Probability distributions EBP038A05
Study Guide

Nicky D. van Foreest, Ruben van Beesten
Joost Doornbos, Wietze Koops, Mikael Makonnen, Zexuan Yuan

August 15, 2022

CONTENTS

INTRODUCTION

This study guide contains material organized per chapter of BH.

1. The simple questions and exercises are based on each section of BH and are meant to practice while you read. These questions are (often much) simpler than exam questions, but just help you to read and study well. It takes time and attention to understand definitions and notation. Mind that good notation and good understanding strongly correlate.

2. The part related to the obligatory exercises of BH provides motivational comments, hints and solutions.

3. The third part contains challenges. These problems are (quite a bit) above exam level, hence optional. However, if you like to be intellectually challenged, then you'll like these problems a lot.

In general, when working on the exercises, try first hard to find the answer and *write it down* on paper. Only after having written your answer on paper, meticulously compare your work with ours. Like this you' ll get a lot of feedback, and you'll see that it is quite hard to get the details right.

Finally, we included many old exam questions. Of course you are not expected to make them all. Instead do a just few until you feel comfortable with the level.

The selection of exercises in the table above are the bare minimum; I advice you to do more. To assure you, I found the problems quite hard at times; probability never 'comes for free'; not for you, not for me, not for anybody. You can expect to spend between 30 minutes (and sometimes more) per problem; if you are serious.

Here is a list of good, and important, advice when making the exercises. (As a student I did not always do this, partly because I was not aware about how useful this advice is. Hopefully you are smart enough to avoid making the same mistakes as I did as a student.)

- Read an example in the book. Close the book, and try to redo the example. When I try, I often fail. Why is that? Simple: I did not really think about the example while just reading it, I skimmed it. But you should get used to the fact that reading requires pen and paper.

- Before trying to solve an exercise, read all parts of it, i.e., part a, b, etc. Ensure you understand the problem.

- Before actually solving an exercise, *make a plan on how to solve it.* A first step is to look for simple corner cases (set things to zero, make certain probabilities equal to one, and so on), make extra assumptions that simplify the problem, and solve the problem under these simplifying (stronger) assumptions. Then drop an assumption, and try to generalize to a pattern or some property you expect to hold. You'll be astonished to see how many problems you can actually solve by following this strategy. And even if you cannot solve it with this approach, the corner cases help to check throughout whether you're still working in the right direction. Also, reduce the problem to simpler cases you do understand. Try to solve the simpler problem first, and then generalize.

- Carry out your plan, and *relax* if you cannot directly find the answer.
- Look back after solving the problem, and try to find a general pattern you used to solve the problem. Can you use this for other problems too?
- Look back again at the problem some time later. In other words, do not solve a problem just once, but also a few weeks later again. This is often very revealing.
- Work every day a reasonable amount of time. This is much more effective than working 10 h on one day, and not at all the next. The concept is often called 'Kaizen': try to improve every day a little bit. Over the course of time, you'll be amazed how much you can achieve.
- Finally, when I am stuck, this piece of advice of Jim Rohn (an author on personal development) helps: 'Don't wish it was easier, wish you were better.'

7 CHAPTER 7: EXERCISES AND REMARKS

7.1 7.1

Ex 7.1. In your own words, explain what is

1. a joint PMF, PDF, CDF;
2. a conditional PMF, PDF, CDF;
3. a marginal PMF, PDF, CDF.

s.7.1. Check the definitions of the book.

Mistake: To say that $\mathsf{P}\{X = x\}$ is the PMF for a continuous random variable is wrong, because $\mathsf{P}\{X = x\} = 0$ when X is continuous.

Why is $\mathsf{P}\{1 < x \leq 4\}$ wrong notation? hint: X should be a capital. What is the difference between X and x?

Ex 7.2. Suppose the probability of obtaining a head twice out of two coin flips is $\mathsf{P}\{X_1 = H, X_2 = H\}$. What has this to do with joint PMFs? Can you generalize this idea to other examples?

s.7.2. This example shows why joint distributions are important! In any experiment that involves a sequence of measurements, such as multiple throws of a coin, or the weighing of a bunch of chimpanzees, we have to deal with joint CDFs and PMFs.

Ex 7.3. In the previous exercise, suppose the outcome of the second throw is always equal to that of the first. Specify the joint PMF.

s.7.3. Here, we deal with two rvs, and we have to specify how they depend. In the present case $\mathsf{P}\{X_1 = H, X_2 = H\} = \mathsf{P}\{X_1 = H\}$ and $\mathsf{P}\{X_1 = T, X_2 = T\} = \mathsf{P}\{X_1 = T\}$, $\mathsf{P}\{X_1 = H, X_2 = T\} = \mathsf{P}\{X_1 = T, X_2 = H\} = 0$. Note that with this, we specified the joint PMF on all possible outcomes.

Ex 7.4. We have the random vector $(X, Y) \in [0,1]^2$ (here $[0,1]^2 = [0,1] \times [0,1]$) consisting of the rvs X and Y with the joint PDF $f_{X,Y}(x, y) = 2\, I_{x \leq y}$.

