Solution Manual for:
Introduction to Probability Models: Eighth Edition
by Sheldon M. Ross.

John L. Weatherwax*

October 26, 2008

Introduction

Chapter 1: Introduction to Probability Theory

Chapter 1: Exercises

Exercise 8 (Bonferroni's inequality)

From the inclusion/exclusion identity for two sets we have

$$P(E \cup F) = P(E) + P(F) - P(EF)\,.$$

Since $P(E \cup F) \leq 1$, the above becomes

$$P(E) + P(F) - P(EF) \leq 1\,.$$

or

$$P(EF) \geq P(E) + P(F) - 1\,,$$

which is known as Bonferroni's inequality. From the numbers given we find that

$$P(EF) \geq 0.9 + 0.8 - 1 = 0.7\,.$$

*wax@alum.mit.edu

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Table 1: The possible values for the sum of the values when two die are rolled.

Exercise 10 (Boole's inequality)

We begin by decomposing the countable union of sets A_i

$$A_1 \cup A_2 \cup A_3 \dots$$

into a countable union of disjoint sets C_j. Define these disjoint sets as

$$\begin{aligned}
C_1 &= A_1 \\
C_2 &= A_2 \backslash A_1 \\
C_3 &= A_3 \backslash (A_1 \cup A_2) \\
C_4 &= A_4 \backslash (A_1 \cup A_2 \cup A_3) \\
&\vdots \\
C_j &= A_j \backslash (A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_{j-1})
\end{aligned}$$

Then by construction

$$A_1 \cup A_2 \cup A_3 \cdots = C_1 \cup C_2 \cup C_3 \cdots ,$$

and the C_j's are disjoint, so that we have

$$\Pr(A_1 \cup A_2 \cup A_3 \cup \cdots) = \Pr(C_1 \cup C_2 \cup C_3 \cup \cdots) = \sum_j \Pr(C_j) .$$

Since $\Pr(C_j) \leq \Pr(A_j)$, for each j, this sum is bounded above by

$$\sum_j \Pr(A_j) ,$$

Problem 11 (the probability the sum of the die is i)

We can explicitly enumerate these probabilities by counting the number of times each occurrence happens, in Table 1 we have placed the sum of the two die in the center of each square. Then by counting the number of squares where are sum equals each number from two to twelve, we have

$$
\begin{aligned}
P_2 &= \frac{1}{36}\,, \quad P_7 = \frac{6}{36} = \frac{1}{6} \\
P_3 &= \frac{2}{36} = \frac{1}{18}\,, \quad P_8 = \frac{5}{36} \\
P_4 &= \frac{3}{36} = \frac{1}{12}\,, \quad P_9 = \frac{4}{36} = \frac{1}{9} \\
P_5 &= \frac{4}{36} = \frac{1}{9}\,, \quad P_{10} = \frac{3}{36} = \frac{1}{12} \\
P_6 &= \frac{5}{36}\,, \quad P_{11} = \frac{2}{36} = \frac{1}{18}\,, \quad P_{12} = \frac{1}{36}\,.
\end{aligned}
$$

Problem 13 (winning at craps)

From Problem 11 we have computed the individual probabilities for various sum of two random die. Following the hint, let E_i be the event that the initial die sum to i and that the player wins. We can compute some of these probabilities immediately $P(E_2) = P(E_3) = P(E_{12}) = 0$, and $P(E_7) = P(E_{11}) = 1$. We now need to compute $P(E_i)$ for $i = 4, 5, 6, 8, 9, 10$. Again following the hint define $E_{i,n}$ to be the event that the player initial sum is i and wins on the n-th **subsequent** roll. Then

$$
P(E_i) = \sum_{n=1}^{\infty} P(E_{i,n})\,,
$$

since if we win, it must be either on the first, or second, or third, etc roll *after the initial roll.* We now need to calculate the $P(E_{i,n})$ probabilities for each n. As an example of this calculation first lets compute $P(E_{4,n})$ which means that we initially roll a sum of four and the player wins on the n-th subsequent roll. We will win if we roll a sum of a four or loose if we roll a sum of a seven, while if roll anything else we continue, so to win when $n = 1$ we see that

$$
P(E_{4,1}) = \frac{1+1+1}{36} = \frac{1}{12}\,,
$$

since to get a sum of four we can roll pairs consisting of $(1, 3)$, $(2, 2)$, and $(3, 1)$.

To compute $P(E_{4,2})$ the rules of craps state that we will win if a sum of four comes up (with probability $\frac{1}{12}$) and loose if a sum of a seven comes up (with probability $\frac{6}{36} = \frac{1}{6}$) and continue playing if anything else is rolled. This last event (continued play) happens with probability

$$
1 - \frac{1}{12} - \frac{1}{6} = \frac{3}{4}\,.
$$

Thus $P(E_{4,2}) = \left(\frac{3}{4}\right)\frac{1}{12} = \frac{1}{16}$. Here the first $\frac{3}{4}$ is the probability we don't roll a four or a seven on the $n = 1$ roll and the second $\frac{1}{12}$ comes from rolling a sum of a four on the second roll (where $n = 2$). In the same way we have for $P(E_{4,3})$ the following

$$
P(E_{4,3}) = \left(\frac{3}{4}\right)^2 \frac{1}{12}\,.
$$

Here the first two factors of $\frac{3}{4}$ are from the two rolls that "keep us in the game", and the factor of $\frac{1}{12}$, is the roll that allows us to win. Continuing in this in this manner we see that

$$P(E_{4,4}) = \left(\frac{3}{4}\right)^3 \frac{1}{12},$$

and in general we find that

$$P(E_{4,n}) = \left(\frac{3}{4}\right)^{n-1} \frac{1}{12} \quad \text{for} \quad n \geq 1\,.$$

To compute $P(E_{i,n})$ for other i, the derivations just performed, only change in the probabilities required to roll the initial sum. We thus find that for other initial rolls (heavily using the results of Problem 24) that

$$\begin{aligned}
P(E_{5,n}) &= \frac{1}{9}\left(1 - \frac{1}{9} - \frac{1}{6}\right)^{n-1} = \frac{1}{9}\left(\frac{13}{18}\right)^{n-1} \\
P(E_{6,n}) &= \frac{5}{36}\left(1 - \frac{5}{36} - \frac{1}{6}\right)^{n-1} = \frac{5}{36}\left(\frac{25}{36}\right)^{n-1} \\
P(E_{8,n}) &= \frac{5}{36}\left(1 - \frac{5}{36} - \frac{1}{6}\right)^{n-1} = \frac{5}{36}\left(\frac{25}{36}\right)^{n-1} \\
P(E_{9,n}) &= \frac{1}{9}\left(1 - \frac{1}{9} - \frac{1}{6}\right)^{n-1} = \frac{1}{9}\left(\frac{13}{18}\right)^{n-1} \\
P(E_{10,n}) &= \frac{1}{12}\left(1 - \frac{1}{12} - \frac{1}{6}\right)^{n-1} = \frac{1}{12}\left(\frac{3}{4}\right)^{n-1}.
\end{aligned}$$

To compute $P(E_4)$ we need to sum the results above. We have that

$$\begin{aligned}
P(E_4) &= \frac{1}{12}\sum_{n\geq 1}\left(\frac{3}{4}\right)^{n-1} = \frac{1}{12}\sum_{n\geq 0}\left(\frac{3}{4}\right)^{n} \\
&= \frac{1}{12}\frac{1}{\left(1 - \frac{3}{4}\right)} = \frac{1}{3}.
\end{aligned}$$

Note that this also gives the probability for $P(E_{10})$. For $P(E_5)$ we find $P(E_5) = \frac{2}{5}$, which also equals $P(E_9)$. For $P(E_6)$ we find that $P(E_6) = \frac{5}{11}$, which also equals $P(E_8)$. Then our probability of winning craps is given by summing all of the above probabilities weighted by the associated priors of rolling the given initial roll. We find by defining I_i to be the event that the initial roll is i and W the event that we win at craps that

$$\begin{aligned}
P(W) &= 0\,P(I_2) + 0\,P(I_3) + \frac{1}{3}P(I_4) + \frac{4}{9}P(I_5) + \frac{5}{9}P(I_6) \\
&+ 1\,P(I_7) + \frac{5}{9}P(I_8) + \frac{4}{9}P(I_9) + \frac{1}{3}P(I_{10}) + 1\,P(I_{11}) + 0\,P(I_{12})\,.
\end{aligned}$$

Using the results of Exercise 25 to evaluate $P(I_i)$ for each i we find that the above summation gives

$$P(W) = \frac{244}{495} = 0.49292\,.$$

These calculations are performed in the Matlab file `chap_1_prob_13.m`.

Exercise 15 (some set identities)

We want to prove that $E = (E \cap F) \cup (E \cap F^c)$. We will do this using the standard proof where we show that each set in the above is a subset of the other. We begin with $x \in E$. Then if $x \in F$, x will certainly be in $E \cap F$, while if $x \notin F$ then x will be in $E \cap F^c$. Thus in either case ($x \in F$ or $x \notin F$) x will be in the set $(E \cap F) \cup (E \cap F^c)$.

If $x \in (E \cap F) \cup (E \cap F^c)$ then x is in either $E \cap F$, $E \cap F^c$, or both by the definition of the union operation. Now x cannot be in both sets or else it would simultaneously be in F and F^c, so x must be in one of the two sets only. Being in either set means that $x \in E$ and we have that the set $(E \cap F) \cup (E \cap F^c)$ is a subset of E. Since each side is a subset of the other we have shown set equality.

To prove that $E \cup F = E \cup (E^c \cap F)$, we will begin by letting $x \in E \cup F$, thus x is an element of E or an element of F or of both. If x is in E at all then it is in the set $E \cup (E^c \cap F)$. If $x \notin E$ then it must be in F to be in $E \cup F$ and it will therefore be in $E^c \cap F$. Again both sides are subsets of the other and we have shown set equality.

Exercise 23 (conditioning on a chain of events)

This result follows for the two set case $P\{A \cap B\} = P\{A|B\}P\{B\}$ by grouping the sequence of E_i's in the appropriate manner. For example by grouping the intersection as

$$E_1 \cap E_2 \cap \cdots \cap E_{n-1} \cap E_n = (E_1 \cap E_2 \cap \cdots \cap E_{n-1}) \cap E_n$$

we can apply the two set result to obtain

$$P\{E_1 \cap E_2 \cap \cdots \cap E_{n-1} \cap E_n\} = P\{E_n|E_1 \cap E_2 \cap \cdots \cap E_{n-1}\}\, P\{E_1 \cap E_2 \cap \cdots \cap E_{n-1}\}\,.$$

Continuing now to peal E_{n-1} from the set $E_1 \cap E_2 \cap \cdots \cap E_{n-1}$ we have the second probability above equal to

$$P\{E_1 \cap E_2 \cap \cdots \cap E_{n-2} \cap E_{n-1}\} = P\{E_{n-1}|E_1 \cap E_2 \cap \cdots \cap E_{n-2}\}P\{E_1 \cap E_2 \cap \cdots \cap E_{n-2}\}\,.$$

Continuing to peal off terms from the back we eventually obtain the requested expression i.e.

$$\begin{aligned}
P\{E_1 \cap E_2 \cap \cdots \cap E_{n-1} \cap E_n\} &= P\{E_n|E_1 \cap E_2 \cap \cdots \cap E_{n-1}\} \\
&\times\ P\{E_{n-1}|E_1 \cap E_2 \cap \cdots \cap E_{n-2}\} \\
&\times\ P\{E_{n-2}|E_1 \cap E_2 \cap \cdots \cap E_{n-3}\} \\
&\ \vdots \\
&\times\ P\{E_3|E_1 \cap E_2\} \\
&\times\ P\{E_2|E_1\} \\
&\times\ P\{E_1\}\,.
\end{aligned}$$

Exercise 30 (target shooting with Bill and George)

warning! not finished...

Let H be the event that the duck is "hit", by either Bill or George's shot. Let B and G be the events that Bill (respectively George) hit the target. Then the outcome of the experiment where both George and Bill fire at the target (assuming that their shots work independently is)

$$\begin{aligned} P(B^c, G^c) &= (1-p_1)(1-p_2) \\ P(B^c, G) &= (1-p_1)p_2 \\ P(B, G^c) &= p_1(1-p_2) \\ P(B, G) &= p_1 p_2 \,. \end{aligned}$$

Part (a): We desire to compute $P(B, G|H)$ which equals

$$P(B, G|H) = \frac{P(B, G, H)}{P(H)} = \frac{P(B, G)}{P(H)}$$

Now $P(H) = (1-p_1)p_2 + p_1(1-p_2) + p_1 p_2$ so the above probability becomes

$$\frac{p_1 p_2}{(1-p_1)p_2 + p_1(1-p_2) + p_1 p_2} = \frac{p_1 p_2}{p_1 + p_2 - p_1 p_2} \,.$$

Part (b): We desire to compute $P(B|H)$ which equals

$$P(B|H) = P(B, G|H) + P(B, G^c|H) \,.$$

Since the first term $P(B, G|H)$ has already been computed we only need to compute $P(B, G^c|H)$. As before we find it to be

$$P(B, G^c|H) = \frac{p_1(1-p_2)}{(1-p_1)p_2 + p_1(1-p_2) + p_1 p_2} \,.$$

So the total result becomes

$$P(B|H) = \frac{p_1 p_2 + p_1(1-p_2)}{(1-p_1)p_2 + p_1(1-p_2) + p_1 p_2} = \frac{p_1}{p_1 + p_2 - p_1 p_2} \,.$$

Exercise 33 (independence in class)

Let S be a random variable denoting the sex of the randomly selected person. The S can take on the values m for male and f for female. Let C be a random variable representing denoting the class of the chosen student. The C can take on the values f for freshman and s for sophomore. We want to select the number of sophomore girls such that the random variables S and C are independent. Let n denote the number of sophomore girls. Then counting up the number of students that satisfy each requirement we have

$$\begin{aligned} P(S=m) &= \frac{10}{16+n} \\ P(S=f) &= \frac{6+n}{16+n} \\ P(C=f) &= \frac{10}{16+n} \\ P(C=s) &= \frac{6+n}{16+n} \,. \end{aligned}$$

The joint density can also be computed and are given by

$$\begin{aligned} P(S=m, C=f) &= \frac{4}{16+n} \\ P(S=m, C=s) &= \frac{6}{16+n} \\ P(S=f, C=f) &= \frac{6}{16+n} \\ P(S=f, C=s) &= \frac{n}{16+n} \,. \end{aligned}$$

Then to be independent we must have $P(C,S) = P(S)P(C)$ for all possible C and S values. Considering the point case where $(S=m, C=f)$ we have that n must satisfy

$$\begin{aligned} P(S=m, C=f) &= P(S=m)P(C=f) \\ \frac{4}{16+n} &= \left(\frac{10}{16+n}\right)\left(\frac{10}{16+n}\right) \end{aligned}$$

which when we solve for n gives $n=9$. Now one should check that this value of n works for all other equalities that must be true, for example one needs to check that when $n=9$ the following are true

$$\begin{aligned} P(S=m, C=s) &= P(S=m)P(C=s) \\ P(S=f, C=f) &= P(S=f)P(C=f) \\ P(S=f, C=s) &= P(S=f)P(C=s) \,. \end{aligned}$$

As these can be shown to be true, $n=9$ is the correct answer.

Exercise 36 (boxes with marbles)

Let B be the event that the drawn ball is black and let X_1 (X_2) be the event that we select the first (second) box. Then to calculate $P(B)$ we will condition on the box drawn from as

$$P(B) = P(B|X_1)P(X_1) + P(B|X_2)P(X_2) \,.$$

Now $P(B|X_1) = 1/2$, $P(B|X_2) = 2/3$, $P(X_1) = P(X_2) = 1/2$ so

$$P(B) = \frac{1}{2}\left(\frac{1}{2}\right) + \frac{1}{2}\left(\frac{2}{3}\right) = \frac{7}{12} \,.$$

Exercise 37 (observing a white marble)

If we see that the ball is white (i.e. it is not black i.e event B^c has happened) we now want to compute that it was drawn from the first box i.e.

$$P(X_1|B^c) = \frac{P(B^c|X_1)P(X_1)}{P(B^c|X_1)P(X_1) + P(B^c|X_2)P(X_2)} = \frac{3}{5}\,.$$

Problem 40 (gambling with a fair coin)

Let F denote the event that the gambler is observing results from a fair coin. Also let O_1, O_2, and O_3 denote the three observations made during our experiment. We will assume that before any observations are made the probability that we have selected the fair coin is 1/2.

Part (a): We desire to compute $P(F|O_1)$ or the probability we are looking at a fair coin given the first observation. This can be computed using Bayes' theorem. We have

$$\begin{aligned} P(F|O_1) &= \frac{P(O_1|F)P(F)}{P(O_1|F)P(F) + P(O_1|F^c)P(F^c)} \\ &= \frac{\frac{1}{2}\left(\frac{1}{2}\right)}{\frac{1}{2}\left(\frac{1}{2}\right) + 1\left(\frac{1}{2}\right)} = \frac{1}{3}\,. \end{aligned}$$

Part (b): With the second observation and using the "posteriori's become priors" during a recursive update we now have

$$\begin{aligned} P(F|O_2, O_1) &= \frac{P(O_2|F, O_1)P(F|O_1)}{P(O_2|F, O_1)P(F|O_1) + P(O_2|F^c, O_1)P(F^c|O_1)} \\ &= \frac{\frac{1}{2}\left(\frac{1}{3}\right)}{\frac{1}{2}\left(\frac{1}{3}\right) + 1\left(\frac{2}{3}\right)} = \frac{1}{5}\,. \end{aligned}$$

Part (c): In this case because the two-headed coin cannot land tails we can immediately conclude that we have selected the fair coin. This result can also be obtained using Bayes' theorem as we have in the other two parts of this problem. Specifically we have

$$\begin{aligned} P(F|O_3, O_2, O_1) &= \frac{P(O_3|F, O_2, O_1)P(F|O_2, O_1)}{P(O_3|F, O_2, O_1)P(F|O_2, O_1) + P(O_3|F^c, O_2, O_1)P(F^c|O_2, O_1)} \\ &= \frac{\frac{1}{2}\left(\frac{1}{5}\right)}{\frac{1}{2}\left(\frac{1}{5}\right) + 0} = 1\,. \end{aligned}$$

Verifying what we know must be true.

