_x|$$

where  $|C_x|$  is the volume of the ball. Thus,

$$\lambda = 10 \left( \frac{4}{3} \pi r^3 \right), \text{ where } r = 1$$

$$\lambda = 10 \left( \frac{4}{3} \pi 1^3 \right) = \frac{40\pi}{3} = 41.8879020479$$

This is also the mean and the variance of the number of points in the ball centered at the origin of radius 1.  
Mean and Variance:

$$\text{Mean and Variance} = \lambda \cdot |C_x| \cdot \text{volume of the box} \cdot \frac{1}{8} = 10 \cdot \left( \frac{4}{3} \pi 1^3 \right) \cdot 8 \cdot \frac{1}{8} = \frac{40\pi}{3} = 41.8879020479$$

Consider this code to follow the simulation for the estimated mean and variance for the number of points in the ball centered at the origin of radius 1:

```
set.seed(123)
num_trails2 <- 100000
lambda <- 10
box_vol <- 8
r <- 1
ball_vol <- (4/3) * pi * r^3

sim_points_in_ball <- function(){
  npoints <- rpois(1, lambda * box_vol)

  x <- runif(npoints, -1, 1)
  y <- runif(npoints, -1, 1)
  z <- runif(npoints, -1, 1)

  sum(x^2 + y^2 + z^2 <= r^2)
}

counts <- replicate(num_trails2, sim_points_in_ball())

mean43 <- mean(counts)
var43 <- var(counts)
```

From the above code chunk, we know that the estimated mean for the number of points in the ball centered at origin with radius 1 is 41.93038 and the estimated variance was 42.0339934.

## Part c

### 100 Path Simulation With The Rate as the Number of Letters in Last Name

**Question:** Simulate 100 paths of the Poisson process with intensity  $\lambda = \text{"number of letters in your last name"}$  via the exponential inter-arrival times construction method.

**Solution:**

Follow the below code chunk for the desired path simulation:

```
set.seed(123)

last_name <- "Garaye"
lambdaC <- nchar(last_name)
num_paths <- 100 # Given
T <- 10

simulate_poisson_path <- function(lambdaC, T){
  times <- cumsum(rexp(1e5, rate = lambdaC))
  times[times <= T]
}

poisson_paths <- lapply(1:num_paths, function(i) simulate_poisson_path(lambdaC, T))

plot(NA, xlim = c(0, T), ylim = c(0, 50), xlab = "Time", ylab = "Cumulative Count",
     main = "100 Poisson Process Paths (lambda = 6)")
for (i in 1:min(10, num_paths)) {
  lines(c(0, poisson_paths[[i]]), 0:length(poisson_paths[[i]]), type = "s", col = i)
}

arrival_counts <- sapply(poisson_paths, length)
```

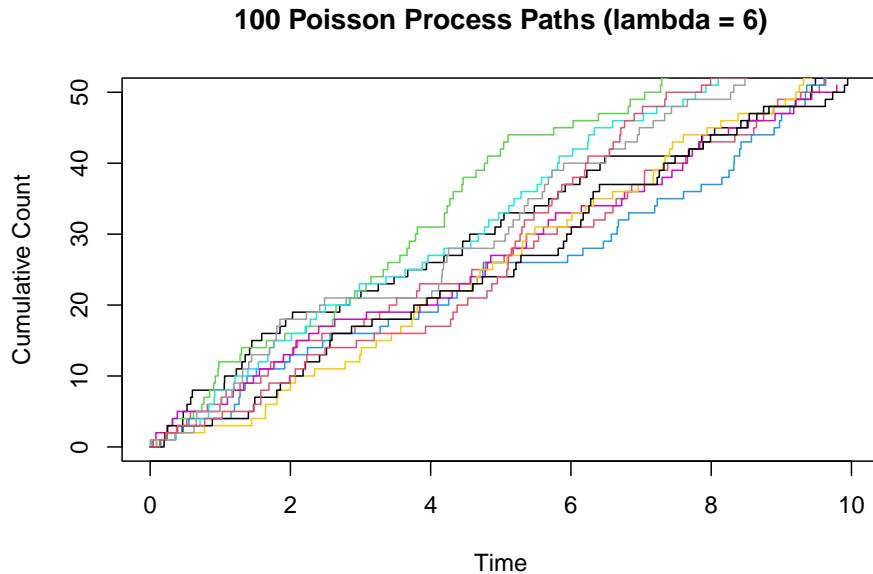


Figure 2: This plot illustrates 10 simulated paths of a Poisson process with an intensity of  $\lambda = 6$ , from the letters in my last name. The x-axis represents time, capped at  $T = 10$ , while the y-axis shows the cumulative number of events. Each step-like line corresponds to one possible path of the process. The simulation demonstrates randomness in the timing and count of events across different paths.

From the above simulation, the mean number of arrivals by  $T = 59.4$  and variance of arrivals by  $T = 44.0606061$ .