1. Are X and Y independent?
2. Compute $F_{X,Y}(x, y)$.

s.7.4.

$$f_X(x) = \int_0^1 f_{X,Y}(x,y)\,\mathrm{d}y = 2\int_0^1 I_{x\leq y}\,\mathrm{d}y = 2\int_x^1 \mathrm{d}y = 2(1-x) \tag{7.1.1}$$

$$f_Y(y) = \int_0^1 f_{X,Y}(x,y)\,\mathrm{d}x = 2\int_0^1 I_{x\leq y}\,\mathrm{d}x = 2\int_0^y \mathrm{d}y = 2y. \tag{7.1.2}$$

But $f_{X,Y}(x,y) \neq f_X(x) f_Y(y)$, hence X, Y are dependent.

$$F_{X,Y}(x,y) = \int_0^x \int_0^y f_{X,Y}(u,v)\,\mathrm{d}v\,\mathrm{d}u \tag{7.1.3}$$
$$= 2\int_0^x \int_0^y I_{u\leq v}\,\mathrm{d}v\,\mathrm{d}u \tag{7.1.4}$$
$$= 2\int_0^x \int I_{u\leq v} I_{0\leq v\leq y}\,\mathrm{d}v\,\mathrm{d}u \tag{7.1.5}$$
$$= 2\int_0^x \int I_{u\leq v\leq y}\,\mathrm{d}v\,\mathrm{d}u \tag{7.1.6}$$
$$= 2\int_0^x [y-u]^+\,\mathrm{d}u, \tag{7.1.7}$$

because $u > y \implies I_{u\leq v\leq y} = 0$. Now, if $y > x$,

$$2\int_0^x [y-u]^+\,\mathrm{d}u = 2\int_0^x (y-u)\,\mathrm{d}u = 2yx - x^2, \tag{7.1.8}$$

while if $y \leq x$,

$$2\int_0^x [y-u]^+\,\mathrm{d}u = 2\int_0^y (y-u)\,\mathrm{d}u = 2y^2 - y^2 = y^2 \tag{7.1.9}$$

Make a drawing of the support of $f_{X,Y}$ to help to understand this better.

Ex 7.5. We have two continuous rvs X, Y. Suppose the joint CDF factors into the product of the marginals, i.e., $F_{X,Y}(x,y) = F_X(x)F_Y(y)$. Can it still be possible in general that the joint PDF does not factor into a product of marginals PDFs of X and Y, i.e., $f_{X,Y}(x,y) \neq f_X(x) f_Y(y)$?

s.7.5.

$$\partial_x \partial_y F_{X,Y}(x,y) = \partial_x \partial_y F_X(x) F_Y(y) = \partial_x F_X(x) \partial_y F_Y(y) = f_X(x) f_Y(y).$$

Ex 7.6. BH define the conditional CDF given an event A on page 416 as $F(y|A)$. Use this definition to write $F_{X,Y}(x,y)/F_X(x)$ as a conditional CDF. Is this equal to the conditional CDF of X and Y?

s.7.6.

$$\frac{F_{X,Y}(x,y)}{F_X(x)} = \frac{\mathsf{P}\left\{X \leq x, Y \leq y\right\}}{\mathsf{P}\left\{X \leq x\right\}} = \mathsf{P}\left\{Y \leq y, X \leq x | X \leq x\right\} = \mathsf{P}\left\{Y \leq y | X \leq x\right\}. \tag{7.1.10}$$

It is a big mistake to write $F_{X,Y}(x,y) = \mathsf{P}\left\{X = x, Y = y\right\}$. If you wrote this, recheck the definitions of BH.

Ex 7.7. Let X be uniformly distributed on the set $\{0,1,2\}$ and let $Y \sim \text{Bern}(1/4)$; X and Y are independent.

1. Present a contingency table for X and Y.
2. What is the interpretation of the column sums of the table?
3. What is the interpretation of the row sums of the table?
4. Suppose you would change some of the entries in the table. Are X and Y still independent?

s.7.7. $\mathsf{P}\{X = 0, Y = 0\} = 1/3 \cdot 3/4$, $\mathsf{P}\{X = 0, Y = 1\} = 1/3 \cdot 1/4$, and so on.

If we have one column with $Y = 0$ and the other with $Y = 1$, then the sum over the columns are $\mathsf{P}\{Y = 0\}$ and $\mathsf{P}\{Y = 1\}$. The row sum for row i are $\mathsf{P}\{X = i\}$.

Changing the values will (most of the time) make X and Y dependent. But, what if we changes the values such that $\mathsf{P}\{X = 0, Y = 0\} = 1$? Are X and Y then again independent? Check the conditions again.

Ex 7.8. A machine makes items on a day. Some items, independent of the other items, are failed (i.e., do not meet the quality requirements). What are N and p in the context of the chicken-egg story of BH? What are the 'eggs' in this context, and what is the meaning of 'hatching'? What type of 'hatching' do we have here?

s.7.8. The number of produced items (laid eggs) is N. The probability of hatching is p, that is, an item is ok. The hatched eggs are the good items.

Ex 7.9. We have two rvs X and Y on $\mathbb{R}^+$. It is given that $F_{X,Y}(x, y) = F_X(x)F_Y(y)$ for $x, y \leq 1/3$. It is true that then X and Y are necessarily independent.

s.7.9. For X, Y to be independent, it is necessary that $F_{X,Y}(x, y) = F_X(x)F_Y(y)$ for all x, y, not just one particular choice. (This is an example that satisfying a necessary condition is not necessarily sufficient.)

Ex 7.10. I select a random guy from the street, his height $X \sim \text{Norm}(1.8, 0.1)$, and I select a random woman from the street, her height is $Y \sim \text{Norm}(1.7, 0.08)$. I claim that since I selected the man and the woman independently, their heights are independent. Briefly comment on this claim.

s.7.10. Many answers are possible here, depending on extra assumptions you make. Here is one. Suppose, just by change, the fraction of taller guys in the street is a bit higher than the population fraction. Assuming that taller (shorter) people prefer taller (shorter) spouses, there must be a dependence between the height of the men and the women. This is because when selecting a man, I can also select his wife.