Problem 46 (a prisoners' dilemma)

I will argue that the jailers reasoning is sound. Before asking his question the probability of event A (A is executed) is $P(A) = 1/3$. If prisoner A is told that B (or C) is to be set free then we need to compute $P(A|B^c)$. Where A, B, and C are the events that prisoner A, B, or C is to be executed respectively. Now from Bayes' rule

$$P(A|B^c) = \frac{P(B^c|A)P(A)}{P(B^c)}.$$

We have that $P(B^c)$ is given by

$$P(B^c) = P(B^c|A)P(A) + P(B^c|B)P(B) + P(B^c|C)P(C) = \frac{1}{3} + 0 + \frac{1}{3} = \frac{2}{3}.$$

So the above probability then becomes

$$P(A|B^c) = \frac{1(1/3)}{2/3} = \frac{1}{2} > \frac{1}{3}.$$

Thus the probability that prisoner A will be executed has increased as claimed by the jailer.

Chapter 4: Markov Chains

Chapter 4: Exercises

Exercise 6 (an analytic calculation of $P^{(n)}$)

Given the transition probability matrix $P = \left[\begin{array}{cc} p & 1-p \\ 1-p & p \end{array}\right]$, by matrix multiplication we see that $P^{(2)}$ is given as

$$P^{(2)} = \left[\begin{array}{cc} p & 1-p \\ 1-p & p \end{array}\right] \left[\begin{array}{cc} p & 1-p \\ 1-p & p \end{array}\right] = \left[\begin{array}{cc} p^2 + (1-p)^2 & 2p(1-p) \\ 2p(1-p) & p^2 + (1-p)^2 \end{array}\right] .$$

We desire to prove that $P^{(n)}$ is given as

$$P^{(n)} = \left[\begin{array}{cc} \frac{1}{2} + \frac{1}{2}(2p-1)^n & \frac{1}{2} - \frac{1}{2}(2p-1)^n \\ \frac{1}{2} - \frac{1}{2}(2p-1)^n & \frac{1}{2} + \frac{1}{2}(2p-1)^n \end{array}\right] . \tag{1}$$

We will do this by mathematical induction. We begin by verifying that the above formula is valid for $n = 1$. Evaluating the above expression for $n = 1$ we find

$$P^{(1)} = \left[\begin{array}{cc} \frac{1}{2} + \frac{1}{2}(2p-1) & \frac{1}{2} - \frac{1}{2}(2p-1) \\ \frac{1}{2} - \frac{1}{2}(2p-1) & \frac{1}{2} + \frac{1}{2}(2p-1) \end{array}\right] = \left[\begin{array}{cc} p & 1-p \\ 1-p & p \end{array}\right] ,$$

as required. Next we assume the relationship in Equation 1 is true for all $n \leq k$ and we desire to show that it is true for $n = k+1$. Since $P^{(k+1)} = P^{(1)}P^{(k)}$ by matrix multiplication we have that

$$\begin{aligned}
P^{(k+1)} &= P^{(1)}P^{(k)} = \left[\begin{array}{cc} p & 1-p \\ 1-p & p \end{array}\right] \left[\begin{array}{cc} \frac{1}{2} + \frac{1}{2}(2p-1)^k & \frac{1}{2} - \frac{1}{2}(2p-1)^k \\ \frac{1}{2} - \frac{1}{2}(2p-1)^k & \frac{1}{2} + \frac{1}{2}(2p-1)^k \end{array}\right] \\
&= \left[\begin{array}{cc} \frac{p}{2} + \frac{p}{2}(2p-1)^k + \frac{1-p}{2} - \frac{1-p}{2}(2p-1)^k & \frac{p}{2} - \frac{p}{2}(2p-1)^k + \frac{1-p}{2} + \frac{1-p}{2}(2p-1)^k \\ \frac{1-p}{2} + \frac{1-p}{2}(2p-1)^k + \frac{p}{2} - \frac{p}{2}(2p-1)^k & \frac{1-p}{2} - \frac{1-p}{2}(2p-1)^k + \frac{p}{2} + \frac{p}{2}(2p-1)^k \end{array}\right] \\
&= \left[\begin{array}{cc} \frac{1}{2} + \frac{1}{2}(p-(1-p))(2p-1)^k & \frac{1}{2} + \left(-\frac{p}{2} + \frac{1-p}{2}\right)(2p-1)^k \\ \frac{1}{2} + \frac{1}{2}(1-p-p)(2p-1)^k & \frac{1}{2} + \frac{1}{2}(-(1-p)+p)(2p-1)^k \end{array}\right] \\
&= \left[\begin{array}{cc} \frac{1}{2} + \frac{1}{2}(2p-1)^{k+1} & \frac{1}{2} - \frac{1}{2}(2p-1)^{k+1} \\ \frac{1}{2} - \frac{1}{2}(2p-1)^{k+1} & \frac{1}{2} + \frac{1}{2}(2p-1)^{k+1} \end{array}\right] ,
\end{aligned}$$

which is the desired result for $n = k+1$.

Chapter 5: The Exponential Distribution and the Poisson Process

Chapter 5: Exercises

Exercise 1 (exponential repair times)

We are told that T is distributed with a exponential probability distribution function with mean $1/2$. This means that the distribution function of T, $f_T(t)$, is given by

$$f_T(t) = \begin{cases} 2e^{-2t} & t \geq 0 \\ 0 & t < 0 \end{cases}$$

Part (a): Now for the repair time to take longer than $1/2$ an hour will happen with probability that is the complement of the probability that it will take *less* than $1/2$ of an hour. We have

$$\begin{aligned} 1 - P\{T < 1/2\} &= 1 - \int_0^{1/2} 2e^{-2t} dt \\ &= 1 - \left. \frac{2e^{-2t}}{-2} \right|_0^{1/2} \\ &= 1 + \left(e^{-1} - 1\right) = e^{-1} . \end{aligned}$$

Part (b): Since the exponential distribution has no memory, the fact that the repair is still going after 12 hours is irrelevant. Thus we only need to compute the probability that the repair will last at least $1/2$ more. This probability is the same as that calculated in Part (a) of this problem and is equal to e^{-1}.

Exercise 2 (the expected bank waiting time)

By the memoryless property of the exponential distribution now the fact that one person is being served makes no difference. We will have to wait until an amount of time given by

$$T = \sum_{i=1}^{6} X_i ,$$

where X_i are independent exponential random variables with rate μ. Taking the expectation of the variable T we have

$$E[T] = \sum_{i=1}^{6} E[X_i] = \sum_{i=1}^{6} \frac{1}{\mu} = \frac{6}{\mu} .$$

We sum to six to allow the teller to service the five original customers and then you.

Exercise 3 (expectations with exponential random variables)

I will argue that $E[X^2|X > 1] = E[(X+1)^2]$. By the memoryless property of the exponential random variable the fact that we are conditioning on the event that $X > 1$ makes no difference relative to the event $X > 0$ (i.e. no restriction on the random variable X). Removing the conditional expectation is equivalent to "starting" the process at $x = 0$. This can be performed as long as we "shift" the expectation's argument accordingly i.e. from X^2 to $(X + 1)^2$. The other two expressions violate the nonlinearity of the function X^2. We can prove that this result is correct by explicitly evaluating the original expectation. We find

$$E[X^2|X > 1] = \int_0^\infty \xi^2 p_X(\xi|X > 1)d\xi\,.$$

Now this conditional probability density is given by

$$\begin{aligned}
p_X(\xi|X > 1) &= \frac{p(X = \xi, X > 1)}{p(X > 1)} \\
&= \frac{p(X = \xi, X > 1)}{1 - p(X < 1)} \\
&= \frac{p(X = \xi, X > 1)}{e^{-\lambda}} \\
&= \frac{\lambda e^{-\lambda\xi} H(\xi - 1)}{e^{-\lambda}} = \lambda e^{-\lambda(\xi-1)} H(\xi - 1)\,.
\end{aligned}$$

Here $H(\cdot)$ is the Heaviside function defined as

$$H(x) = \begin{cases} 0 & x < 0 \\ 1 & x > 0\,, \end{cases}$$

and this function enforces the constraint that $X > 1$. With this definition we then have that

$$\begin{aligned}
E[X^2|X > 1] &= \int_0^\infty \xi^2 \lambda e^{-\lambda(\xi-1)} H(\xi - 1)d\xi \\
&= \int_1^\infty \xi^2 \lambda e^{-\lambda(\xi-1)} d\xi\,,
\end{aligned}$$

Letting $u = \xi - 1$, so that $du = d\xi$ the above becomes

$$\int_0^\infty (u + 1)^2 \lambda e^{-\lambda u} du = E[(X + 1)^2]\,,$$

where X is an exponential random variable with rate parameter λ. This is the expression we argued at the beginning of this problem should hold true.

Exercise 4 (the post office)

Part (a): In this case, it is not possible for A to still be in the post office because in ten minutes time A and B will both finish their service times and exit the service station together. Thus there is no way for C to get served before A finishes.

Part (b): A will still be in the post office if both B and C are served before A. If we let A, B, and C be the amount of time that each respective person spends with their clerks, then the event that A is the last person in the post office is equivalent to the constraint that $A \geq B+C$. Here we have assumed (by using $\geq$) that an equality constraint is acceptable for determining if A leaves last. For notational purposes we will define the event that A leaves last as E. To compute the probabilities that this event happens we can condition on the possible sums of B and C (the times B and C spend at their clerks). We have

$$\begin{aligned} P(E) &= P(E|B+C=2)P(B+C=2) + P(E|B+C=3)P(B+C=3) \\ &+ P(E|B+C=4)P(B+C=4) + P(E|B+C=5)P(B+C=5) \\ &+ P(E|B+C=6)P(B+C=6)\,. \end{aligned}$$

Now $P(E|B+C=4) = P(E|B+C=5) = P(E|B+C=6) = 0$, since A will certainly finish in a time less than four units. Also

$$P(E|B+C=2) = \frac{2}{3}\,,$$

since to have $A \geq B+C$, A can finishes in two *or* three units. While finally

$$P(E|B+C=3) = \frac{1}{3}\,,$$

since to have $A \geq B+C$, A can finish in three units. Now we have that our priors are given by $P(B+C=2) = \frac{1}{9}$ and $P(B+C=3) = \frac{2}{9}$, which gives for $P(E)$ using the above formula

$$P(E) = \frac{4}{27}\,.$$

If we want to work this problem assuming strict inequality in the time relationships i.e. that A will leave last only if $A > B+C$, we find that our conditional probabilities must be adjusted. For example

$$P(E|B+C=2) = \frac{1}{3}\,,$$

since A must now finish in three units of time. Also $P(E|B+C=3) = 0$. These then give in the same way that

$$P(E) = \frac{1}{3} \cdot \frac{1}{9} = \frac{1}{27}\,.$$

Part (c): For this part we assume that the service time of each clerk is exponentially distributed. Now because the random variables are continuous there is no need to differentiate between greater than and greater than or equal signs in the inequality denoting the event that A finishes last. Thus we can take as an expression of the event that A will be the last person served if $A > B+C$. This means that the time to service A takes more time than to service both B and C. To evaluate the probability that this event happens we will condition on the random variable which is the sum of B and C, i.e.

$$P\{A > B+C\} = \int p\{A > B+C|B+C=t\}P\{B+C=t\}dt\,.$$

Now since both B and C are exponential random variables with the *same rate* the sum of them is a random variable distributed with a gamma distribution (see the section on further properties of the exponential distribution). We thus have that the distribution of the sum of B and C is given by

$$f_{B+C}(t) = \frac{1!e^{-\mu t}(\mu t)}{1!} = \mu^2 t e^{-\mu t}\,.$$

So that our integral is given by

$$P\{A > B + C\} = \int P_A\{A > t\}\mu^2 t e^{-\mu t} dt\,.$$

Now $P_A\{A > t\} = 1 - (1 - e^{-\mu t}) = e^{-\mu t}$, so that this integral becomes

$$\begin{aligned} P\{A > B + C\} &= \int_0^\infty \mu^2 t e^{-2\mu t} dt \\ &= \mu^2 \frac{t e^{-2\mu t}}{(-2\mu)}\Big|_0^\infty - \frac{\mu^2}{(-2\mu)}\int_0^\infty e^{-2\mu t} dt \\ &= \frac{\mu}{2}\left(\frac{e^{-2\mu t}}{(-2\mu)}\Big|_0^\infty = \frac{1}{4}\,. \end{aligned}$$

Exercise 5 (old radios)

Because of the memoryless property of the exponential distribution the fact that the radio is ten years old makes no difference. Thus the probability that the radio is working after an additional ten years is given by

$$P\{X > 10\} = \int_{10}^\infty \lambda e^{-\lambda t} dt = 1 - F(10) = 1 - \frac{1}{10}e^{-\frac{1}{10}(10)} = 1 - \frac{1}{10}e^{-1} = 0.96321\,.$$

Exercise 6 (the probability that Smith is last)

Now one of Mr. Jones or Mr. Brown will finish their service first. Once this person's service station is open, Smith will begin his processing there. By the memoryless property of exponential the fact that Smiths "competitor" (the remaining Mr. Jones or Mr. Brown who did not finish first) is in the middle of their service has no effect on whether Smith or his comp editor finishes first. Let E be the event that Smith is finishes last. Then we have that $P(E^c) = 1 - P(E)$, where E^c is the desired event (i.e. that Smith *is* last). Now let JB be the event that Mr. Jones finishes before Mr. Brown (then Mr. Smith will take Mr. Jones' server position), and let JB be the event that Mr. Brown finishes before Mr. Jones (so that Mr. Smith would take Mr. Brown's server positions). We have that by conditioning on the events JB and BJ that

$$P(E) = P(E|JB)P(JB) + P(E|BJ)P(BJ)\,.$$

Now

$$P(JB) = \frac{\lambda_J}{\lambda_J + \lambda_B} \quad \text{and} \quad P(BJ) = \frac{\lambda_B}{\lambda_J + \lambda_B}\,,$$

where λ_J and λ_B are the rates of the servers who are initially servicing Mr. Jones and Mr. Brown when Mr. Smith enters the post office. By the discussion in the text on the section of the text entitled "Further properties of the exponential distribution", we have that $P(E|JB)$ is the probability that Mr. Smith is last given that he is being serviced by Mr. Jones' old server. This will happen if Mr. Brown finishes first and thus has a probability of

$$\frac{\lambda_B}{\lambda_B + \lambda_J}.$$

In the same way we have

$$P(E|BJ) = \frac{\lambda_J}{\lambda_B + \lambda_J}.$$

So that $P(E)$ is given by

$$P(E) = \frac{\lambda_B}{\lambda_B + \lambda_J} \cdot \frac{\lambda_J}{\lambda_B + \lambda_J} + \frac{\lambda_J}{\lambda_B + \lambda_J} \cdot \frac{\lambda_B}{\lambda_B + \lambda_J} = \frac{2\lambda_B \lambda_J}{(\lambda_B + \lambda_J)^2}.$$

The probability that we want is $P(E^c)$ and is given by

$$\begin{aligned} P(E^c) &= 1 - P(E) \\ &= \frac{(\lambda_B + \lambda_J)^2}{(\lambda_B + \lambda_J)^2} - \frac{2\lambda_B \lambda_J}{(\lambda_B + \lambda_J)^2} \\ &= \frac{\lambda_B^2}{(\lambda_B + \lambda_J)^2} + \frac{\lambda_J^2}{(\lambda_B + \lambda_J)^2}, \end{aligned}$$

as we where to show.