From this exercise you should memorize that *independence is a property of the joint CDF, not of the rvs.*

Mistake: $\mathsf{P}\{Y\}$ is wrong notation wrong because we can only compute the probability of an event, such as $\{Y \leq y\}$. But Y itself is not an event.

Ex 7.11. For any two rvs X and Y on $\mathbb{R}^+$ with marginals F_X and F_Y, can it hold that $\mathsf{P}\{X \leq x, Y \leq y\} = F_X(x)F_Y(y)$?

s.7.11. Only when X, Y are independent.

Mistake: independence of X and Y is not the same as the linear independence. Don't confuse these two types of dependene.

Ex 7.12. Theorem 7.1.11. What is the meaning of the notation $X|N = n$?

s.7.12. Given $N = n$, the random variable X has a certain distribution, here binomial.

Ex 7.13. Let X, Y be two discrete rvs with CDF $F_{X,Y}$. Can we compute the PDF as $\partial_x\partial_y F_{X,Y}(x, y)$?

s.7.13. This claim is incorrect, because X, Y are discrete, hence they have a PMF, not a PDF.

Mistake: Someone said that $\partial_x\partial_y$ is not correct notation; however, it is correct! It's a (much used) abbreviation of the much heaver $\partial^2/\partial x\partial y$. Next, the derivative of the PMF is not well-defined (at least, not within this course. If you object, ok, but then show that you passed a decent course on measure theory.)

Ex 7.14. Redo BH.7.1.24 with indicator functions and the fundamental bridge (recall, $\mathsf{P}\{A\} = \mathsf{E}[I_A]$ for an event A). (Indicators are often easy to use, and prevent many mistakes, as is demonstrated with this example.)

s.7.14.

$$\begin{aligned}
\mathsf{P}\{T_1 < T_2\} = \mathsf{E}\left[I_{T_1<T_2}\right] = &= \int_0^\infty \int_0^\infty I_{t_1<t_2} f_{T_1,T_2}(t_1, t_2)\, \mathrm{d}t_1\, \mathrm{d}t_2 \\
&= \int_0^\infty \int_{t_1}^\infty \lambda_1 e^{-\lambda_1 t_1} \lambda_2 e^{-\lambda_2 t_2}\, \mathrm{d}t_2\, \mathrm{d}t_1 \\
&= \int_0^\infty \lambda_1 e^{-\lambda_1 t_1} \lambda_2 \int_{t_1}^\infty e^{-\lambda_2 t_2}\, \mathrm{d}t_2\, \mathrm{d}t_1 \\
&= \int_0^\infty \lambda_1 e^{-\lambda_1 t_1} e^{-\lambda_2 t_1}\, \mathrm{d}t_1 \\
&= \int_0^\infty \lambda_1 e^{-\lambda_1 t_1 - \lambda_2 t_1}\, \mathrm{d}t_1 \\
&= \frac{\lambda_1}{\lambda_1 + \lambda_2}.
\end{aligned}$$

7.2 SECTION 7.2

Ex 7.15. BH.7.2.2. Write down the integral to compute $\mathsf{E}\left[(X-Y)^2\right]$, and solve it.

s.7.15. We have

$$\begin{aligned}
\mathsf{E}\left[(X-Y)^2\right] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-y)^2 f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y && (7.2.1)\\
&= \int_0^1\int_0^1 (x-y)^2\,\mathrm{d}x\,\mathrm{d}y && (7.2.2)\\
&= \int_0^1\int_0^1 (x^2-2xy+y^2)\,\mathrm{d}x\,\mathrm{d}y && (7.2.3)\\
&= \int_0^1\int_0^1 x^2\,\mathrm{d}x\,\mathrm{d}y - 2\int_0^1\int_0^1 xy\,\mathrm{d}x\,\mathrm{d}y + \int_0^1\int_0^1 y^2\,\mathrm{d}x\,\mathrm{d}y && (7.2.4)\\
&= \int_0^1 x^2\,\mathrm{d}x - 2\int_0^1\int_0^1 xy\,\mathrm{d}x\,\mathrm{d}y + \int_0^1 y^2\,\mathrm{d}y && (7.2.5)\\
&= 1/3 - 2\cdot 1/2\cdot 1/2 + 1/3. && (7.2.6)
\end{aligned}$$

Ex 7.16. Explain that for a continuous r.v. X with CDF F and a and b (so it might be that $a > b$),

$$\mathsf{P}\{a < X < b\} = [F(b)-F(a)]^+. \tag{7.2.7}$$

h.7.16. Recall that $F \in [0,1]$.

s.7.16.

$$\begin{aligned}
a < b \implies \mathsf{P}\{a < X < b\} &= F(b)-F(a) = [F(b)-F(a)]^+ && (7.2.8)\\
a \geq b \implies \mathsf{P}\{a < X < b\} &= 0 = [F(b)-F(a)]^+, && (7.2.9)
\end{aligned}$$

where the last equality follows from the fact that F is increasing.

Remark 7.17. If you are like me, you underestimate at first the importance of using indicator functions. In fact, they are extremely useful for several reasons.

1. They help to keep your formulas clean.

2. You can use them in computer code as logical conditions, or to help counting relevant events, something you need when numerically estimating multi-D integrals, for machine learning for instance.

3. Even though figures give geometrical insight into how to integrate over an 2D area, when it comes to reversing the sequence of integration, indicators are often easier to use.

4. In fact, *expectation is the fundamental concept in probability theory,* and the *probability of an event is defined* as
$$\mathsf{P}\{A\} := \mathsf{E}[I_A]. \tag{7.2.10}$$
Thus, the fundamental bridge is actually an application of LOTUS to indicator functions. Hence, reread BH.4.4!

Ex 7.18. What is $\int_{-\infty}^{\infty} I_{0\leq x\leq 3}\,\mathrm{d}x$?

s.7.18.

$$\int_{-\infty}^{\infty} I_{0\leq x\leq 3}\,\mathrm{d}x = \int_0^3 \mathrm{d}x = 3.$$

Ex 7.19. What is
$$\int x\, I_{0\leq x\leq 4}\,\mathrm{d}x? \tag{7.2.11}$$

s.7.19.

$$\int x\, I_{0\leq x\leq 4}\,\mathrm{d}x = \int_0^4 x\,\mathrm{d}x = 16/2 = 8.$$

When we do an integral over a 2D surface we can first integrate over the x and then over the y, or the other way around, whatever is the most convenient. (There are conditions about how to handle multi-D integral, but for this course these are irrelevant.)

Ex 7.20. What is
$$\iint xy\, I_{0\leq x\leq 3}\, I_{0\leq y\leq 4}\,\mathrm{d}x\,\mathrm{d}y? \tag{7.2.12}$$

s.7.20.

$$\begin{aligned}\iint xy\, I_{0\leq x\leq 3}\, I_{0\leq y\leq 4}\,\mathrm{d}x\,\mathrm{d}y &= \int_0^3 x \int_0^4 y\,\mathrm{d}y\,\mathrm{d}x \\ &= \int_0^3 x \frac{y^2}{2}\bigg|_0^4 \mathrm{d}x \\ &= \int_0^3 x\cdot 8\,\mathrm{d}x = 8\cdot 9/2 = 4\cdot 9.\end{aligned}$$

Ex 7.21. What is
$$\iint I_{0\leq x\leq 3}\, I_{0\leq y\leq 4}\, I_{x\leq y}\,\mathrm{d}x\,\mathrm{d}y? \tag{7.2.13}$$

s.7.21. Two solutions. First we integrate over y.

$$\begin{aligned}
\iint I_{0\leq x\leq 3}\, I_{0\leq y\leq 4}\, I_{x\leq y}\,\mathrm{d}x\,\mathrm{d}y &= \int I_{0\leq x\leq 3} \int I_{0\leq y\leq 4}\, I_{x\leq y}\,\mathrm{d}y\,\mathrm{d}x && (7.2.14)\\
&= \int I_{0\leq x\leq 3} \int I_{\max\{x,0\}\leq y\leq 4}\,\mathrm{d}y\,\mathrm{d}x && (7.2.15)\\
&= \int_0^3 \int_{\max\{x,0\}}^4 \mathrm{d}y\,\mathrm{d}x && (7.2.16)\\
&= \int_0^3 y\Big|_{\max\{x,0\}}^4 \mathrm{d}x && (7.2.17)\\
&= \int_0^3 (4-\max\{x,0\})\,\mathrm{d}x && (7.2.18)\\
&= 12 - \int_0^3 \max\{x,0\}\,\mathrm{d}x && (7.2.19)\\
&= 12 - \int_0^3 x\,\mathrm{d}x && (7.2.20)\\
&= 12 - 9/2. && (7.2.21)
\end{aligned}$$

Let's now instead first integrate over x.

$$\begin{aligned}
\iint I_{0\leq x\leq 3}\, I_{0\leq y\leq 4}\, I_{x\leq y}\,\mathrm{d}x\,\mathrm{d}y &= \int I_{0\leq y\leq 4} \int I_{0\leq x\leq 3}\, I_{x\leq y}\,\mathrm{d}x\,\mathrm{d}y && (7.2.22)\\
&= \int_0^4 \int I_{0\leq x\leq \min\{3,y\}}\,\mathrm{d}x\,\mathrm{d}y && (7.2.23)\\
&= \int_0^4 \int_0^{\min\{3,y\}} \mathrm{d}x\,\mathrm{d}y && (7.2.24)\\
&= \int_0^4 \min\{3,y\}\,\mathrm{d}y && (7.2.25)\\
&= \int_0^3 \min\{3,y\}\,\mathrm{d}y + \int_3^4 \min\{3,y\}\,\mathrm{d}y && (7.2.26)\\
&= \int_0^3 y\,\mathrm{d}y + \int_3^4 3\,\mathrm{d}y && (7.2.27)\\
&= 9/2 + 3. && (7.2.28)
\end{aligned}$$

Ex 7.22. Take $X \sim \text{Unif}([1,3])$, $Y \sim \text{Unif}([2,4])$ and independent. Compute

$$\mathsf{P}\{Y \leq 2X\}. \qquad (7.2.29)$$

s.7.22. Take c the normalization constant (why is $c = 1/4$), then using the previous exercise

$$\begin{aligned} \mathsf{P}\{Y \leq 2X\} &= \mathsf{E}[I_{Y\leq 2X}] && (7.2.30)\\ &= c\int_1^3\int_2^4 I_{y\leq 2x}\,\mathrm{d}y\,\mathrm{d}x && (7.2.31)\\ &= c\int_1^3\int I_{2\leq y\leq \min\{4,2x\}}\,\mathrm{d}y\,\mathrm{d}x && (7.2.32)\\ &= c\int_1^3 (\min\{4,2x\}-2)\,\mathrm{d}x && (7.2.33)\end{aligned}$$

Now make a drawing of the function $(\min\{4,2x\}-2)$ on the interval $[1,3]$ to see that

$$\int_1^3 (min\{4,2x\}-2)\,\mathrm{d}x = \int_1^2 (2x-2)\,\mathrm{d}x + \int_2^3 (4-2)\,\mathrm{d}x. \qquad (7.2.34)$$

I leave the rest of the computation to you.