Exercise 7 (the probability $X_1 < X_2$ given the minimum value of X_1 and X_2)

For this exercise we will be considering the expression $P\{X_1 < X_2 | \min(X_1, X_2) = t\}$. Using the definition of conditional probability we have that this is equal to

$$\begin{aligned} P\{X_1 < X_2 | \min(X_1, X_2) = t\} &= \frac{P\{X_1 < X_2, \min(X_1, X_2) = t\}}{P\{\min(X_1, X_2) = t\}} \\ &= \frac{P\{X_1 < X_2, X_1 = t\}}{P\{\min(X_1, X_2) = t\}} \\ &= \frac{P\{t < X_2, X_1 = t\}}{P\{\min(X_1, X_2) = t\}} \\ &= \frac{P\{X_2 > t\} P\{X_1 = t\}}{P\{\min(X_1, X_2) = t\}}. \end{aligned}$$

Where in the last step we used the independence of X_1 and X_2. We now need to be able to relate the expressions above in terms of the failure rates of the random variables X_1 and X_2. Given the failure rate of a random variable as a function of t say $r(t)$ from the discussion in the book the distribution function $F(\cdot)$ in this case is given by

$$F(t) = e^{-\int_0^t r(\tau) d\tau}.$$

Taking the derivative of this expression we find that the density function $f(\cdot)$ for our random variable with failure rate $r(t)$ is given by

$$f(t) = r(t)e^{-\int_0^t r(\tau)d\tau}\,.$$

With these two expressions we can evaluate part of the above fraction. We find that

$$\frac{P\{X_2 > t\}P\{X_1 = t\}}{P\{\min(X_1, X_2) = t\}} = \frac{(e^{-\int_0^t r_2(\tau)d\tau})r_1(t)e^{-\int_0^t r_1(\tau)d\tau}}{P\{\min(X_1, X_2) = t\}}\,.$$

Lets now work on evaluating the denominator of the above expression, which we do in the same way as in the book. We find that

$$\begin{aligned}
P\{\min(X_1, X_2) > t\} &= P\{X_i > t\,, \forall i\} \\
&= \prod_{i=1}^{2} P\{X_i > t\} \\
&= e^{-\int_0^t r_1(\tau)d\tau} e^{-\int_0^t r_2(\tau)d\tau} \\
&= e^{-\int_0^t (r_1(\tau)+r_2(\tau))d\tau}\,.
\end{aligned}$$

From which we see that $\min(X_1, X_2)$ is a random variable with a failure rate given by $r_1(t)+r_2(t)$. This then means that our *density* function for this random variable $\min(X_1, X_2)$ is given by

$$P\{\min(X_1, X_2) = t\} = (r_1(t) + r_2(t))e^{-\int_0^t (r_1(\tau)+r_2(\tau))d\tau}\,.$$

Using this result in our desired expression we finally conclude that

$$\begin{aligned}
P\{X_1 < X_2 | \min(X_1, X_2) = t\} &= \frac{r_1(t)e^{-\int_0^t (r_1(\tau)+r_2(\tau))d\tau}}{(r_1(t) + r_2(t))e^{-\int_0^t (r_1(\tau)+r_2(\tau))d\tau}} \\
&= \frac{r_1(t)}{r_1(t) + r_2(t)}\,,
\end{aligned}$$

as we were to show.

Exercise 8 (the expectation of $1/r(x)$)

We will use an equivalent expression for the expectation of X which is derived in many books on probability, for example see **??**, which is that

$$E[X] = \int P\{X > x\}dx\,.$$

From this expression and with the following manipulations to introduce the failure rate function $r(X)$ we find that

$$\begin{aligned}E[X] &= \int P\{X > x\}dx \\ &= \int \frac{1-F(x)}{f(x)} f(x)dx \\ &= \int \frac{1}{r(x)} f(x)dx \\ &= E[\frac{1}{r(X)}].\end{aligned}$$

Here we have used $f(\cdot)$ for the probability density function, $F(\cdot)$ for the cumulative density function and $r(\cdot)$ for the failure rate function, i.e. $r(x) \equiv \frac{f(x)}{1-F(x)}$.

Exercise 9 (the probability we fail first after working t time)

We are told that machine one is currently working and at time t, machine two will be put into service. At that point if the first machine is still working they will be working in tandem. We want to calculate the probability that machine one is the first machine to fail. Let E be the event that machine one fails first and F the event that machine one fails in the time $(0, t)$. Now conditioning on F and F^c we have that

$$P(E) = P(E|F)P(F) + P(E|F^c)P(F^c).$$

Now introducing X_1 and X_2 be the random variables denoting the *time* when machine one/two fails.

$$\begin{aligned}P(F) &= P\{X_1 < t\} = 1 - e^{-\lambda_1 t} \\ P(F^c) &= P\{X_1 > t\} = e^{-\lambda_1 t}.\end{aligned}$$

Now by the memoryless property of the exponential distribution when the second machine is brought on-line with the first machine (assuming that machine one has not failed in the time $(0, t)$) the probability that machine one fails before machine two is given by

$$\frac{\lambda_1}{\lambda_1 + \lambda_2},$$

which is the probability $P(E|F^c)$. In addition we have that $P(E|F) = 1$, so that $P(E)$ becomes

$$\begin{aligned}P(E) &= 1 \cdot (1 - e^{-\lambda_1 t}) + \frac{\lambda_1}{\lambda_1 + \lambda_2} \cdot e^{-\lambda_1 t} \\ &= 1 - \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right) e^{-\lambda_1 t}\end{aligned}$$

Exercise 10 (some expectations)

Part (a): We want to evaluate $E[MX|M = X]$. Note that if we try to evaluate this as

$$E[MX|M = X] = E[X^2]\,,$$

this result cannot be correct since we have effectively lost the information that our minimum of X and Y is X since the last expression involves only X. Instead we find that

$$E[MX|M = X] = E[M^2] = \int_0^\infty m^2 e^{-(\lambda+\mu)m} dm = \frac{2}{(\lambda+\mu)^2}\,,$$

when we perform the integration.

Part (b): To evaluate $E[MX|M = Y]$, we will stress the information that $Y = M$ is less that X by writing $X = M + X'$. Now by the memoryless property of the exponential distribution X' is another exponential random variable with failure rate λ. With this we find that

$$\begin{aligned}
E[MX|M = Y] &= E[M(M + X')] \\
&= E[M^2] + E[MX'] \\
&= E[M^2] + E[M]E[X'] \\
&= \frac{2}{(\lambda+\mu)^2} + \frac{1}{\lambda+\mu} \cdot \frac{1}{\lambda}\,.
\end{aligned}$$

Part (c): We want to evaluate $\mathrm{Cov}(X, M)$, which we will do by simplifying the definition of the covariance to find

$$\mathrm{Cov}(X, M) = E[XM] - E[X]E[M]\,.$$

To compute $E[XM]$ we can condition on whether the minimum is X or Y. We find that

$$\begin{aligned}
E[XM] &= E[XM|M = X]P\{M = X\} + E[XM|M = Y]P\{M = Y\} \\
&= \frac{2}{(\lambda+\mu)^2}\left(\frac{\lambda}{\lambda+\mu}\right) + \left(\frac{2}{(\lambda+\mu)^2} + \frac{1}{\lambda(\lambda+\mu)}\right)\frac{\mu}{\lambda+\mu} \\
&= \frac{2\lambda+\mu}{\lambda(\lambda+\mu)^2}\,,
\end{aligned}$$

so that with this result $\mathrm{Cov}(X, M)$ is then given by

$$\begin{aligned}
\mathrm{Cov}(X, M) &= \frac{2\lambda+\mu}{\lambda(\lambda+\mu)^2} - \frac{1}{\lambda+\mu}\left(\frac{1}{\lambda}\right) \\
&= \frac{1}{(\lambda+\mu)^2}\,.
\end{aligned}$$

Exercise 12 (probabilities with three exponential random variables)

Part (a): We can evaluate $P\{X_1 < X_2 < X_3\}$ using the definition of this expression. We find.

$$\begin{aligned}
P\{X_1 < X_2 < X_3\} &= \int_{x_1=0}^{\infty} \int_{x_2=x_1}^{\infty} \int_{x_3=x_2}^{\infty} p(x_1, x_2, x_3) dx_3 dx_2 dx_1 \\
&= \int_{x_1=0}^{\infty} \int_{x_2=x_1}^{\infty} \int_{x_3=x_2}^{\infty} p_{X_1}(x_1) p_{X_2}(x_2) p_{X_3}(x_3) dx_3 dx_2 dx_1 \\
&= \int_{x_1=0}^{\infty} p_{X_1}(x_1) \int_{x_2=x_1}^{\infty} p_{X_2}(x_2) \int_{x_3=x_2}^{\infty} p_{X_3}(x_3) dx_3 dx_2 dx_1 \,,
\end{aligned}$$

by the independence of the random variables X_i. Because X_i are exponential with failure rate λ_i the above becomes

$$P\{X_1 < X_2 < X_3\} = \int_{x_1=0}^{\infty} \lambda_1 e^{-\lambda_1 x_1} \int_{x_2=x_1}^{\infty} \lambda_2 e^{-\lambda_2 x_2} \int_{x_3=x_2}^{\infty} \lambda_3 e^{-\lambda_3 x_3} dx_3 dx_2 dx_1 \,.$$

To compute the above probability We can perform the above integrations one at a time from the outside inward. The first integral is with respect to x_3, the second with respect to x_2, and finally the third integral is with respect to x_1. When this is done we obtain

$$P\{X_1 < X_2 < X_3\} = \frac{\lambda_1 \lambda_2}{(\lambda_2 + \lambda_3)(\lambda_1 + \lambda_2 + \lambda_3)} \,.$$

See the **Mathematica** file `chap_5_prob_12.nb` for this algebra.

Part (b): We want to evaluate $P\{X_1 < X_2 | \max(X_1, X_2, X_3) = X_3\}$. Using the definition of the conditional expectation we find that

$$\begin{aligned}
P\{X_1 < X_2 | \max(X_1, X_2, X_3) = X_3\} &= \frac{P\{X_1 < X_2, \max(X_1, X_2, X_3) = X_3\}}{P\{\max(X_1, X_2, X_3) = X_3\}} \\
&= \frac{P\{X_1 < X_2 < X_3\}}{P\{\max(X_1, X_2, X_3) = X_3\}} \,,
\end{aligned}$$

since the two events $X_1 < X_2$ *and* $\max(X_1, X_2, X_3) = X_3$ imply that $X_1 < X_2 < X_3$. The probability of this event was calculated in Part (a) of this problem. Now the event $\max(X_1, X_2, X_3) = X_3$ is equivalent to the union of the disjoint events

$$X_1 < X_2 < X_3 \quad \text{and} \quad X_2 < X_1 < X_3 \,.$$

Each of these events can be computed using the results from Part (a). Specifically

$$\begin{aligned}
P\{X_1 < X_2 < X_3\} &= \frac{\lambda_1 \lambda_2}{(\lambda_2 + \lambda_3)(\lambda_1 + \lambda_2 + \lambda_3)} \\
P\{X_2 < X_1 < X_3\} &= \frac{\lambda_1 \lambda_2}{(\lambda_1 + \lambda_3)(\lambda_1 + \lambda_2 + \lambda_3)} \,,
\end{aligned}$$

so that

$$P\{\max(X_1, X_2, X_3) = X_3\} = P\{X_1 < X_2 < X_3\} + P\{X_2 < X_1 < X_3\} \,.$$

Thus using these results we can calculate our desired probability

$$
\begin{aligned}
P\{X_1 < X_2 | \max(X_1, X_2, X_3) = X_3\} &= \frac{\frac{\lambda_1\lambda_2}{(\lambda_2+\lambda_3)(\lambda_1+\lambda_2+\lambda_3)}}{\frac{\lambda_1\lambda_2}{(\lambda_2+\lambda_3)(\lambda_1+\lambda_2+\lambda_3)} + \frac{\lambda_1\lambda_2}{(\lambda_1+\lambda_3)(\lambda_1+\lambda_2+\lambda_3)}} \\
&= \frac{\lambda_1 + \lambda_3}{\lambda_1 + \lambda_2 + 2\lambda_3}.
\end{aligned}
$$

Part (c): We want to evaluate $E[\max(X_i)|X_1 < X_2 < X_3]$. From the definition of conditional expectation we have that (if we define the random variable M to be $M = \max(X_i)$)

$$
E[\max(X_i)|X_1 < X_2 < X_3] = \int_0^\infty m p_M(m|X_1 < X_2 < X_3) dm\,.
$$

So that we see that to evaluate this expectation we need to be able to compute the conditional density $p_M(m|X_1 < X_2 < X_3)$. Using the definition of conditional density we have that

$$
p_M(m|X_1 < X_2 < X_3) = \frac{P\{M = m, X_1 < X_2 < X_3\}}{P\{X_1 < X_2 < X_3\}}\,.
$$

From Part (a) of this problem we know how to compute $P\{X_1 < X_2 < X_3\}$, so our problem becomes how to calculate the expression $P\{M = m, X_1 < X_2 < X_3\}$. We find that

$$
\begin{aligned}
P\{M = m, X_1 < X_2 < X_3\} &= P\{\max(X_1, X_2, X_3) = m, X_1 < X_2 < X_3\} \\
&= P\{X_3 = m, X_1 < X_2 < X_3\} \\
&= P\{X_3 = m, X_1 < X_2 < m\} \\
&= P\{X_3 = m\}P\{X_1 < X_2 < m\}\,,
\end{aligned}
$$

using the independence of the variables X_i. Since X_3 is an exponential random variable we know that $P\{X_3 = m\} = \lambda_3 e^{-\lambda_3 m}$. We can also compute the second probability in the product above as

$$
\begin{aligned}
P\{X_1 < X_2 < m\} &= \int\int_{\Omega:\{X_1<X_2<m\}} p_{X_1}(x_1)p_{X_2}(x_2)dx_2dx_1 \\
&= \int_{x_2=0}^{m}\int_{x_1=0}^{x_2} \lambda_1 e^{-\lambda_1 x_1}\lambda_2 e^{-\lambda_2 x_2} dx_1 dx_2\,.
\end{aligned}
$$

We can perform the above integrations one at a time from the outside inward. The first integral is with respect to x_1 and the second is with respect to x_2. When this is done we obtain

$$
P\{X_1 < X_2 < m\} = \frac{\lambda_1}{\lambda_1 + \lambda_2} - e^{-\lambda_2 m} + \frac{\lambda_2}{\lambda_1 + \lambda_2} e^{-(\lambda_1+\lambda_2)m}\,.
$$

See the **Mathematica** file `chap_5_prob_12.nb` for this algebra. Now multiplying this expression by $P\{X_3 = m\} = \lambda_3 e^{-\lambda_3 m}$ we have (doing the steps very slowly) that

$$
\begin{aligned}
P\{M = m, X_1 < X_2 < X_3\} &= \frac{\lambda_1}{\lambda_1 + \lambda_2}\cdot\lambda_3 e^{-\lambda_3 m} - \lambda_3 e^{-(\lambda_2+\lambda_3)m} + \frac{\lambda_2\lambda_3}{\lambda_1 + \lambda_2} e^{-(\lambda_1+\lambda_2+\lambda_3)m} \\
&= \frac{\lambda_1}{\lambda_1 + \lambda_2}\cdot\lambda_3 e^{-\lambda_3 m} - \frac{\lambda_3}{\lambda_2 + \lambda_3}(\lambda_2 + \lambda_3)e^{-(\lambda_2+\lambda_3)m} \\
&+ \frac{\lambda_2\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3}(\lambda_1 + \lambda_2 + \lambda_3)e^{-(\lambda_1+\lambda_2+\lambda_3)m}\,.
\end{aligned}
$$

The manipulations performed above were to enable us to recognize the above expression as a linear combination of exponential random variables. We explicitly demonstrate this fact by introducing normalized exponential random variables in the second line above. With this representation, taking the expectation of this expression is simple to do. At this point in the calculation we do *not* yet need divide by $P\{X_1 < X_2 < X_3\}$ since it is just a scalar multiplier of the above and can be done after the integration. Performing the integration over m in the above expression and remembering the expression for the mean of a exponentially distributed random variable we find that

$$\begin{aligned} E[\max(X_i)|X_1 < X_2 < X_3] &\propto \frac{\lambda_1}{\lambda_1+\lambda_2} \cdot \frac{1}{\lambda_3} - \frac{\lambda_3}{\lambda_2+\lambda_3} \cdot \frac{1}{\lambda_2+\lambda_3} \\ &+ \frac{\lambda_2\lambda_3}{\lambda_1+\lambda_2+\lambda_3} \cdot \frac{1}{\lambda_1+\lambda_2+\lambda_3} . \end{aligned}$$

So that when we divide by $P\{X_1 < X_2 < X_3\}$ we find that we obtain for $E[\max(X_i)|X_1 < X_2 < X_3]$ the following

$$\frac{(\lambda_2+\lambda_3)(\lambda_1+\lambda_2+\lambda_3)}{\lambda_1\lambda_2} \times \left(\frac{\lambda_1}{\lambda_3(\lambda_1+\lambda_2)} - \frac{\lambda_3}{(\lambda_2+\lambda_3)^2} + \frac{\lambda_3}{(\lambda_1+\lambda_2+\lambda_3)^2} \right) .$$

Which one might be able to simplify further but we stop here.

Part (d): To evaluate $E[\max(X_1, X_2, X_3)]$ we will use the mins for maxs identity which is

$$\begin{aligned} \max(X_1, X_2, X_3) &= \sum_{i=1}^{3} X_i - \sum_{i<j} \min(X_i, X_j) + \sum_{i<j<k} \min(X_i, X_j, X_k) \\ &= X_1 + X_2 + X_3 - \min(X_1, X_2) - \min(X_1, X_3) - \min(X_2, X_3) \\ &+ \min(X_1, X_2, X_3) , \end{aligned}$$

Sine the X_i's are independent exponential distributed random variables the random variables which are the minimization's of such random variables are themselves exponential random variables with failure rates given by the sum of the failure rates of their various components. Specifically the random variable $\min(X_1, X_2)$ is an exponential random variable with failure rate give by $\lambda_1 + \lambda_2$. Taking the expectation of this expression we find that

$$\begin{aligned} E[\max(X_1, X_2, X_3)] &= \frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \frac{1}{\lambda_3} - \frac{1}{\lambda_1+\lambda_2} - \frac{1}{\lambda_1+\lambda_3} - \frac{1}{\lambda_2+\lambda_3} \\ &+ \frac{1}{\lambda_1+\lambda_2+\lambda_3} . \end{aligned}$$

Note that by using the various pieces computed in this problem one could compute $E[\max(X_i)]$ by conditioning on events like in Part (c). For example

$$\begin{aligned} E[\max(X_i)] &= E[\max(X_i)|X_1 < X_2 < X_3]P\{X_1 < X_2 < X_3]\} \\ &+ E[\max(X_i)|X_1 < X_3 < X_2]P\{X_1 < X_3 < X_2]\} \\ &+ E[\max(X_i)|X_2 < X_3 < X_1]P\{X_2 < X_3 < X_1]\} \\ &+ \cdots \end{aligned}$$

The two approaches should give the same answer but the second seems more tedious.

Exercise 14 (are we less than c?)

Part (a): We have from the definition of conditional expectation that

$$E[X|X < c] = \int x p_X(x|X < c) dx \,.$$

Now the density function in the above expression is by definition given by

$$p_X(x|X < c) = \frac{P\{X = x, X < c\}}{P\{X < c\}} \,.$$

Now in the above expression the numerator is given by

$$P\{X = x, X < c\} = H(-(x - c))p_X(x) = H(c - x)p_X(x) \,.$$

here $H(\cdot)$ is the Heaviside function and enforces the requirement that $X < c$. The denominator is then given by the cumulative distribution function for an exponential random variable i.e.

$$P\{X < c\} = F_X(c) = 1 - \lambda e^{-\lambda c} \,.$$

Using both of these expressions the above becomes

$$\begin{aligned}
\int_0^\infty x \left(\frac{H(c - x)p_X(x)}{1 - \lambda e^{-\lambda c}} \right) dx &= \frac{1}{1 - \lambda e^{-\lambda c}} \int_0^c x \lambda e^{-\lambda x} dx \\
&= \frac{1}{\lambda(1 - \lambda e^{-\lambda c})} (1 - e^{-\lambda c} - c\lambda e^{-\lambda c}) \,,
\end{aligned}$$

for the requested expression.

Part (b): We now want to find the $E[X|X < c]$ by conditioning on $E[X]$. To begin we have

$$E[X] = E[X|X < c]P\{X < c\} + E[X|X > c]P\{X > c\} \,.$$

Now from the properties of an exponential distribution we know several of these pieces. We know that

$$\begin{aligned}
E[X] &= \frac{1}{\lambda} \\
P\{X < c\} &= 1 - \lambda e^{-\lambda c} \\
P\{X > c\} &= \lambda e^{-\lambda c} \,.
\end{aligned}$$

Thus using all of this information in the first expression derived in this part of this problem we find

$$\frac{1}{\lambda} = E[X|X < c](1 - \lambda e^{-\lambda c}) + E[X|X > c](\lambda e^{-\lambda c}) \,.$$

Now by the memoryless property of the exponential distribution we have the key observation that

$$E[X|X > c] = E[X + c] = E[X] + c = \frac{1}{\lambda} + c \,,$$

so the formula above becomes

$$\frac{1}{\lambda} = E[X|X < c](1 - \lambda e^{-\lambda c}) + \left(\frac{1}{\lambda} + c\right)(\lambda e^{-\lambda c}),$$

which when we solve for $E[X|X < c]$ we find that

$$E[X|X < c] = \frac{1 - (1 + c\lambda)e^{-\lambda c}}{\lambda(1 - \lambda e^{-\lambda c})},$$

the same expression as in Part (a).

Exercise 18 (mins and maxes of same rate exponentials)

Part (a): Since $X_{(1)}$ is defined as $X_{(1)} = \min(X_1, X_2)$ it is an exponential random variable with rate 2μ so we find that

$$E[X_{(1)}] = \frac{1}{2\mu}.$$

Part (b): Since $X_{(1)}$ is defined as $X_{(1)} = \min(X_1, X_2)$ it is an exponential random variable with rate 2μ so we find that

$$\text{Var}(X_{(1)}) = \frac{1}{(2\mu)^2} = \frac{1}{4\mu^2}.$$

Part (c): To compute moments of $X_{(2)} = \max(X_1, X_2)$ recall from the discussion on order statistics that the density function for the ith order (of n independent random variables with density/distribution function f/F) is given by

$$f_{X_{(i)}}(x) = \frac{n!}{(n-i)!(i-1)!} f(x)F(x)^{i-1}(1 - F(x))^{n-i},$$

so the distribution of the maximum $X_{(n)}$ is given by

$$f_{X_{(n)}}(x) = \frac{n!}{0!(n-1)!} f(x)F(x)^{n-1} = nf(x)F(x)^{n-1}.$$

When there are only two independent exponential random variables this becomes

$$f_{X_{(2)}}(x) = 2\mu e^{-\mu x} - 2\mu e^{-2\mu x}.$$

Notice that this is a linear combination of two exponential densities i.e. one with rate μ and one with rate 2μ. This observation helps in computing various quantities. For example for the expectation we find that

$$E[X_{(2)}] = 2\left(\frac{1}{\mu}\right) - \frac{1}{2\mu} = \frac{3}{2\mu}.$$

We note that this problem could also be solved by using the "mins" for "maxes" relationship which for two variables is given by

$$\max(X_1, X_2) = X_1 + X_2 - \min(X_1, X_2)\,, \tag{2}$$

where we know that since X_i is distributed as an exponential random variable with rate μ the random variable $\min(X_1, X_2)$ is an exponential random variable with rate 2μ. Thus taking the expectation of the above (and using the known value of the expectation for an exponential random variable) we find that

$$E[\max(X_1, X_2)] = \frac{1}{\mu} + \frac{1}{\mu} - \frac{1}{2\mu} = \frac{3}{2\mu}\,,$$

as before. The variance calculation using this approach would be more complicated because it would involve products of terms containing X_1 and $\min(X_1, X_2)$ which are not independent and would have to be computed in some way. Finally a third way to compute this expectation is to recall that by the memoryless property of the exponential distribution that the random variable $X_{(2)}$ is related to the random variable $X_{(1)}$ by an "offset" random variable (say A) in such a way that $X_{(2)} = X_{(1)} + A$, where A is an exponential random variable with rate μ and $X_{(1)}$ and A are independent. Thus we then find using this method that

$$E[X_{(2)}] = E[X_{(1)}] + E[A] = \frac{1}{2\mu} + \frac{1}{\mu} = \frac{3}{2\mu}\,,$$

the same as earlier.

Part (d): To compute the variance of $X_{(2)}$ we can use

$$\mathrm{Var}(X_{(2)}) = E[X_{(2)}^2] - E[X_{(2)}]^2\,,$$

and we can calculate $E[X_{(2)}^2]$ from the density given in Part (c). Remembering that the second moment of an exponential random variable (say X) with failure rate λ is given by

$$E[X^2] = \frac{2}{\lambda^2}\,,$$

given the density for $X_{(2)}$ derived above we find that

$$E[X_{(2)}^2] = 2\left(\frac{2}{\mu^2}\right) - \frac{2}{(2\mu)^2} = \frac{7}{2\mu^2}\,.$$

Using this result we see that $\mathrm{Var}(X_{(2)})$ is given by

$$\mathrm{Var}(X_{(2)}) = \frac{7}{2\mu^2} - \frac{9}{4}\frac{1}{\mu^2} = \frac{5}{4\mu^2}\,.$$

Another way to compute this variance is to recall that by the memoryless property of the exponential distribution that the random variable $X_{(2)}$ is related to the random variable $X_{(1)}$ by an "offset" random variable (say A) in such a way that $X_{(2)} = X_{(1)} + A$, where A is an exponential random variable with rate μ and $X_{(1)}$ and A are independent. Thus we then find using this method that

$$\mathrm{Var}(X_{(2)}) = \mathrm{Var}(X_{(1)}) + \mathrm{Var}(A) = \frac{1}{(2\mu)^2} + \frac{1}{\mu^2} = \frac{5}{4\mu}\,,$$

the same as before.

Exercise 19 (mins and maxes of different rate exponentials)

Part (a): Using much of the discussion from Exercise 18 we have that $X_{(1)}$ is an exponential random variable with rate $\mu_1 + \mu_2$, so that

$$E[X_{(1)}] = \frac{1}{\mu_1 + \mu_2}.$$

Part (b): As in Part (a) the variance of $X_{(1)}$ is simple to calculate and we find that

$$\text{Var}(X_{(1)}) = \frac{2}{(\mu_1 + \mu_2)^2}.$$

Part (c): To compute $E[X_{(2)}]$ we condition on whether $X_1 < X_2$ or not. We have

$$\begin{aligned}
E[X_{(2)}] &= E[X_{(2)}|X_1 < X_2]P\{X_1 < X_2\} + E[X_{(2)}|X_1 > X_2]P\{X_1 > X_2\} \\
&= E[X_{(2)}|X_1 < X_2]\left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right) + E[X_{(2)}|X_1 > X_2]\left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right).
\end{aligned}$$

We will now evaluate $E[X_{(2)}|X_1 < X_2]$. We first define the random variable $M = \max(X_1, X_2)$ and use the definition of conditional expectation. We have

$$E[X_{(2)}|X_1 < X_2] = \int_0^\infty m p_M(m|X_1 < X_2) dm.$$

The conditional probability distribution $p_M(m|X_1 < X_2)$ is given by some simple manipulations. We find

$$\begin{aligned}
p_M(m|X_1 < X_2) &= \frac{P\{M = m, X_1 < X_2\}}{P\{X_1 < X_2\}} \\
&= \frac{P\{\max(X_1, X_2) = m, X_1 < X_2\}}{P\{X_1 < X_2\}} \\
&= \frac{P\{X_2 = m, X_1 < X_2\}}{P\{X_1 < X_2\}} \\
&= \frac{P\{X_2 = m, X_1 < m\}}{P\{X_1 < X_2\}} \\
&= \frac{P\{X_2 = m\}P\{X_1 < m\}}{P\{X_1 < X_2\}},
\end{aligned}$$

where we have used the independence of X_1 and X_2 in the last step. From earlier we know that the denominator of the above is given by $P\{X_1 < X_2\} = \frac{\lambda_1}{\lambda_1+\lambda_2}$, and the numerator is given by

$$P\{X_2 = m\}P\{X_1 < m\} = \lambda_2 e^{-\lambda_2 m}(1 - e^{-\lambda_1 m}) = \lambda_2 e^{-\lambda_2 m} - \lambda_2 e^{-(\lambda_1+\lambda_2)m}.$$

Combining all of this information we find that our probability density is given by

$$p_M(m|X_1 < X_2) = \left(\frac{\lambda_1 + \lambda_2}{\lambda_1}\right)(\lambda_2 e^{-\lambda_2 m} - \lambda_2 e^{-(\lambda_1+\lambda_2)m}).$$

Multiplying by m and integrating we can use the fact that $p_M(m|X_1 < X_2)$ is a linear combination of exponential densities to find that

$$E[X_{(2)}|X_1 < X_2] = \frac{\lambda_1 + \lambda_2}{\lambda_1 \lambda_2} - \frac{\lambda_2}{\lambda_1}\left(\frac{1}{\lambda_1 + \lambda_2}\right).$$

By the same arguments (effectively exchanging 1 with 2) we find that

$$E[X_{(2)}|X_1 > X_2] = \frac{\lambda_1 + \lambda_2}{\lambda_1 \lambda_2} - \frac{\lambda_1}{\lambda_2}\left(\frac{1}{\lambda_1 + \lambda_2}\right).$$

So combining these two results we can finally obtaining $E[X_{(2)}]$ as

$$E[X_{(2)}] = \frac{1}{\lambda_1} + \frac{1}{\lambda_2} - \frac{1}{\lambda_1 + \lambda_2}.$$

Note that this could also have been obtained (with a lot less work) by using the mins for max identity given by Equation 2.

Part (d): We want to evaluate $\text{Var}(X_{(2)})$, which we will do by evaluating $E[X_{(2)}^2] - E[X_{(2)}]^2$ using the results from Part (c) above. As before we can evaluate $E[X_{(2)}^2]$ by conditioning on whether $X_1 < X_2$ or not. From the expression for $p_M(m|X_1 < X_2)$ derived in Part (c) we find that

$$E[X_{(2)}^2|X_1 < X_2] = \frac{\lambda_1 + \lambda_2}{\lambda_1}\left(\frac{2}{\lambda_2^2} - \frac{\lambda_2}{\lambda_1 + \lambda_2} \cdot \frac{2}{(\lambda_1 + \lambda_2)^2}\right),$$

and in the same way we then have

$$E[X_{(2)}^2|X_1 > X_2] = \frac{\lambda_1 + \lambda_2}{\lambda_2}\left(\frac{2}{\lambda_1^2} - \frac{\lambda_1}{\lambda_1 + \lambda_2} \cdot \frac{2}{(\lambda_1 + \lambda_2)^2}\right).$$

So that using the results above, the expectation of $X_{(2)}^2$ is given by

$$E[X_{(2)}^2] = \frac{2}{\lambda_1^2} + \frac{2}{\lambda_2^2} - \frac{2}{(\lambda_1 + \lambda_2)^2}.$$

Giving finally the desired result for $\text{Var}(X_{(2)})$ of

$$\begin{aligned}
\text{Var}(X_{(2)}) &= \frac{2}{\lambda_1^2} + \frac{2}{\lambda_2^2} - \frac{2}{(\lambda_1 + \lambda_2)^2} - \left(\frac{1}{\lambda_1} + \frac{1}{\lambda_2} - \frac{1}{\lambda_1 + \lambda_2}\right)^2 \\
&= \frac{1}{\lambda_1^2} + \frac{1}{\lambda_2^2} - \frac{3}{(\lambda_1 + \lambda_2)^2},
\end{aligned}$$

after some algebra.

Exercise 21 (waiting time at a two station queue)

For simplicity let X_i for $i = 1, 2$ be an exponential random variables with parameter μ_1, and Y_i for $i = 1, 2$ be exponential random variables with parameter μ_2. When we enter the system we will have to first wait for the first server to finish. By the memoryless property of exponential random variables we have that the amount of time we have to wait for this to happen will be X_1. At this point we will be at server ones station while the person ahead of us will be at sever number two's station. The total time for us to wait until we move to server number two's station is given by the maximum amount of time spent at either server number one or server number two and again using the memory less property of the exponential distribution is represented as the random variable $\max(X_2, Y_1)$. Once this amount of time has completed we will be a the second server where we will have to wait an additional Y_2 amount of time. Thus if T denotes the random variable representing the total amount of time we will have to wait we have that

$$T = X_1 + \max(X_2, Y_1) + Y_2 \,.$$

To evaluate this we need to compute the distribution function of a random variable A defined by $A = \max(X_2, Y_1)$ when X_2 and Y_2 are distributed as above. To do this consider the distribution function of this random variable where we drop the subscripts on X and Y we have

$$F(a) = P\{\max(X, Y) \leq a\} = \int\int_{\Omega_a} f(x, y) dx dy$$

where Ω_a is the set of points in the XY plane where $\max(X, Y) \leq a$. If we imagine a the random variable X along the x axis of a Cartesian grid and the random variable Y along the y axis of a Cartesian grid then the set of points where $\max(X, Y) \leq a$ is $0 \leq X \leq a$ and $0 \leq Y \leq a$. Thus we can evaluate $F(a)$ (using the independence of X and Y) as

$$\begin{aligned}
F(a) &= \int_0^a \int_0^a f(x, y) dx dy \\
&= \int_0^a \int_0^a \mu_1 e^{-\mu_1 x} \mu_2 e^{-\mu_2 y} dx dy \\
&= \mu_1 \mu_2 \int_0^a e^{-\mu_1 x} dx \int_0^a e^{-\mu_2 y} dy \\
&= (1 - e^{-\mu_1 a})(1 - e^{-\mu_2 a}) \\
&= 1 - e^{-\mu_1 a} - e^{-\mu_2 a} + e^{-(\mu_1 + \mu_2) a}
\end{aligned}$$

Thus our density function is then given by

$$f(a) = \mu_1 e^{-\mu_1 a} + \mu_2 e^{-\mu_2 a} - (\mu_1 + \mu_2) e^{-(\mu_1 + \mu_2) a} \,.$$

Showing easily that the expectation of the random variable A is given by

$$E[A] = \frac{1}{\mu_1} + \frac{1}{\mu_2} - \frac{1}{\mu_1 + \mu_2} \,.$$

Thus we can now evaluate the expected length of time in the system T as

$$\begin{aligned}
E[T] &= E[X_1] + E[\max(X_2, Y_1)] + E[Y_2] \\
&= \left(\frac{1}{\mu_1}\right) + \left(\frac{1}{\mu_1} + \frac{1}{\mu_2} - \frac{1}{\mu_1 + \mu_2}\right) + \left(\frac{1}{\mu_2}\right) \\
&= \frac{2}{\mu_1} + \frac{2}{\mu_2} - \frac{1}{\mu_1 + \mu_2} \,.
\end{aligned}$$

Exercise 22 (more waiting at a two station queue)

In the same way as in the previous problem we will let X_i be an exponential random variable with parameter μ_1 and Y_i be an exponential random variable with parameter μ_2. Then to have server one become *available* we must wait an amount $\max(X_1, Y_1)$. After this amount of time you move to server number one. To move to server number two you must wait another $\max(X_2, Y_2)$ amount of time. Finally you must wait for server number two to finish (an additional amount of time Y_3) serving you. Thus the total time T spent in the system is given by

$$T = \max(X_1, Y_1) + \max(X_2, Y_2) + Y_3 .$$

So the expected time in the station is given by (using the results from the previous problem)

$$\begin{aligned} E[T] &= \left(\frac{1}{\mu_1} + \frac{1}{\mu_2} - \frac{1}{\mu_1 + \mu_2}\right) + \left(\frac{1}{\mu_1} + \frac{1}{\mu_2} - \frac{1}{\mu_1 + \mu_2}\right) + \left(\frac{1}{\mu_2}\right) \\ &= \frac{2}{\mu_1} + \frac{3}{\mu_2} - \frac{2}{\mu_1 + \mu_2} . \end{aligned}$$

Exercise 23 (flashlight batteries)

By the memory less property of the exponential distribution, when two batteries are being used simultaneously the lifetime of either battery is given by an exponential distribution that does not depend on how long the battery has been operational up to that point. Thus if X and Y are exponential random variables denoting the life expectancy of two batteries with the same parameter μ the probability that X will fail before Y is given by

$$\begin{aligned} P\{X < Y\} &= \int\int_{\Omega_{X<Y}} f_{X,Y}(x,y)dxdy \\ &= \int_{x=0}^{\infty}\int_{y=0}^{x} f_{X,Y}(x,y)dydx \\ &= \int_{x=0}^{\infty}\int_{y=0}^{x} f_{X,Y}(x,y)dydx \\ &= \int_{x=0}^{\infty}\int_{y=0}^{x} \mu e^{-\mu x}\mu e^{-myy}dydx \\ &= \mu\int_{x=0}^{\infty} e^{-\mu x}(1 - e^{-\mu x})dx \\ &= \int_{x=0}^{\infty} \mu e^{-\mu x}dx - \frac{1}{2}\int_{x=0}^{\infty}(2\mu)e^{-2\mu x}dx \\ &= 1 - \frac{1}{2} = \frac{1}{2} . \end{aligned}$$

Thus when two batteries are in place due to the memoryless property of the exponential distribution it is equally likely that either one of them will expire first. With this information we can calculate the various probabilities.