7.3 SECTION 7.3

Ex 7.23. Give a brief example of a situation where it might be more convenient to employ the correlation than the covariance. Explain why.

s.7.23. The covariance might be a large number, which may suggest that the rvs X and Y are 'very' dependent. However, when $\mathsf{V}[X]$ and $\mathsf{V}[Y]$ are also large, the correlation can be small. Thus, correlation is a scaled type of covariance.

Ex 7.24. In queueing theory the concept of squared coefficient of variance SCV of a rv X is very important. It is defined as $C = \mathsf{V}[X]/(\mathsf{E}[X])^2$. Is the SCV of X equal to $\text{Corr}(X, X)$? Can it happen that $C > 1$?

s.7.24. Answers: no and yes.

We have

$$C = \frac{\mathsf{V}[X]}{(\mathsf{E}[X])^2}, \tag{7.3.1}$$

which does not equal

$$\text{Corr}(X, X) = \frac{\mathsf{Cov}[X, X]}{\sqrt{\mathsf{V}[X]\,\mathsf{V}[X]}} = 1 \tag{7.3.2}$$

in general (for instance, consider a degenerate random variable $X \equiv 1$). Next, consider a $N(1, 100)$ random variable. Then,

$$C = 100/(1^2) = 100 > 1. \tag{7.3.3}$$

Ex 7.25. Prove the key properties 1–5 of the covariance below BH.7.3.2.

s.7.25.

1. We have

$$\mathsf{Cov}[X, X] = \mathsf{E}[XX] - \mathsf{E}[X]\,\mathsf{E}[X] = \mathsf{E}\left[X^2\right] - \mathsf{E}[X]^2 = \mathsf{V}[X]. \tag{7.3.4}$$

2. We have

$$\mathsf{Cov}[X, Y] = \mathsf{E}[XY] - \mathsf{E}[X]\,\mathsf{E}[Y] = \mathsf{E}[YX] - \mathsf{E}[Y]\,\mathsf{E}[X] = \mathsf{Cov}[Y, X]. \tag{7.3.5}$$

3. We have

$$\mathsf{Cov}[X, c] = \mathsf{E}[Xc] - \mathsf{E}[X]\,\mathsf{E}[c] = c\,\mathsf{E}[X] - c\,\mathsf{E}[X] = 0. \tag{7.3.6}$$

4. We have

$$\mathsf{Cov}[aX, Y] = \mathsf{E}[aXY] - \mathsf{E}[aX]\,\mathsf{E}[Y] = a\big(\mathsf{E}[XY] - \mathsf{E}[X]\,\mathsf{E}[Y]\big) = a\,\mathsf{Cov}[X, Y]. \tag{7.3.7}$$

5. We have

$$\begin{aligned}
\mathsf{Cov}[X+Y,Z] &= \mathsf{E}[(X+Y)Z] - \mathsf{E}[X+Y]\,\mathsf{E}[Z] && (7.3.8)\\
&= \mathsf{E}[XZ+YZ] - \big(\mathsf{E}[X]+\mathsf{E}[Y]\big)\mathsf{E}[Z] && (7.3.9)\\
&= \mathsf{E}[XZ] - \mathsf{E}[X]\,\mathsf{E}[Z] + \mathsf{E}[YZ] - \mathsf{E}[Y]\,\mathsf{E}[Z] && (7.3.10)\\
&= \mathsf{Cov}[X,Z] + \mathsf{Cov}[Y,Z]. && (7.3.11)
\end{aligned}$$

Ex 7.26. Using the definition of Covariance (BH.7.3.1) derive the expression $\mathsf{Cov}[X,Y] = \mathsf{E}[XY] - \mathsf{E}[X]\,\mathsf{E}[Y]$. Use this to show why independence of X and Y implies their uncorrelatedness (Note that the converse does not hold).

s.7.26. We have

$$\begin{aligned}
\mathsf{Cov}[X,Y] &= \mathsf{E}[(X-\mathsf{E}[X])(Y-\mathsf{E}[Y])] && (7.3.12)\\
&= \mathsf{E}[XY - X\,\mathsf{E}[Y] - Y\,\mathsf{E}[X] + \mathsf{E}[X]\,\mathsf{E}[Y]] && (7.3.13)\\
&= \mathsf{E}[XY] - \mathsf{E}[X]\,\mathsf{E}[Y] - \mathsf{E}[Y]\,\mathsf{E}[X] + \mathsf{E}[X]\,\mathsf{E}[Y] && (7.3.14)\\
&= \mathsf{E}[XY] - \mathsf{E}[X]\,\mathsf{E}[Y]. && (7.3.15)
\end{aligned}$$

When X and Y are independent, then $\mathsf{E}[XY] = \mathsf{E}[X]\,\mathsf{E}[Y]$, and then $\mathsf{Cov}[X,Y] = 0$.

Ex 7.27. Let U, V be two rvs and let $a, b \in \mathbb{R}$. Use the previous question to express $\mathsf{Cov}[a(U+V), b(U-V)]$ in terms of $\mathsf{V}[U]$, $\mathsf{V}[V]$ and $\mathsf{Cov}[U,V]$.

s.7.27. By linearity of the covariance wea have

$$\begin{aligned}
\mathsf{Cov}[a(U+V), b(U-V)] &= a\Big(\mathsf{Cov}[U, b(U-V)] + \mathsf{Cov}[V, b(U-V)]\Big) && (7.3.16)\\
&= a\Big(b\big(\mathsf{Cov}[U,U] - \mathsf{Cov}[U,V]\big) + b\big(\mathsf{Cov}[V,U] - \mathsf{Cov}[V,V]\Big) && (7.3.17)\\
&= a\Big(b\big(\mathsf{Cov}[U,U] - \mathsf{Cov}[U,V]\big) + b\big(\mathsf{Cov}[V,U] - \mathsf{Cov}[V,V]\Big) && (7.3.18)\\
&= ab\Big(\mathsf{V}[U] - \mathsf{Cov}[U,V] + \mathsf{Cov}[V,U] - \mathsf{V}[V]\Big) && (7.3.19)\\
&= ab\Big(\mathsf{V}[U] - \mathsf{V}[V]\Big). && (7.3.20)
\end{aligned}$$

Alternatively one can also use the result from BH.7.1.26, according to which $\mathsf{Cov}[X,Y] = \mathsf{E}[XY] - \mathsf{E}[X]\,\mathsf{E}[Y]$.

Ex 7.28. The solution of BH.7.3.6 is a somewhat tricky; I would have not found this trick myself. Here is an approach that is trick free.

Neglecting the event $\{X = Y\}$ as this has zero probability, we know that $M = X, L = Y$ or $M = Y, L = X$. Use this idea and the formula $\mathsf{Cov}[M,L] = \mathsf{E}[ML] - \mathsf{E}[M]\,\mathsf{E}[L]$ to derive the result of this example.

h.7.28. Realize that $\mathsf{E}[ML] = \mathsf{E}[XY]$.

s.7.28. With the hint: $\mathsf{E}[XY] = 1/\lambda^2$, when $X, Y \sim \text{Exp}(\lambda)$. Then, $L \sim \text{Exp}(2\lambda)$, since $f_L(x) = 2f_X(x)(1-F_Y(x)) = 2\lambda e^{-2\lambda x}$. Therefore, $\mathsf{E}[L] = 1/2\lambda$. Also, by memoryless, $\mathsf{E}[M] = \mathsf{E}[L] + \mathsf{E}[X] = 3/2\lambda$. Hence, $\mathsf{E}[M]\,\mathsf{E}[L] = 3/4\lambda^2$. Hence, $\mathsf{E}[ML] - \mathsf{E}[M]\,\mathsf{E}[L] = 1/\lambda^{2-}3/4\lambda^2 = 1/4\lambda^2$.

7.4 SECTION 7.4

Ex 7.29. Come up with a short illustrative example in which the random vector $\mathbf{X} = (X_1, \ldots, X_6)$ follows a Multinomial Distribution with parameters $n = 10$ and $\mathbf{p} = (\frac{1}{6}, \ldots, \frac{1}{6}) \in \mathbb{R}^6$.

s.7.29. We throw 10 fair dice. X_i denotes the number of dice that show the number i, $i = 1, \ldots, 6$.

7.5 SECTION 7.5

Ex 7.30. Is the following claim correct? If the rvs X, Y are both normally distributed, then (X, Y) follows a Bivariate Normal distribution.

s.7.30. No, this does not always hold, see BH.7.5.2. However, it does hold when X and Y are independent.

Ex 7.31. Let X, Y, Z be iid $\mathcal{N}(0,1)$. Determine whether or not the random vector

$$\mathbf{W} = (X + 2Y, 3X + 4Z, 5Y + 6Z, 2X - 4Y + Z, X - 9Z, 12X + \sqrt{3}Y - \pi Z + 18)$$

is Multivariate Normal. (Explain in words, don't do a lot of tedious math here!)

s.7.31. Since X, Y, Z are independent normally distributed variables, (X, Y, Z) is multivariate normally distributed. Hence, every linear combination of X, Y, Z is normally distributed. Note that every linear combination of the elements of W can be written as a linear combination of X, Y, Z. Hence, every linear combination of the elements of W is normally distributed. Hence, W is multivariate normally distributed.

7.6 EXERCISES ON 2D INTEGRATION

Here is some extra material for you practice on 2D integration, indicators and 2D LOTUS. These exercises are old exam questions, hence quite a bit harder than the above. They form important training.

Ex 7.32. Let X and Y be continuous random variables. Furthermore, $F(x,y)$ is the joint cumulative distribution function of X and Y. This function has the following properties.

$$
\begin{aligned}
&1. \quad F(x,y) = \frac{1}{8}(x-1)^2(y-2) && \text{for } 1 < x < 3 \text{ and } 2 < y < 4,\\
&2. \quad \frac{\partial F(x,y)}{\partial x} = 0 && \text{for } x \notin (1,3),\\
&3. \quad \frac{\partial F(x,y)}{\partial y} = 0 && \text{for } y \notin (2,4).
\end{aligned}
$$

Use these properties to answer the following questions.