Part (a): $P\{X = n\}$ is the probability that the last non failed battery is number n. Since when the n the battery is placed in the flashlight, with probability $1/2$ it will last longer than the other battery. Thus $P\{X = n\} = 1/2$.

Part (b): $P\{X = 1\}$ is the probability that when the last working battery is observed it happens to be the first battery on the battery list. This will happen if the first battery lasts longer than the second battery, the third battery, the fourth batter, etc. up to the nth battery. Since the event that the first battery lasts longer than these $n - 1$ other batteries this will happen with probability $(1/2)^{n-1}$, thus

$$P\{X = 1\} = \left(\frac{1}{2}\right)^{n-1} .$$

Part (c): To evaluate $P\{X = i\}$ means that when the ith battery is placed in the flashlight it proceeds to last longer than the remaining $I, i + 1, i + 2, \cdots, n$ other batteries. Here I is the index of the other battery in the flashlight when the ith battery is placed in the flashlight $(1 \le I \le i - 1)$. Thus the ith battery has $n - (i + 1) + 1 + 1 = n - i + 1$ comparisons which it must win. Thus

$$P\{X = i\} = \left(\frac{1}{2}\right)^{n-i+1} \qquad 2 \le i \le n$$

The reason this expression does not work for $i = 1$ is that in that case there is no battery "before" this one to compare to. Thus the index in the above expression is one two large for the case $i = 1$.

Part (d): The random variable T (representing the entire lifetime of our flashlight) can be expressed in terms of sums of random variables representing the serial lifetimes of two individual batteries. For example, the first two batteries will produced a functional flashlight until the first one burns out. If we let T_1 and T_2 represent the exponential random variables representing the lifetime of the first and second battery then the flashlight will function for a random amount of time $F_1 = \min(T_1, T_2)$. When one of these two batteries burns out it is immediately replaces. By the memoryless property of the exponential distribution the battery that is *not* replaced has a distribution as if it just started at this replacement time. The next increment of time before the *fourth* battery is used is given by $F_2 = \min(T_*, T_3)$, where T_* is the battery not replaced earlier. Note that each of these minimizations is itself a random variable with a rate given by 2μ. This process continues $n - 1$ times. Thus if T is the lifetime of the total flashlight system we see that it is given by

$$T = \sum_{i=1}^{n-1} F_i .$$

Since for each F_i we have $E[F_i] = \frac{1}{2\mu}$, the expectation of T is given by

$$E[T] = \sum_{i=1}^{n-1} E[F_i] = \frac{n-1}{2\mu} .$$

Part (e): Since T is the sum of $n-1$ exponential random variables with rates 2μ the distribution of T is given by a gamma distribution with parameters $n-1$ and 2μ, specifically

$$f_T(t) = 2\mu e^{-2\mu t}\frac{(2\mu t)^{n-2}}{(n-2)!}.$$

Exercise 24 (Laplace distributions)

Part (a): We want to show that the given f is a density function. We will integrate and see if we obtain unity. We have

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x)dx &= \int_{-\infty}^{0} \frac{\lambda}{2}e^{\lambda x}dx + \int_{0}^{\infty} \frac{\lambda}{2}e^{-\lambda x}dx \\
&= \frac{\lambda}{2}\left(\left.\frac{e^{\lambda x}}{\lambda}\right|_{-\infty}^{0}\right. + \frac{\lambda}{2}\left(\left.\frac{e^{-\lambda x}}{(-\lambda)}\right|_{0}^{\infty}\right. \\
&= \frac{1}{2}(1) - \frac{1}{2}(-1) = 1\,.
\end{aligned}$$

Part (b): The distribution function $F(\cdot)$ is given by

$$F(x) = \int_{-\infty}^{x} \frac{1}{2}\lambda e^{\lambda\xi}d\xi = \frac{1}{2}\left.e^{\lambda\xi}\right|_{-\infty}^{x} = \frac{1}{2}e^{\lambda x} \quad \text{for} \quad x < 0\,,$$

and

$$F(x) = \frac{1}{2} + \int_{0}^{x} \frac{1}{2}\lambda e^{-\lambda\xi}d\xi = \frac{1}{2} + \frac{\lambda}{2}\left.\frac{e^{\lambda\xi}}{(-\lambda)}\right|_{0}^{x} = 1 - \frac{1}{2}e^{-\lambda x} \quad \text{for} \quad x > 0\,,$$

Part (c): We are told that X and Y are exponential random variables then we want to prove that $X - Y$ is given by a Laplace (double exponential) distribution. To do this define the random variable $Z \equiv X - Y$, which has a distribution function given by the convolution of the random variables X and $-Y$. This means that

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(\xi)f_{-Y}(z-\xi)d\xi\,.$$

Now the density $f_{-Y}(y)$ is given by

$$f_{-Y}(y) = \begin{cases} 0 & y > 0 \\ \lambda e^{\lambda y} & y < 0 \end{cases}$$

So to evaluate our convolution given above we begin by letting $z < 0$ and then since $f_{-Y}(z-\xi) = f_{-Y}(-(\xi-z))$, plotted as a function of ξ this is the function $f_{-Y}(-\xi)$ shifted to the *right* by z. Plotting the function $f_{-Y}(-(\xi - z))$, for $z = -1$ and $z = +1$ we have the figure **??**. With this understanding if $z < 0$ our function $f_Z(z)$ is given by

$$f_Z(z) = \int_{0}^{\infty} \lambda e^{-\lambda\xi}\cdot\lambda e^{\lambda(z-\xi)}d\xi = \frac{\lambda}{2}e^{\lambda z} \quad \text{for} \quad z < 0\,,$$

after some algebra. If $z > 0$ then $f_Z(z)$ is given by

$$f_Z(z) = \int_z^\infty \lambda e^{-\lambda\xi} \cdot \lambda e^{\lambda(z-\xi)} d\xi = \frac{\lambda}{2} e^{-\lambda z} \quad \text{for} \quad z < 0\,,$$

Combining these results we find that

$$f_Z(z) = \begin{cases} \frac{\lambda}{2} e^{\lambda z} & z < 0 \\ \frac{\lambda}{2} e^{-\lambda z} & z > 0 \end{cases}$$

showing that Z is a Laplace or double exponential random variable.

Part (d): Now to show that IX is a Laplace random variable we remember that $I = \pm 1$ both with probability $1/2$. Defining $Y \equiv IX$ the probability distribution of Y can be found by conditioning on the value of I. We have

$$\begin{aligned} p_Y(y) &= P\{IX = y | I = -1\} P\{I = -1\} + P\{IX = y | I = +1\} P\{I = +1\} \\ &= P\{IX = y | I = -1\} \frac{1}{2} + P\{IX = y | I = +1\} \frac{1}{2} \\ &= P\{X = -y | I = -1\} \frac{1}{2} + P\{X = y | I = +1\} \frac{1}{2} \\ &= P\{X = -y\} \frac{1}{2} + P\{X = y\} \frac{1}{2}\,, \end{aligned}$$

by the independence of X and I. Since

$$P\{X = -y\} = \begin{cases} 0 & y > 0 \\ \lambda e^{\lambda y} & y < 0 \end{cases}$$

we have that $p_Y(y)$ is given by

$$p_Y(y) = \begin{cases} \frac{1}{2} \lambda e^{\lambda y} & y < 0 \\ \frac{1}{2} \lambda e^{-\lambda y} & y > 0 \end{cases}$$

or a Laplace random variable.

Part (e): We define the random variable W as

$$W = \begin{cases} X & I = 1 \\ -Y & I = -1 \end{cases}$$

then this probability distribution can be calculated as in a Part (d) of this problem. For we have

$$p_W(w) = \frac{1}{2} p_X(w) + \frac{1}{2} p_{-Y}(w) = \frac{1}{2} H(w) \lambda e^{-\lambda w} + \frac{1}{2} H(-w) \lambda e^{\lambda w}\,,$$

with $H(\cdot)$ the Heaviside function. From which we see that the random variable W is a Laplace random variable.

Exercise 29 (replacement kidneys)

Part (a): From the information given the time till a new kidney's arrival, T, is distributed as an exponential random variable with rate λ. Thus as long as A is still living when it finally arrives she will receive it. If we denote L_A as the random variable representing the lifetime of A then A will be alive if $L_A > T$. This event happens with probability

$$\frac{\lambda}{\mu_A + \lambda},$$

with μ_A the exponential rate of the lifetime of A. This result as derived in the section of the book entitled "Further Properties of the Exponential Distribution".

Part (b): For B to receive a new kidney we must have had A expire, and B still alive when the kidney arrives. Mathematically if we denote L_B as the random variable the lifetime of B this is the event that

$$L_B > L_A \quad \text{and} \quad L_B > T .$$

By the independence of these random variables the probability of this event is

$$P\{L_B > L_A\}P\{L_B > T\} = \left(\frac{\mu_A}{\mu_A + \mu_B}\right)\left(\frac{\lambda}{\mu_B + \lambda}\right) .$$

Problem 30 (pet lifetimes)

Let L be the random variable denoting the additional lifetime of the surviving pet. We can compute the expectation of L by conditioning on which pet dies first. If we let L_D and L_C be random variables denoting the lifetimes of the dog and cat respectively, then the event that the dog is the first do die is mathematically stated as $L_D < L_C$. With these definitions we can compute $E[L]$ as

$$\begin{aligned}
E[L] &= E[L|L_D < L_C]P\{L_D < L_C\} + E[L|L_D > L_C]P\{L_D > L_C\} \\
&= E[L|L_D < L_C]\left(\frac{\lambda_d}{\lambda_d + \lambda_c}\right) + E[L|L_D > L_C]\left(\frac{\lambda_c}{\lambda_d + \lambda_c}\right) \\
&= E[L_C]\left(\frac{\lambda_d}{\lambda_d + \lambda_c}\right) + E[L_D]\left(\frac{\lambda_c}{\lambda_d + \lambda_c}\right) \\
&= \frac{1}{\lambda_c}\left(\frac{\lambda_d}{\lambda_d + \lambda_c}\right) + \frac{1}{\lambda_d}\left(\frac{\lambda_c}{\lambda_d + \lambda_c}\right) = \left(\frac{1}{\lambda_d + \lambda_c}\right)\left(\frac{\lambda_d}{\lambda_c} + \frac{\lambda_c}{\lambda_d}\right) .
\end{aligned}$$

In the above we have been able to make the substitution that $E[L|L_D < L_C] = E[L_C]$, because of the memoryless property of the exponential random variables involved.

Problem 31 (doctors appointments)

If the 1:00 appointment is over before the 1:30 appointment then the amount of time the 1:30 patient spends in the office is given by an exponential random variable with mean 30 minutes. If the 1:00 appointment runs over (into the 1:30 patients time slot) the 1:30 patient will spend more time in the office due to the fact that his appointment cannot start on time. Thus letting T_1 and T_2 be exponential random variables denoting the length of time each patient requires with the doctor, the expected time the 1:30 appointment spends at the doctors office (L) is given by conditioning on the length of time the 1:00 patient requires with the doctor. We find

$$E[L] = E[L|T_1 < 30]P\{T_1 < 30\} + E[L|T_1 > 30]P\{T_1 > 30\} .$$

We can compute each of these expressions. We find $E[L|T_1 < 30] = E[T_2] = 30$ minutes,

$$P\{T_1 < 30\} = 1 - e^{-(1/30)(30)} = 1 - e^{-1} ,$$

and

$$P\{T_1 > 30\} = e^{-(1/30)(30)} = e^{-1} .$$

To evaluate $E[L|T_1 > 30]$ we can write it as follows.

$$\begin{aligned} E[L|T_1 > 30] &= E[(T_1 - 30) + T_2|T_1 > 30] \\ &= E[T_1 - 30|T_1 > 30] + E[T_2|T_1 > 30] . \end{aligned}$$

By the memoryless property of the exponential random variable we have that

$$E[T_1 - 30|T_1 > 30] = E[T_1] = 30 .$$

Thus we find

$$E[L|T_1 > 30] = 30 + E[T_2|T_1 > 30] = 30 + 30 = 60 ,$$

and finally obtain

$$E[L] = 30(1 - e^{-1}) + 60(e^{-1}) = 41.03 ,$$

minutes for the expected amount of time the 1:30 patient spends at the doctors office.

Problem 34 (X given $X + Y$)

Part (a): Let begin by computing $f_{X|X+Y=c}(x|c)$. We have that

$$\begin{aligned} f_{X|X+Y=c}(x|c) &= P\{X = x|X + Y = c\} \\ &= \frac{P\{X = x, X + Y = c\}}{P\{X + Y = c\}} \\ &= \frac{P\{X = x, Y = c - x\}}{P\{X + Y = c\}} \\ &= \frac{P\{X = x\}P\{Y = c - x\}}{P\{X + Y = c\}} , \end{aligned}$$

By the independence of X and Y. Now defining the random variable Z as $Z = X + Y$, we see that Z has a distribution function given by the appropriate convolution

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(\xi)f_Y(z - \xi)d\xi ,$$

so that if $z < 0$ we have the above equal to

$$\begin{aligned} f_Z(z) &= \int_0^z \lambda e^{-\lambda\xi}\mu e^{-\mu(z-\xi)} \\ &= \frac{\lambda\mu}{\lambda-\mu}(e^{-\mu z} - e^{-\lambda z}), \end{aligned}$$

when we do the integration. Thus with this result we see that the expression for $f_{X|X+Y=c}(x|c)$ is given by

$$\begin{aligned} f_{X|X+Y=c}(x|c) &= \frac{(\lambda e^{-\lambda x})(\mu e^{-\mu(c-x)})}{\frac{\lambda\mu}{\lambda-\mu}(e^{-\mu c} - e^{-\lambda c})} \\ &= \frac{(\lambda-\mu)e^{-(\lambda-\mu)x}}{1 - e^{-(\lambda-\mu)c}} \quad \text{for} \quad 0 < x < c. \end{aligned}$$

The book gives a slightly different result (which I believe is a typo). I believe the result above is correct. One can integrate the above over the range $(0, c)$ to verify that this function integrates to one.

Part (b): From Part (a) of this problem we can integrate this expression to compute $E[X|X + Y = c]$. We find that

$$\begin{aligned} E[X|X+Y=c] &= \frac{(\lambda-\mu)}{(1 - e^{-(\lambda-\mu)c})} \int_0^c x e^{-(\lambda-\mu)x} dx \\ &= \frac{1}{\lambda-\mu} - \frac{ce^{-(\lambda-\mu)c}}{1 - e^{-(\lambda-\mu)c}}. \end{aligned}$$

Part (c): To find $E[Y|X + Y = c]$ we can use an idea like that in Exercise 14. Consider the expression $E[X + Y|X + Y = c]$, which we know must equal c since we are conditioning on the event $X + Y = c$. By linearity of the expectation it must also equals

$$E[X|X+Y=c] + E[Y|X+Y=c].$$

So solving for the desired $E[Y|X + Y = c]$ we see that

$$\begin{aligned} E[Y|X+Y=c] &= c - E[X|X+Y=c] \\ &= c - \frac{1}{\lambda-\mu} + \frac{ce^{-(\lambda-\mu)c}}{1 - e^{-(\lambda-\mu)c}} \\ &= c\left(\frac{1 - e^{-(\lambda-\mu)c}}{1 - e^{-(\lambda-\mu)c}}\right) - \frac{1}{\lambda-\mu} + \frac{ce^{-(\lambda-\mu)c}}{1 - e^{-(\lambda-\mu)c}} \\ &= \frac{c}{1 - e^{-(\lambda-\mu)c}} - \frac{1}{\lambda-\mu}. \end{aligned}$$

Problem 35 (equivalent definitions of a Poisson process)

We are asked to prove the equivalence of two definitions for a Poisson process. The first definition (Definition 5.1 in the book) is the following

The counting process $\{N(t), t \geq 0\}$ is said to be a Poisson process if:

i $N(0) = 0$

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

iii The number of events in any interval of length t has a Poisson distribution with mean λt. That is for $s, t \geq 0$, we have

$$P\{N(t+s) - N(s) = n\} = e^{-\lambda t}\frac{(\lambda t)^n}{n!} \quad n \geq 0$$

We want to show that this definition is equivalent to the following (which is Definition 5.3 in the book)

i $N(0) = 0$

ii $\{N(t), t \geq 0\}$ has stationary, independent increments

iii $P\{N(t) \geq 2\} = o(t)$

iv $P\{N(t) = 1\} = \lambda t + o(t)$

We begin by noting that both definitions require $N(0) = 0$. From (ii) in Definition 5.1 we have the required independent increments needed in Definition 5.3 (ii). From (iii) in Definition 5.1 we have that the distributions of $X(t_2 + s) - X(t_1 + s)$ is given by a Poisson distribution with mean $\lambda(t_2 - t_1)$ and the distribution of random variable $X(t_2) - X(t_1)$ is also given by a Poisson distribution with mean $\lambda(t_2 - t_1)$ showing that the process $\{N(t)\}$ also has *stationary* increments and thus satisfies the totality of Definition 5.3 (ii).