1. What is $F(2,5)$?
2. Determine the joint probability density function of X and Y.
3. Determine $\mathrm{P}(2 < X < 3, 2 < Y < 4)$.
4. Determine the joint probability $\mathrm{P}(Y < 2X, 2X + Y > 6)$. Clearly draw the area over which you integrate.

s.7.32. a.

$$F(2,5) = P(X \leq 2, Y \leq 5) = P(X \leq 2, Y \leq 4) = F(2,4) = \frac{1}{4}$$

The second step is valid since the cumulative distribution function does not change by changing y from 5 to 4 thanks to property 3.

b. To obtain the joint pdf, use that $f_{X,Y}(x,y) = \frac{\partial^2}{\partial y \partial x} F(x,y) = \frac{\partial}{\partial y}\left(\frac{\partial}{\partial x} F(x,y)\right)$.

Since $\frac{\partial}{\partial x} F(x,y) = \frac{1}{4}(x-1)(y-2)$ for $1 < x < 3$, and $\frac{\partial}{\partial x} F(x,y) = 0$ for other values of x, we have that

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{4}(x-1), & \text{for } 1 < x < 3 \text{ and } 2 < y < 4,\\ 0, & \text{elsewhere.} \end{cases}$$

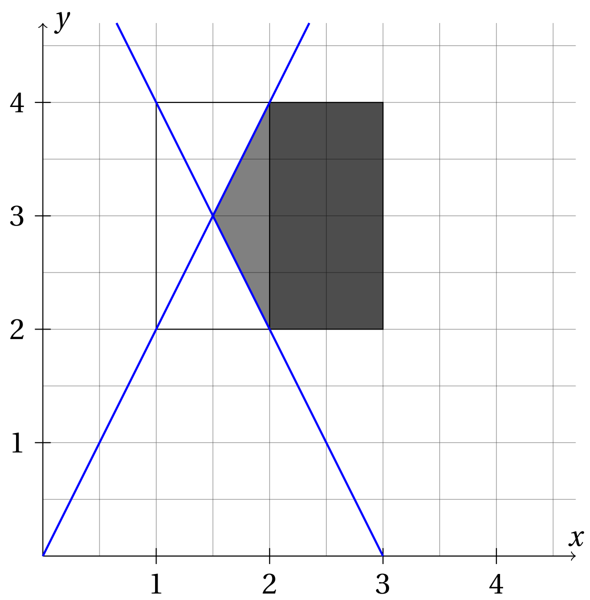
c. The simplest way of solving this question is by writing

$$
\begin{aligned}
P(2 < X < 3, 2 < Y < 4) &= F(3,4) - F(2,4) - F(3,2) + F(2,2)\\
&= 1 - \frac{1}{4} - 0 + 0 = \frac{3}{4}.
\end{aligned}
$$

Alternatively, one can integrate over the pdf from (b) to obtain the same result:

$$
\begin{aligned}
P(2 < X < 3, 2 < Y < 4) &= \int_2^3 \int_2^4 f(x,y)\mathrm{d}y\mathrm{d}x \\
&= \frac{1}{4}\int_2^3 (x-1)y\Big|_{y=2}^{y=4}\mathrm{d}x \\
&= \frac{1}{4}\left(x^2 - 2x\Big|_2^3\right) = \frac{1}{4}(9-6-4+4) = \frac{3}{4}.
\end{aligned}
$$

d. First, draw the integration area:



The domain on which the density is non-zero, is the complete shaded area. The downward-sloping line represents $2x + y = 6$ and the upward-sloping line is $y = 2x$.

We already know the integral over the dark shaded area from the previous subquestion. What remains is the lighter shaded triangular part on the left side.

First, we need to calculate the intersection of the two curves, which can be found by solving $6 - 2x = 2x$, which gives $x = \frac{3}{2}$, and consequently $y = 3$.

The integral of the triangular part of the dark shaded region is

$$\begin{aligned}
\int_{3/2}^{2}\int_{6-2x}^{2x}\frac{1}{4}(x-1)\,dy\,dx &= \int_{3/2}^{2}\frac{1}{4}(x-1)y\Big|_{6-2x}^{2x}dx \\
&= \frac{1}{4}\int_{3/2}^{2}(x-1)2x-(x-1)(6-2x)\,dx \\
&= \frac{1}{4}\int_{3/2}^{2}(x-1)(4x-6)\,dx \\
&= \frac{1}{4}[\frac{4}{3}x^3-5x^2+6x]_{3/2}^{2} \\
&= \frac{5}{48}\approx 0.1042
\end{aligned}$$

Finally, the joint probability asked for in the question is given by

$$P(Y<2X, 2X+Y>6)=\frac{5}{48}+\frac{3}{4}=\frac{41}{48}\approx 0.8542$$

Ex 7.33. Suppose X and V are independent, $X\sim \text{Expo}(\lambda)$ and $V\sim\text{Expo}(\mu)$. Define the ratio $R=X/V$ and derive the cumulative distribution function (CDF) of R. Provide at least two checks on the CDF to make sure that your result is indeed a valid CDF. Note: There is no need to derive the probability density function (PDF) of R.

s.7.33. For $r\geq 0$, we have

$$\begin{aligned}
F_R(r) &= P(R\leq r) \\
&= P(X\leq Vr) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{vr} f_{X,V}(x,v)\,dx\,dv \\
&= \lambda\mu\int_{0}^{\infty}\int_{0}^{vr} e^{-\lambda x}e^{-\mu v}\,dx\,dv \\
&= \lambda\mu\int_{0}^{\infty} e^{-\mu v}\left[-\frac{1}{\lambda}e^{-\lambda x}\right]_{0}^{vr}dv \\
&= -\mu\int_{0}^{\infty} e^{-\mu v}(e^{-\lambda vr}-1)\,dv \\
&= -\mu\left[-\frac{1}{\mu+\lambda r}e^{-(\mu+\lambda r)v}+\frac{1}{\mu}e^{-\mu v}\right]_{0}^{\infty} \\
&= -\mu\left[\frac{1}{\mu+\lambda r}-\frac{1}{\mu}\right] \\
&= \frac{\lambda r}{\mu+\lambda r},
\end{aligned}$$

while $F_R(r)=0$ when $r<0$ since both X and V are nonnegative.