From (iii) in Definition 5.1 we have with $s = 0$ (and the fact that $N(0) = 0$) that

$$P\{N(t) = n\} = \frac{e^{-\lambda t}(\lambda t)^n}{n!}.$$

So that

$$\begin{aligned} P\{N(t) \geq 2\} &= \sum_{n=2}^{\infty} \frac{e^{-\lambda t}(\lambda t)^n}{n!} \\ &= e^{-\lambda t}\left[\sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} - 1 - \lambda t\right] \\ &= e^{-\lambda t}\left[e^{\lambda t} - 1 - \lambda t\right] \\ &= 1 - e^{-\lambda t} - \lambda t e^{-\lambda t}, \end{aligned}$$

which (we claim) is a function that is $o(t)$. To show that this is true consider the limit as t goes to zero. Thus we want to evaluate

$$\lim_{t \to 0} \frac{1 - e^{-\lambda t} - \lambda t e^{-\lambda t}}{t}.$$

Since this is an indeterminate limit of type $0/0$ we must use L'Hospital's rule which gives that the above limit is equal to the limit of the derivative of the top and bottom of the above or

$$\lim_{t \to 0} \frac{\lambda e^{-\lambda t} - \lambda e^{-\lambda t} + \lambda^2 t e^{-\lambda t}}{1} = \lambda - \lambda = 0\,.$$

Proving that this expression is $o(t)$ (since this limit equaling zero is the definition) and proving that $P\{N(t) \geq 2\} = o(t)$. The final condition required for Definition 5.3 is (iv). We have from Definition 5.1 (iii) that

$$P\{N(t) = 1\} = \frac{e^{-\lambda t}(\lambda t)}{1!} = \lambda t e^{-\lambda t}$$

To show that this expression has the correct limiting behavior as $t \to 0$, we first prove that

$$e^{-\lambda t} = 1 - \lambda t + o(t) \quad \text{as} \quad t \to 0\,,$$

Which we do by evaluating the limit

$$\lim_{t \to 0} \frac{e^{-\lambda t} - 1 + \lambda t}{t} = \lim_{t \to 0} \frac{-\lambda e^{-\lambda t} + \lambda}{1} = -\lambda + \lambda = 0\,.$$

Where we have used L'Hospital's rule again. With this result we see that

$$\begin{aligned} P\{N(t) = 1\} &= \lambda t(1 - \lambda t + o(t)) \\ &= \lambda t - \lambda^2 t^2 + o(t^2) \\ &= \lambda t + o(t)\,, \end{aligned}$$

showing the truth of condition (iv) in Definition 5.3.

Exercise 37 (helping Reb cross the highway)

At the point where Reb wants to cross the highway the number of cars that cross is a Poisson process with rate $\lambda = 3$, the probability that k appear is given by

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

Thus Reb will have no problem if no cars come during her crossing. If her crossing takes s second this will happen with probability

$$P\{N = 0\} = e^{-\lambda s} = e^{-3s}\,.$$

Note that this is the density function for a Poisson random variable (or the cumulative distribution function of a Poisson random variable with $n = 0$). This expression is tabulated for $s = 2, 5, 10, 20$ seconds in `chap_5_prob_37.m`.

Exercise 38 (helping a nimble Reb cross the highway)

Following the results from Exercise 29, Reb will cross unhurt, with probability

$$P\{N = 0\} + P\{N = 1\} = e^{-\lambda s} + e^{-\lambda s}(\lambda s) = e^{-3s} + 3se^{-3s}\,.$$

Not that this is the cumulative distribution function for a Poisson random variable. This expression is tabulated for $s = 5, 10, 20, 30$ seconds in `chap_5_prob_38.m`.

Exercise 40 (the sum of two Poisson processes)

We will first prove that the sum of two Poisson random variables is a Poisson random variable. Let X and Y be Poisson random variables with parameters λ_1 and λ_2 respectively. We can evaluate the distribution of $X+Y$ by computing the characteristic function of $X+Y$. Since X and Y are independent Poisson random variables the characteristic functions of $X+Y$ is given by

$$\begin{aligned}
\phi_{X+Y}(u) &= \phi_X(u)\phi_Y(u) \\
&= e^{\lambda_1(e^{iu}-1)}e^{\lambda_2(e^{iu}-1)} \\
&= e^{(\lambda_1+\lambda_2)(e^{iu}-1)}\,.
\end{aligned}$$

From the direct connection between characteristic functions to and probability density functions we see that the random variable $X+Y$ is a Poisson random variable with parameter $\lambda_1+\lambda_2$, the sum of the Poisson parameters of the random variables X and Y.

Now for the problem at hand, since $N_1(t)$ and $N_2(t)$ are both Poisson random variables with parameters $\lambda_1 t$ and $\lambda_2 t$ respectively, then from the above discussion the random variable $N(t)$ defined by $N_1(t)+N_2(t)$ is a Poisson random variable with parameter $\lambda_1 t + \lambda_2 t$ and thus has a probability of the event $N(t) = j$ given by

$$P\{N(t) = j\} = \frac{e^{-(\lambda_1 t+\lambda_2 t)}(\lambda_1 t + \lambda_2 t)^j}{j!} = \frac{e^{-(\lambda_1+\lambda_2)t}((\lambda_1+\lambda_2)t)^j}{j!}\,,$$

showing that $N(t)$ is a Poisson process with rate $\lambda_1+\lambda_2$.

Exercise 41 (the probability that N_1 hits first)

For this problem we are asked to evaluate

$$P\{N_1(t) = 1, N_2(t) = 0 | N_1(t) + N_2(t) = 1\}\,,$$

which we can do by using the definition of conditional probabilities as

$$\begin{aligned}
P\{N_1(t) = 1, N_2(t) = 0 | N_1(t) + N_2(t) = 1\} &= \frac{P\{N_1(t) = 1, N_2(t) = 0\}}{P\{N_1(t) + N_2(t) = 1\}} \\
&= \frac{P\{N_1(t) = 1\}P\{N_2(t) = 0\}}{P\{N_1(t) + N_2(t) = 1\}}\,.
\end{aligned}$$

In the above we have used the independence of the process $N_1(\cdot)$ and $N_2(\cdot)$. The above then equals

$$\frac{\frac{e^{-\lambda_1 t}(\lambda_1 t)^1}{1!} \cdot e^{-\lambda_2 t}}{\frac{e^{-(\lambda_1+\lambda_2)t}((\lambda_1+\lambda_2)t)^1}{1!}} = \frac{\lambda_1}{\lambda_1 + \lambda_2}\,.$$

Exercise 42 (arrival times for a Poisson process)

Part (a): For a Poisson process the inter-arrival times T_i are distributed as exponential random variables with parameter λ. Thus the expectation of T_i is well known since

$$E[T_i] = \frac{1}{\lambda}\,.$$

With this results we see that the expected value of the fourth arrival time $E[S_4]$ is given by

$$E[S_4] = E[\sum_{i=1}^{4} T_i] = \sum_{i=1}^{4} E[T_i] = \sum_{i=1}^{4} \frac{1}{\lambda} = \frac{4}{\lambda}\,.$$

From this we see that in general we have that

$$E[S_n] = \frac{n}{\lambda}\,.$$

Part (b): To calculate $E[S_4|N(1) = 2]$ we will use the memoryless property of the exponential distribution. Now since we are told that at $t = 1$ we have seen two of the four events (not more) from the time $t = 1$ onward the evaluation of the above expression we only need to compute the time for two more events to elapse. This last argument relies on the memoryless property of the exponential distribution, i.e. no matter how much time has elapsed between the last event seen, the time needed to elapse to the next event is given by an exponential distribution that is unchanged. Mathematically we have

$$E[S_4|N(1) = 2] = 1 + E[S_2] = 1 + \frac{2}{\lambda}\,,$$

using the results from Part (a) of this problem.

Part (c): To calculate $E[N(4) - N(2)|N(1) = 3]$ we will specifically use the *stationary* increments property of the Poisson process. Subtracting one from each $N(\cdot)$ in the above expectation we find that

$$\begin{aligned}
E[N(4) - N(2)|N(1) = 3] &= E[N(3) - N(1)|N(1) = 3] \\
&= E[N(3) - 3] \\
&= E[N(3)] - 3 \\
&= 3\lambda - 3\,.
\end{aligned}$$

Using the result that $E[N(t)] = \lambda t$ which was shown in the book.

Exercise 50 (waiting for the train)

Part (a): We will solve this problem by conditioning on the time of arrival of the train t. Specifically if X is the random variable denoting the number of passengers who get on the next train the we have

$$E[X] = E[E[X|T]]\,,$$

where T is the random variable denoting the arrival time of the next train after the previous train has left the station (we are told that this random variable is uniform). We begin by computing $E[X|T]$. Now assuming that the number of people that arrive to wait for the next train is a Poisson process, the probability that there are n people waiting at the train stop at time $T = t$ is given by the standard expression for a Poisson process i.e.

$$P\{X = n|T = t\} = \frac{e^{-\lambda t}(\lambda t)^n}{n!}\,,$$

where from the problem statement we have $\lambda = 7$ (per hour). This expression has an expected value is given by

$$E[X|T = t] = \lambda t\,.$$

The expectation of X can now be computed by conditioning on the random variable T. We have

$$E[X] = E[E[X|T]] = E[\lambda T] = \lambda E[T] = \frac{\lambda}{2}\,.$$

Since we know T to be a uniform random variable over $(0, 1)$.

Part (b): To compute the variance of X we will use the conditional variance formula given by

$$\mathrm{Var}(X) = E[\mathrm{Var}(X|T)] + \mathrm{Var}(E[X|T])\,.$$

We will compute each term on the right hand side. We begin with the second term: $\mathrm{Var}(E[X|T])$. Since from the above the random variable $E[X|T]$ is given by λT, the variance of this expression is related to the variance of T and is given by

$$\mathrm{Var}(E[X|T]) = \mathrm{Var}(\lambda T) = \lambda^2 \mathrm{Var}(T) = \frac{\lambda^2}{12}\,.$$

Now for the *first* term on the right hand side of the conditional variance expansion (and the properties of the Poisson distribution), in exactly the same way as we computed the expectation $E[X|T = t]$, the conditional variance is given by

$$\mathrm{Var}(X|T = t) = \lambda t\,,$$

so that the expectation of this expression then gives

$$E[\mathrm{Var}(X|T)] = \lambda E[T] = \frac{\lambda}{2}\,.$$

When we combine these two sub-results we finally conclude that

$$\mathrm{Var}(X) = \frac{\lambda}{2} + \frac{\lambda^2}{12}\,.$$

Exercise 57 (Poisson events)

Part (a): Since our events are generated by a Poisson process we know that

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

so in one hour ($t = 1$) we have that

$$P\{N(t) = 0\} = \frac{e^{-\lambda}(\lambda)^0}{0!} = e^{-\lambda} = e^{-2} = 0.1353,$$

Part (b): A Poisson process has the time between events given by exponential random variables with parameter λ. Thus the fourth event will occur at the time $S_4 = \sum_{i=1}^{4} X_i$, where each X_i is an exponential random variable with parameter λ. Thus

$$E[S_4] = \sum_{i=1}^{4} E[X_i] = \sum_{i=1}^{4} \frac{1}{\lambda} = \frac{4}{\lambda} = 2.$$

Since λ is measured in units of reciprocal hours, this corresponds to 2 P.M.

Part (c): The probability that two or more events occur between 6 P.M. and 8 P.M. (a two hour span) is the complement of the probability that less than two events occur in this two hour span. This latter probability is given by

$$\begin{aligned} P\{N(2) = 0\} + P\{N(2) = 1\} &= \frac{e^{-2\lambda}(2\lambda)^0}{0!} + \frac{e^{-2\lambda}(2\lambda)^1}{1!} \\ &= e^{-2\lambda} + e^{-2\lambda}2\lambda = 5e^{-4}. \end{aligned}$$

Thus the probability we seek is given by $1 - 5e^{-4} = 0.9084$.

Exercise 58 (a Geiger counter)

To solve this problem we will use the following fact. If $\{N(t), t \geq\}$ is a Poisson process with rate λ then the process that counts each of the events from the process $N(t)$ with probability p is another Poisson process with rate $p\lambda$.

Now for this specific problem the arrival of the particles at a rate of three a minute can be modeled as a Poisson process with a rate $\lambda = 3$ per minute. The fact that only 2/3 of the particles that arrive are actually measured means that this combined event can be modeled using a Poisson process with a rate $\frac{2}{3}\lambda = 2$ particles per minute. Thus the process $X(t)$ is a Poisson process with rate 2 particles per minute. With this information the various parts of the problem can be solved.

Part (a): In this case we have

$$P\{X(t) = 0\} = \frac{e^{-2t}(2t)^0}{0!} = e^{-2t}.$$

Part (b): In this case we have that

$$E[X(t)] = 2t\,.$$

Exercise 60 (bank arrivals)

Each part of this problem can be solved with the information that given n events have been observed by time t, the location of any specific event is *uniformly* distributed over the time interval $(0, t)$.

Part (a): We desire that *both* events land in the first $20/60 = 1/3$ of the total one hour time. This will happen with probability

$$\left(\frac{1}{3}\right)^2 = \frac{1}{9}\,.$$

Part (b): In this case we desire that at least one of the events fall in the first third of the total time. This can happen in several ways. The first (or second) event falls in this interval while the other event does not or both events happen in the first third of our total interval. This would give for a probability (reasoning like above) of

$$\frac{1}{3}\left(\frac{2}{3}\right) + \frac{1}{3}\left(\frac{2}{3}\right) + \frac{1}{3}\left(\frac{1}{3}\right) = \frac{5}{9}\,.$$

We note that the situation we desire the probability for is also the complement of the situation where both events land in the last $2/3$ of time. Using this idea the above probability could be calculated as

$$1 - \left(\frac{2}{3}\right)^2 = \frac{5}{9}\,.$$

Problem 81 (the distribution of the event times in a nonhomogenous process)

Part (a): Following the same strategy that the book used to compute the distribution function for a homogeneous Poisson process, we will begin by assuming an ordered sequence of arrival times

$$0 < t_1 < t_2 < \cdots < t_n < t_{n+1} = t$$

and let h_i be small increments such that $t_i + h_i < t_{i+1}$, for $i = 1, 2, \cdots n$. Then the probability that a random sample of n arrival times S_i happen at the times t_i and conditioned on the fact that we have n arrivals by time t can be computed by considering

$$P\{t_i \le S_i \le t_i + h_i, \quad \text{for} \quad i = 1, 2, \cdots, n | N(t) = n\}\,.$$

Which if we define the event A to be the event that we have exactly *one* event in $[t_i, t_i + h_i]$ for $i = 1, 2, \cdots, n$ and *no* events in the other regions then (by definition) the above equals the following expression

$$\frac{P\{A\}}{P\{N(t)=n\}}\,.$$

The probability that we have one event in $[t_i, t_i + h_i]$ is given by the fourth property in the definition of a nonhomogenous Poisson and is given by

$$P\{N(t_i + h_i) - N(t_i) = 1\} = \lambda(t_i)h_i + o(h_i) \tag{3}$$

To calculate the probability that we have *no* events in a given interval, we will derive this from the four properties in the definition a nonhomogenous Poisson process. Specifically, since the total probability must sum to one we have the constraint on increment variables over the range of time $[t_l, t_r]$,

$$P\{N(t_r) - N(t_l) = 0\} + P\{N(t_r) - N(t_l) = 1\} + P\{N(t_r) - N(t_l) \geq 2\} = 1\,.$$

or using properties (ii) and (iii) in the definition of a nonhomogenous Poisson process the above becomes (solving for $P\{N(t_r) - N(t_l) = 0\}$) the following

$$P\{N(t_r) - N(t_l) = 0\} = 1 - \lambda(t_l)(t_r - t_l) + o(t_r - t_l)\,. \tag{4}$$

This result will be used in what follows. To evaluate $P\{A\}$ we recognized that in the intervals

$$(0, t_1), (t_1 + h_1, t_2), (t_2 + h_2, t_3), \cdots, (t_n + h_n, t_{n+1})\,,$$

no events occurs, while in the intervals

$$(t_1, t_1 + h_1), (t_2, t_2 + h_2), (t_3, t_3 + h_3), \cdots, (t_n, t_n + h_n)\,,$$

one event occurs. By the independent increments property of nonhomogenous process the event A can be computed as the product of the probabilities of each of the above intervals event. The contributed probability $P(A_1)$ in the evaluation of $P\{A\}$ from the intervals where the count increase by *one* is given by

$$P(A_1) = \prod_{i=1}^{n}\{\lambda(t_i)h_i + o(h_i)\} = \prod_{i=1}^{n}\lambda(t_i)h_i + o(h)\,,$$

where we have used Eq. 3 and the term $o(h)$ represents terms higher than first order in any of the h_i's. By analogy with this result the contributed probability in the evaluation of $P\{A\}$ from the intervals where the count does not increase $P(A_0)$ is given by

$$P(A_0) = (1 - \lambda(0)(t_1) + o(t_1))\prod_{i=1}^{n}\{1 - \lambda(t_i + h_i)(t_{i+1} - t_i - h_i) + o(t_{i+1} - t_i - h_i)\}\,.$$

This expression will take some manipulations to produce a desired expression. We begin our sequence of manipulations by following the derivation in the book and recognizing that we will eventually be taking the limits as $h_i \to 0$. Since this expression has a finite limit we can take the limit of the above expression as is and simplify some of the notation. Taking the limit $h_i \to 0$ and defining $t_0 = 0$ the above expression becomes

$$P(A_0) = \prod_{i=0}^{n}\{1 - \lambda(t_i)(t_{i+1} - t_i) + o(t_{i+1} - t_i)\}\,.$$

We can simplify this product further by observing that the individual linear expressions we multiply can be written as an exponential which will facilitate our evaluation of this product. Specifically, it can be shown (using Taylor series) that

$$e^{-\lambda(t_i)(t_{i+1}-t_i)} = 1 - \lambda(t_i)(t_{i+1}-t_i) + o(t_{i+1}-t_i)\,.$$

With this substitution the product above becomes a sum in the exponential and we have

$$P(A_0) = \prod_{i=0}^{n} e^{-\lambda(t_i)(t_{i+1}-t_i)} = \exp\left\{-\sum_{i=0}^{n}\lambda(t_i)(t_{i+1}-t_i)\right\}\,.$$

Recognizing the above summation as an approximation to the integral of $\lambda(\cdot)$, we see that the above is approximately equal to the following

$$P(A_0) = \exp\left\{-\sum_{i=0}^{n}\lambda(t_i)(t_{i+1}-t_i)\right\} \approx \exp\left\{-\int_0^t \lambda(\tau)d\tau\right\} = e^{-m(t)}\,.$$

With these expressions for $P(A_1)$ and $P(A_0)$, we can now evaluate our target expression

$$\begin{aligned}
\frac{P\{A\}}{P\{N(t)=n\}} &= \frac{P(A_1)P(A_0)}{P\{N(t)=n\}} \\
&= \left(\frac{n!}{e^{-m(t)}m(t)^n}\right) P(A_1)P(A_0) \\
&= \left(\frac{n!}{e^{-m(t)}m(t)^n}\right)\left(\left(\prod_{i=1}^{n}\lambda(t_i)h_i + o(h)\right)e^{-m(t)}\right) \\
&= n!\left(\prod_{i=1}^{n}\frac{\lambda(t_i)}{m(t)}h_i + o(h)\right)\,.
\end{aligned}$$

It is this final result we were after. After dividing by $\prod_{i=1}^{n} h_i$ and taking the limit where $h_i \to 0$, we can conclude that the probability of drawing a specific sample of n event times (i.e. obtaining a draw of the random variables S_i) for a *nonhomogenous* Poisson process with rate $\lambda(t)$ given that we have seen n events by time t is given by

$$f_{S_1,S_2,\cdots,S_n}(t_1,t_2,\cdots,t_n|N(t)=n) = n!\left(\prod_{i=1}^{n}\frac{\lambda(t_i)}{m(t)}\right) \quad 0 < t_1 < t_2 < \cdots < t_n < t \tag{5}$$

We recognized that this expression is the same distribution as would be obtained for the order statistics corresponding to n independent random variables uniformly distributed with probability density function $f(\cdot)$ and a cumulative distribution function $F(\cdot)$ given by

$$f(x) = \frac{\lambda(x)}{m(t)} \quad \text{and} \quad F'(x) = f(x)\,.$$

By the definition of the function $m(\cdot)$ we have that $\lambda(x) = m'(x)$, so that an equation for our cumulative distribution function F is given by

$$F'(x) = \frac{m'(x)}{m(t)}\,.$$

This can be integrated to give

$$F(x) = \frac{m(x)}{m(t)}\,,$$

which can only hold if $x \le t$, while if $x > t$, $F(x) = 1$. This is the desired result.

Problem 82 (selecting events from an nonhomogenous Poisson process)

Part (a): We claim that the process $\{N_c(t), t \geq 0\}$ is a nonhomogenous process with an intensity function given by $p(t)\lambda(t)$.

Part (b): The statement made in Part (a) can be proven by showing that $N_c(t)$ satisfies the four conditions in the *definition* of a nonhomogenous Poisson process with intensity function $p(t)\lambda(t)$. We begin, by defining $N(t)$ to be the nonhomogenous Poisson process with intensity function $\lambda(t)$. We then have

- Because $N(t)$ is a nonhomogenous Poisson process we have $N(0) = 0$. Because no events have occurred at time $t = 0$ we then must necessarily have $N_c(0) = 0$.
- As $N(t)$ has the property of independent increments $N_c(t)$ will inherit this property also. This can be reasoned by recognizing that if the increment variables for our original nonhomogenous Poisson process $N(t)$ are independent then when we select a *subset* of these events by considering the process $N_c(t)$ and involving the probability function $p(\cdot)$ we cannot introduce dependencies.
- We will compute the expression $P\{N_c(t+h) - N_c(t) = 1\}$. Note that the desired event $N_c(t+h) - N_c(t) = 1$ only holds true if our original process has an event $N(t+h) - N(t) = 1$. Thus since we desire this probability we begin by defining the event E_t to be the event that (as h goes to zero) we count the *single* event that happens in the interval $(t, t+h)$ from our original nonhomogenous Poisson process $N(t)$. We then have conditioning on E_t that

$$
\begin{aligned}
P\{N_c(t+h) - N_c(t) = 1\} &= P\{N(t+h) - N(t) = 1|E_t\}P(E_t) \\
&= P\{N(t+h) - N(t) = 1\}p(t) \\
&= (\lambda(t)h + o(h))p(t) \\
&= \lambda(t)p(t)h + o(h)\,.
\end{aligned}
$$

  Where we have used independence of E_t and $N(t)$.
- To compute $P\{N_c(t+h) - N_c(t) > 1\}$ define E_G to be the event that in the interval $(t, t+h)$ we count at least one event. This is the complement of the event we count no events in this interval. Then we have

$$
\begin{aligned}
P\{N_c(t+h) - N_c(t) > 1\} &= P\{N(t+h) - N(t) > 1|E_G\}P(E_G) \\
&= P\{N(t+h) - N(t) > 1\}P(E_G) \\
&= o(h)P(E_G) = o(h)\,.
\end{aligned}
$$

Thus we have shown that $N_c(t)$ satisfies the four requirements in the definition of a nonhomogenous Poisson process with intensity function $p(t)\lambda(t)$.

Problem 83 (time rescaling results in a nonhomogenous Poisson processes)

We will show that $N(t)$ is an nonhomogenous Poisson process by showing that all four of the required properties hold. We have

- $N(0) = N_0(m(0)) = N_0\left(\int_0^0 \lambda(s)ds\right) = N_0(0) = 0$, since $N_0(t)$ is a Poisson process.
- To show that $N(t)$ has independent increments consider two non overlapping intervals of time (t_1, t_2) and (t_3, t_4). Then the since $\lambda(\cdot)$ is a non-negative function and the t's are ordered as
$$t_1 < t_2 \leq t_3 < t_4$$
we have that $m(t) = \int_0^t \lambda(s)ds$ is an increasing function of t. Thus our ordering of time above introduced an ordering of $m_i \equiv m(t_i)$ and we see
$$m(t_1) < m(t_2) \leq m(t_3) < m(t_4)\,,$$
so the independent increments property of $N_0(t)$ on the "transformed" intervals $(m(t_1), m(t_2))$ and $(m(t_3), m(t_4))$ imply that same property for $N(t)$.
- Consider
$$\begin{aligned} P\{N(t+h) - N(t) = 1\} &= P\{N_0(m(t+h)) - N_0(m(t)) = 1\} \\ &= P\{N_0(m(t) + m'(t)h + o(h)) - N_0(m(t)) = 1\} \\ &= 1 \cdot (m'(t)h + o(h)) \\ &= m'(t)h + o(h)\,. \end{aligned}$$
since $m(t) = \int_0^t \lambda(s)ds$ we have that $m'(t) = \lambda(t)$ and the above becomes
$$P\{N(t+h) - N(t) = 1\} = \lambda(t)h + o(h)\,.$$
- As in the above we find
$$\begin{aligned} P\{N(t+h) - N(t) > 1\} &= P\{N_0(m(t) + m'(t)h + o(h)) - N_0(m(t)) > 1\} \\ &= o(m'(t)h + o(h)) = o(h)\,. \end{aligned}$$

Thus we have shown that $N(t)$ satisfies all four requirements for a nonhomogenous Poisson process with rate function $\lambda(t)$ as we were asked to show.

Problem 72 (the departure of cable car riders)

Part (a): By definition the last rider will depart after all $n-1$ other riders have. Lets assume that we start our cable car with all n passengers at the time $t = 0$ and define the random variables X_i to be the amount of time between successive stops. This means that we will stop to let off the first passenger at a time of X_1, we will stop for the second passenger to get off at a time $X_1 + X_2$, for the third passenger to get off at a time of $X_1 + X_2 + X_3$, etc. In all of these cases X_i is an exponentially distributed random variable with rate λ. Thus the last or nth passenger gets off the cable car at a time T given by $T = \sum_{i=1}^{n} X_i$. A sum of n exponentially distributed random variables each with a rate of λ is distributed as Gamma random variable with parameters (n, λ). That is the distribution function for T is given by

$$f_T(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} .$$

Part (b): To solve this problem introduce the random variables Y_i to be the time it takes passenger i to arrive home after he/she is dropped off. Then for the *next to last* passenger (the $n-1$st rider) to be home before the last passenger (the nth rider) gets off requires $Y_{n-1} < X_n$. Since Y_{n-1} is an exponential random variable with rate μ the probability this is true is given by $\frac{\mu}{\mu+\lambda}$. Given that the $n-1$ passenger makes it home before the nth passenger gets off lets now consider the $n-2$nd passenger. Using the same logic for the $n-2$ passenger to get home before the nth one gets off requires that

$$Y_{n-2} < X_n + X_{n-1} .$$

In general, the requirement that the $n-k$th passenger get home before the nth rider has departed is

$$Y_{n-k} < \sum_{i=0}^{k-1} X_{n-i} \quad \text{for} \quad k = 1, 2 \cdots n-1 .$$

Since each Y_i is an exponential random variable and each sum is a Gamma random variable, to evaluate the probability of each of these events we need the following lemma.

Lemma: If Y is an exponential random variable with parameter μ and X is a Gamma random variable with parameters (λ, n), independent of Y, we have

$$\begin{aligned} P\{Y < X\} &= \int_0^\infty P\{Y < X | X = x\} P\{X = x\} dx \\ &= \int_0^\infty P\{Y < x\} \lambda e^{-\lambda x} \frac{(\lambda x)^{n-1}}{(n-1)!} dx \\ &= \int_0^\infty (1 - e^{-\mu x}) \lambda e^{-\lambda x} \frac{(\lambda x)^{n-1}}{(n-1)!} dx \\ &= 1 - \frac{\lambda^n}{(\lambda + \mu)^n} , \end{aligned}$$

when we perform the required integration.

Combining this result with the discussion above we see that the probability that all the other riders are home (denoted here as P_h) at the time the nth gets off is given by

$$\begin{aligned}
P_h &= \prod_{k=1}^{n-1} P\left\{Y_{n-k} < \sum_{i=0}^{k-1} X_{n-i}\right\} \\
&= P\{Y_{n-1} < X_n\}P\{Y_{n-2} < X_n + X_{n-1}\}\cdots P\{Y_1 < X_n + X_{n-1} + \cdots X_2\} \\
&= \left(\frac{\mu}{\mu+\lambda}\right)\left(1 - \frac{\lambda^2}{(\mu+\lambda)^2}\right)\left(1 - \frac{\lambda^3}{(\mu+\lambda)^3}\right)\cdots\left(1 - \frac{\lambda^{n-1}}{(\mu+\lambda)^{n-1}}\right) \\
&= \prod_{k=1}^{n-1}\left(1 - \frac{\lambda^k}{(\lambda+\mu)^k}\right).
\end{aligned}$$

Problem 73 (shocks to the system)

Part (a): We are told that it takes n shocks to cause our system to fail. Because these events are from a Poisson process the times between individual events X_i are exponentially distributed with a rate λ. Thus the time T can be written $T = \sum_{i=1}^n X_i$, and we have

$$P\{T = t|N = n\} = P\{\sum_{i=1}^n X_i = t|N = n\}.$$

Since the distribution of the sum of n exponential random variables is distributed as a Gamma random variable we have that the above becomes

$$P\{T = t|N = n\} = \lambda e^{-\lambda t}\frac{(\lambda t)^{n-1}}{(n-1)!}.$$

Part (b): We want to compute $P\{N = n|T = t\}$ which we can compute using an application of Bayes' rule. We have

$$P\{N = n|T = t\} = \frac{P\{T = t|N = n\}P\{N = n\}}{P\{T = t\}}.$$

The first factor in the numerator was calculated in Part (a). The second factor in the numerator $P\{N = n\}$ represents the probability it takes n trails to get one "success" where a success represents the failure of the system. This is given by a negative-binomial distribution and in this case is given by $p(1-p)^{n-1}$. The denominator $P\{T = t\}$ can be calculated by conditioning on N as

$$\begin{aligned}
P\{T = t\} &= \sum_{n=1}^{\infty} P\{T = t|N = n\}P\{N = n\} \\
&= \lambda p e^{-\lambda t}\sum_{n=1}^{\infty}\frac{(\lambda t(1-p))^{n-1}}{(n-1)!} \\
&= \lambda p e^{-\lambda t}e^{\lambda t(1-p)} = \lambda p e^{-\lambda p t}.
\end{aligned}$$

Note that this is an exponential distribution with rate λp, which could have been determined without any calculations as follows. Recall that a Poisson process with rate λ that has its events "filtered" with a Bernoulli process with parameter p is another Poisson process with rate λp. Since this second Poisson process represents the process of failures the first failure *time* will be distributed as an exponential random variable with rate λp exactly as found above. Using these parts we then have

$$\begin{aligned}
P\{N = n | T = t\} &= \frac{\left(\frac{\lambda e^{-\lambda t}(\lambda t)^{n-1}}{(n-1)!}\right) p(1-p)^{n-1}}{\lambda p e^{-\lambda t p}} \\
&= e^{-\lambda(1-p)t} \frac{(\lambda(1-p)t)^{n-1}}{(n-1)!} \quad \text{for} \quad n = 1, 2, \cdots .
\end{aligned}$$

This looks exactly like a Poisson distribution with a mean $\lambda(1-p)t$ but starts at the value $n = 1$. Defining the random variable N' as $N' \equiv N - 1$, then from the above expression the distribution of N' is given by

$$P\{N' = n' | T = t\} = e^{-\lambda(1-p)t} \frac{(\lambda(1-p)t)^{n'}}{(n')!} \quad \text{for} \quad n' = 0, 1, 2, \cdots .$$

So N' is a Poisson random variable with a mean $\lambda(1-p)t$ and we have that N is given as $1 + N'$ as we were to show.

Part (c): The desired probability $P\{N = n | T = t\}$ for $n = 1, 2, \cdots$ could have been determined without calculation using the following argument. The probability $P\{N = n | T = t\}$ means that the system failure happened on shock n at the time t. This failure must have been caused by the last shock and requires that the previous $n - 1$ shocks did not result in failure. Since non-failure causing shocks occur according to a Poisson process with rate $\lambda(1-p)$, the *number* of such non-failure causing shocks, $n - 1$, is given by a Poisson random variable with mean $\lambda(1-p)t$, exactly as calculated in Part (b).

Problem 93 (min/max identities)

Part (a): To prove the given identity

$$\max(X_1, X_2) = X_1 + X_2 - \min(X_1, X_2) ,$$

we can appeal to the *law of the excluded middle*. Which is a fancy way of saying we can simply consider the two possible cases for the relationship between X_1 and X_2. For example, if $X_1 \le X_2$, then the left hand side of the above expression is $\max(X_1, X_2) = X_2$ while the right hand side is $X_1 + X_2 - \min(X_1, X_2) = X_1 + X_2 - X_1 = X_2$ which are equal. If in fact the other relationship holds between X_1 and X_2, that is, $X_1 \ge X_2$, a similar calculation shows that the above identity to be true in this case also.

Part (b): I had a difficult time proving this identity directly but can offer the following arguments in support of the given expression when $X_i \ge 0$. Consider the expression $\max(X_1, X_2, \cdots, X_n)$ in comparison to the expression $\sum_{i=1}^n X_i$. Since one of the X_i, say X_{i^*}, is the largest the summation expression will "over count" the maximum by $\sum_{i \ne i^*} X_i$.

That is by all X_i that are less than X_{i^*} and we have

$$\max(X_1, X_2, \cdots, X_n) \leq \sum_{i=1}^{n} X_i \,.$$

We can "adjust" our sum on the right hand side to make it more like the left hand side by subtracting off all terms smaller than X_{i*}. Terms that are smaller than X_{i^*} can be represented as $\min(X_i, X_j)$ with $i < j$. We need to include the requirement that $i < j$ since otherwise considering all possible terms like $\min(X_i, X_j)$ we would double count all possible minimums i.e. $\min(X_i, X_j) = \min(X_j, X_i)$. Thus we now might consider the following

$$\sum_{i=1}^{n} X_i - \sum_{i<j} \min(X_i, X_j) \,.$$

Note that in the above the second expression now subtracts too many terms since smaller terms will certainly appear more than once. For example if $X_1 = 1$, $X_2 = 2$, and $X_3 = 3$ then $\max(X_1, X_2, X_3) = 3$, while the above gives

$$1 + 2 + 3 - (1 + 1 + 2) = 3 - 1 \neq 3 \,.$$

Thus the value for X_1 was subtracted off twice. This logic allows one to conclude that

$$\max(X_1, X_2, \cdots, X_n) \geq \sum_{i=1}^{n} X_i - \sum_{i<j} \min(X_i, X_j) \,.$$

Our approximation to the maximum expression can be improved on by adding terms like $\min(X_i, X_j, X_k)$ to give

$$\max(X_1, X_2, \cdots, X_n) \leq \sum_{i=1}^{n} X_i - \sum_{i<j} \min(X_i, X_j) + \sum_{i<j<k} \min(X_i, X_j, X_k) \,.$$

This process is repeated until we add/subtract $\min(X_1, X_2, \cdots, X_n)$. Whether we add or subtract depends on whether the number of variables we start with n is even or odd.

If we define X_i for $i = 1, 2, \cdots n$ to be an indicator random variables denoting whether or not the event A_i occurred we see that

$$\begin{aligned}
\max(X_i, X_j, \cdots, X_k) &= P(A_i \cup A_j \cdots \cup A_k) \quad \text{and} \\
\min(X_i, X_j, \cdots, X_k) &= P(A_i A_j \cdots A_k) \,.
\end{aligned}$$

With these expressions and using the max/min identity proven in Part (a) above we obtain the well known identity that

$$P\left(\cup_{i=1}^{n} A_i\right) = \sum_{i} P(A_i) - \sum_{i<j} P(A_i A_j) + \sum_{i<j<k} P(A_i A_j A_k) + \cdots + (-1)^{n-1} P(A_1 A_2 \cdots A_n) \,.$$

Part (c): There seemed to be two ways to do this part of the problem. To explicitly use the results from above, let X_i be the random variable denoting the time when the *first* event has occurred in the ith Poisson process. In a similar way, let X be a random variable denoting the time at which and event has occurred in *all* n Poisson processes. Then $X = \max(X_1, X_2, \cdots, X_n)$. We desire to calculate $E[X]$. From the identity in Part (b) we have that X can be expressed in term of minimization's as

$$X = \sum_i X_i - \sum_{i<j} \min(X_i, X_j) + \sum_{i<j<k} \min(X_i, X_j, X_k) + \cdots + (-1)^{n-1} \min(X_1, X_2, \cdots, X_n)\,.$$

Taking the expectation of the above expression we find that

$$\begin{aligned} E[X] &= \sum_i E[X_i] - \sum_{i<j} E[\min(X_i, X_j)] \\ &+ \sum_{i<j<k} E[\min(X_i, X_j, X_k)] + \cdots + (-1)^{n-1} E[\min(X_1, X_2, \cdots, X_n)]\,. \end{aligned}$$

Using the independence of the X_i, the fact that they are exponential with a rate λ_i, and earlier results from the book about the distribution of minimum of exponential random variables we find

$$E[X] = \sum_{i=1}^{n} \frac{1}{\lambda_i} - \sum_{i<j} \frac{1}{\lambda_i + \lambda_j} + \sum_{i<j<k} \frac{1}{\lambda_i + \lambda_j + \lambda_k} + \cdots + \frac{(-1)^{n-1}}{\lambda_1 + \lambda_2 + \cdots + \lambda_n}\,. \tag{6}$$

As another way to solve this problem that results in a quite different formula consider the following where we compute the distribution function for X directly. Using the above definitions we find

$$\begin{aligned} P\{X < t\} &= P\{\max_i(X_i) < t\} = P\{X_i < t \text{ for all } i\} \\ &= \prod_{i=1}^{n} (1 - e^{-\lambda_i t})\,. \end{aligned}$$

Now we can compute $E[X]$ if we know $P\{X > t\}$ as $E[X] = \int_0^\infty P\{X > t\} dt$. From this we have

$$E[X] = \int_0^\infty \left(1 - \prod_{i=1}^{n} (1 - e^{-\lambda_i t}) \right) dt\,. \tag{7}$$

From the above functional form it appears that one should be able to expand the product in the integrand, perform the integration, and show that it agrees with the Equation 6.

Chapter 5: Additional Exercises

The following are some problems that appeared in different editions of this book in the chapter on exponential distributions and Poisson processes.

Expectations of products of pairs of order statistics

The Problem: Let $X_1, X_2, \cdots, X_n$ be independent exponential random variables, each having rate λ. Also let $X_{(i)}$ be the ith smallest of these values $i = 1, \cdots, n$. Find

(a) $E[X_{(1)}X_{(2)}]$

(b) $E[X_{(i)}X_{(i+1)}]$ for $i = 1, 2, \cdots, n-1$.

(c) $E[X_{(i)}X_{(j)}]$ for $i < j$.

The Solution: To evaluate the above expectations we will simply use the joint density of the order statistic $X_{(i)}$ and $X_{(j)}$, when $i < j$ and integrate over the appropriate region. Rather than deriving this density we simply state the result, which can be discussed in [1]. We have

$$\begin{aligned} f_{X_{(i)},X_{(j)}}(x_i, x_j) &= \frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F(x_i)]^{i-1} \\ &\times \; [F(x_j) - F(x_i)]^{j-i-1}[1 - F(x_j)]^{n-j} f(x_i) f(x_j) \quad \text{for} \quad x_i < x_j \,. \end{aligned}$$

When X is an exponential random variable with rate λ we have that our density and distribution functions f and F are given by the following special forms

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

So that our density above becomes in this specific case the following

$$\begin{aligned} f_{X_{(i)},X_{(j)}}(x_i, x_j) &= \frac{\lambda^2 n!}{(i-1)!(j-i-1)!(n-j)!}[1 - e^{-\lambda x_i}]^{i-1} \\ &\times \; [e^{-\lambda x_i} - e^{-\lambda x_j}]^{j-i-1} e^{-\lambda[(n-j+1)x_j + x_i]} \quad \text{for} \quad x_i < x_j \,. \end{aligned}$$

Part (a): When $i = 1$ and $j = 2$ our density function above becomes

$$f_{X_{(1)},X_{(2)}}(x_1, x_2) = \lambda^2 n(n-1) e^{-\lambda((n-1)x_2 + x_1)} \quad \text{for} \quad x_1 < x_2 \,.$$

So that our expectation is then given by

$$\begin{aligned} E[X_{(1)}X_{(2)}] &= \int_{x_1=0}^{\infty} \int_{x_2=x_1}^{\infty} x_1 x_2 f_{X_{(1)},X_{(2)}}(x_1, x_2) dx_2 dx_1 \\ &= \lambda^2 n(n-1) \int_{x_1=0}^{\infty} \int_{x_2=x_1}^{\infty} x_1 x_2 e^{-\lambda((n-1)x_2 + x_1)} dx_2 dx_1 \\ &= \lambda^2 n(n-1) \int_{x_1=0}^{\infty} x_1 e^{-\lambda x_1} \int_{x_2=x_1}^{\infty} x_2 e^{-\lambda(n-1)x_2} dx_2 dx_1 \\ &= \frac{3n-2}{(n-1)} \left(\frac{1}{n\lambda} \right)^2 , \end{aligned}$$

See the Mathematica file `chap_5_eprob_pp.nb` for the algebra for this problem. In addition, for a Monte-Carlo verification of the above see the Matlab file `exp_pp_order_stats_pt_a.m` and Figure 1.

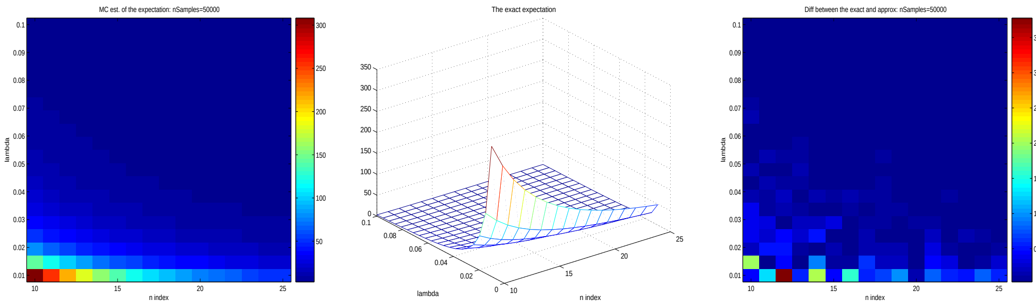

Figure 1: A comparison between a Monte-Carlo simulation of $E[x_{(1)}x_{(2)}]$ and the derived analytical expression. **Left:** the Monte-Carlo simulation plotted as a function of λ and n. **Middle:** The analytical solution derived above. **Right:** The error between the two.

Part (b): When $j = i + 1$ our joint density function becomes

$$f_{X_{(i)},X_{(i+1)}}(x_i, x_{i+1}) = \frac{\lambda^2 n!}{(i-1)!(n-i-1)!}(1-e^{-\lambda x_i})^{i-1}e^{-\lambda[(n-i)x_{i+1}+x_i]} \quad \text{for} \quad x_i < x_{i+1}\,.$$

So that our expectation is then given by

$$\begin{aligned}
E[X_{(i)}X_{(i+1)}] &= \int_{x_i=0}^{\infty}\int_{x_{i+1}=x_i}^{\infty} x_i x_{i+1} f_{X_{(i)},X_{(i+1)}}(x_i, x_{i+1}) dx_{i+1} dx_i \\
&= \frac{\lambda^2 n!}{(i-1)!(n-i-1)!} \\
&\times \int_{x_i=0}^{\infty}\int_{x_{i+1}=x_i}^{\infty} x_i x_{i+1}(1-e^{-\lambda x_i})^{i-1} e^{-\lambda((n-i)x_{i+1}+x_i)} dx_{i+1} dx_i \\
&= \frac{\lambda^2 n!}{(i-1)!(n-i-1)!} \\
&\times \int_{x_i=0}^{\infty} x_i (1-e^{-\lambda x_i})^{i-1} e^{-\lambda x_i} \int_{x_{i+1}=x_i}^{\infty} x_{i+1} e^{-\lambda(n-i)x_{i+1}} dx_{i+1} dx_i\,.
\end{aligned}$$

At this point even performing algebra by a computer becomes difficult. See the Mathematica file `chap_5_eprob_pp.nb` for some of the algebra. A literature search reveals that this is a rather difficult problem to solve exactly, so instead I'll use a Monte-Carlo calculation to compute these expressions. See the Matlab file `exp_pp_order_stats_pt_b.m` and Figure 2 for the computational experiment done for this part of the problem. Since we expect the expression to scale as $1/\lambda^2$, we have assumed $\lambda = 1$ for the plots presented.

Part (c): As in **Part (b)** an exact analytic solution is quite difficult so in the Matlab file `exp_pp_order_stats_pt_c.m` and Figure 3, I have presented the results from the Monte-Carlo calculation done to evaluate the requested expression. Again since we expect the results to scale as $1/\lambda^2$, we have assumed that $\lambda = 1$.

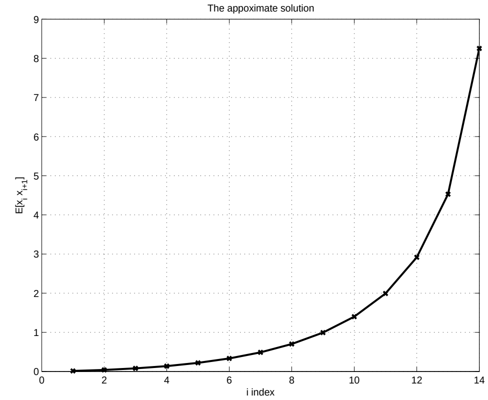

Figure 2: The Monte-Carlo simulation of $E[x_{(i)}x_{(i+1)}]$ as a function of i for $i = 1, 2, \cdots, n-1$.

Figure 3: The Monte-Carlo simulation of $E[x_{(i)}x_{(j)}]$ as a function of i and j for $i < j$.

A Identity with Reciprocal Sums

The Problem: Argue that if λ_i, $i = 1, 2, \cdots, n$ are distinct positive numbers then

$$\sum_{i=1}^{n} \frac{1}{\lambda_i} \prod_{j \neq i} \frac{\lambda_j}{\lambda_j - \lambda_i} = \sum_{i=1}^{n} \frac{1}{\lambda_i} .$$

Hint: Relate this problem to Section 5.2.4.

The Solution: From Section 5.2.4 we know that the random variable X defined as the sum of n exponential random variables $X = \sum_{i=1}^{n} X_i$ is called a hypo-exponential random variable and has a probability density function given by

$$f_{X_1+X_2+\cdots+X_n}(x) = \sum_{i=1}^{n} \prod_{j \neq i} \frac{\lambda_j}{\lambda_j - \lambda_i} \lambda_i e^{-\lambda_i t} .$$

Since X is the sum of n exponential random variables when we multiply both sides of the above by x and integrate the left hand side becomes the expectation of random variable $X = \sum_{i=1}^{n} X_i$, and equals

$$\sum_{i=1}^{n} \frac{1}{\lambda_i} .$$

While the right hand side results in a another expectation of sums of exponential random variables. Taking these expectations we have an expression given by

$$\sum_{i=1}^{n} \frac{1}{\lambda_i} \prod_{j \neq i} \frac{\lambda_j}{\lambda_j - \lambda_i} .$$

When we equate these two results we have the requested theorem.

We can derive the fact that $P\{N(t)=1\}=\lambda t+o(t)$

The Problem: Show that assumption (iv) of Definition 5.3 follows from assumptions (ii) and (iii).

Hint: Derive a functional equation for $g(t)=P\{N(t)=0\}$.

The Solution: Following the hint for this problem we will try to derive a functional relationship for $P\{N(t)=0\}$, by considering $P\{N(t+s)=0\}$. Now if $N(t+s)=0$, this event is equivalent to the event that $N(t)=0$ *and* $N(t+s)-N(t)=0$. so we have that

$$\begin{aligned}
P\{N(t+s)=0\} &= P\{N(t)=0, N(t+s)-N(t)=0\} \\
&= P\{N(t)-N(0)=0, N(t+s)-N(t)=0\} \\
&= P\{N(t)-N(0)=0\}P\{N(t+s)-N(t)=0\} \\
&= P\{N(t)=0\}P\{N(s)=0\}\,.
\end{aligned}$$

When we used the property of stationary independent increments. Thus defining $f(t)\equiv P\{N(t)=0\}$, from the above we see that f satisfies

$$f(t+s)=f(t)f(s)\,.$$

By the discussion in the book the unique continuous solution to this equation is $f(t)=e^{-\lambda t}$, for some λ. Thus we have that $P\{N(t)=0\}=e^{-\lambda t}$. Using (iii) from Definition 5.3 and the fact that probabilities must be normalized (sum to one) we have that

$$P\{N(t)=0\}+P\{N(t)=1\}+P\{N(t)\geq 2\}=1\,.$$

which gives us (solving for $P\{N(t)=1\}$) the following

$$\begin{aligned}
P\{N(t)=1\} &= 1-P\{N(t)=0\}-P\{N(t)\geq 2\} \\
&= 1-e^{-\lambda t}-o(t) \\
&= 1-(1-\lambda t+o(t))-o(t) \\
&= \lambda t+o(t)\,,
\end{aligned}$$

as we were to show.

Cars v.s. Vans on the Highway

The Problem: Cars pass a point on the highway at a Poisson rate of one per minute. If 5 percent of the cars on the road are vans, then

- what is the probability that at least one van passes during the hour?
- given that ten vans have passed by in an hour, what is the expected number of cars to have passes by in that time?

- if 50 cars have passed by in an hour, what is the probability that five of them are vans?

The Solution:

Part (a): The number of vans that have passed by in time $N_v(t)$ is given by a Poisson process with rate $p\lambda$, where p is the fraction of the cars that are vans and λ is the Poisson rate of the cars (we are told these values are $\lambda = 1$ (per minute) and $p = 0.05$). Because of this fact the probability that at least one van passes by in one hour $t = 60$ (minutes) is given by

$$\begin{aligned} P\{N_v(60) \geq 1\} &= 1 - P\{N_v(60) = 0\} \\ &= 1 - e^{-p\lambda 60} \\ &= 1 - e^{-0.05 \cdot 1 \cdot 60} \\ &= 0.9502\,. \end{aligned}$$

Part (b): We can think of the question posed: how many cars need to pass by to given ten vans as equivalent to the statement of the expected number of trials will we need to perform to guarantee ten successes, where the probability of success is given by $p = 0.05$. This is the *definition* of a negative binomial random variable. If we define the random variable X to be the number of trials needed to get r success where each success occurs with probability p we have that

$$E[X] = \frac{r}{p}\,.$$

For this part of the problem at hand $(r, p) = (10, 0.05)$, so the expected number of cars that must have passed is given by

$$\frac{10}{0.05} = 200\,.$$

Part (c): If 50 cars have passed by and we have a 5 percent chance that each car is a van, the probability that n from 50 are vans is a binomial random variable with parameters $(n, p) = (50, 0.05)$, so the answer to the question posed is given by

$$\binom{50}{5}(0.05)^5(0.95)^{45} = 0.06584\,.$$

References

[1] S. Ross. *A First Course in Probability*. Macmillan, 3rd edition, 1988.