We see that (1) $F_R(-\infty)=0$, (2) $F_R(\infty)=1$, and (3) $F_R(r)$ is monotonically increasing in r, so $F_R(r)$ satisfies the conditions for being a valid CDF.

Ex 7.34. Consider the following joint density function

$$f_{X,Y}(x,y) = \begin{cases} cxy & \text{for } 0 \le x < \frac{1}{2} \text{ and } 0 \le y \le x, \\ cxy & \text{for } \frac{1}{2} \le x \le 1 \text{ and } 0 \le y \le 1-x, \\ 0 & \text{otherwise.} \end{cases}$$

1. What is the correct value of the constant c?
2. Derive the conditional probability density function $f_{X|Y}(x|y)$. Verify that your result is indeed a valid density function.

s.7.34. a. We integrate $f_{X,Y}(x,y)$ over its domain

$$\begin{aligned}
&\int_0^{1/2}\int_0^x cxydydx + \int_{1/2}^1\int_0^{1-x} cxydydx \\
=&c\left[\frac{1}{2}\int_0^{1/2} x^3dx + \frac{1}{2}\int_{1/2}^1 x(1-x)^2dx\right] \\
=&\frac{1}{2}c\left\{\left[\frac{1}{4}x^4\right]_0^{1/2} + \left[\frac{1}{2}x^2 - \frac{2}{3}x^3 + \frac{1}{4}x^4\right]_{1/2}^1\right\} \\
&= \frac{1}{2}c\left\{\frac{1}{4}\frac{1}{16} + \frac{1}{2} - \frac{2}{3} + \frac{1}{4} - \frac{1}{2}\frac{1}{4} + \frac{2}{3}\frac{1}{8} - \frac{1}{4}\frac{1}{16}\right\} \\
&= \frac{1}{2}c\frac{12-16+6-3+2}{24} \\
&= \frac{c}{48}
\end{aligned}$$

Since this integral should equal 1, $c = 48$.

Alternative Rewrite the probability density function to

$$f_{X,Y}(x,y) = \begin{cases} cxy & 0 \le y \le \frac{1}{2}, \quad y \le x \le 1-y \\ 0 & \text{otherwise} \end{cases}$$

Then,

$$\begin{aligned}
\int_0^{1/2}\int_y^{1-y} cxydxdy &= c\int_0^{1/2} y\frac{1}{2}\left((1-y)^2 - y^2\right) \\
&= \frac{1}{2}c\int_0^{1/2}\left(y - 2y^2\right)dy \\
&= \frac{1}{2}c\left[\frac{1}{2}y^2 - \frac{2}{3}y^3\right]_0^{1/2} \\
&= \frac{1}{2}c\left[\frac{1}{8} - \frac{2}{3}\frac{1}{8}\right] = \frac{c}{48}
\end{aligned}$$

Since this integral should equal 1, $c = 48$.

b. The conditional density function is given by

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

We first need the marginal density of Y.

$$\begin{aligned}
f_Y(y) &= 48\int_y^{1-y} xy\,dx \\
&= 48y\left[\frac{1}{2}(1-y)^2 - \frac{1}{2}y^2\right] \\
&= 24y\left[1-2y+y^2-y^2\right] \\
&= 24y(1-2y) \quad \text{for } 0 \le y \le \frac{1}{2}
\end{aligned}$$

and $f_Y(y) = 0$ otherwise.

Not required: We can check that this is a valid density function:

$$\begin{aligned}
\int_0^{1/2} 24y(1-2y)\,dy &= 24\left[\frac{1}{2}y^2 - \frac{2}{3}y^3\right]_0^{1/2} \\
&= 24\left[\frac{1}{2}\frac{1}{4} - \frac{2}{3}\frac{1}{8}\right]_0^{1/2} \\
&= 1
\end{aligned}$$

---

b. Now we can obtain the conditional density function

$$\begin{aligned}
f_{X|Y}(x|y) &= \frac{f_{X,Y}(x,y)}{f_Y(y)} \\
&= \frac{48xy}{24y(1-2y)} = \frac{2x}{1-2y} \quad \text{for } y \le x \le 1-y, \text{ and } 0 \le y < \frac{1}{2}
\end{aligned}$$

and $f_{X|Y}(x|y) = 0$ otherwise.

This is a valid density function since $f(X|Y)(x|y) \ge 0$, and

$$\begin{aligned}
\int_y^{1-y} f_{X|Y}(x|y)\,dx &= \frac{2}{1-2y}\left[\frac{1}{2}x^2\right]_y^{1-y} \\
&= 1
\end{aligned}$$

Ex 7.35. Suppose the random variables X and Y have the joint probability density function

$$f_{X,Y}(x,y) = \begin{cases} 2 & \text{if } 0 \le x < 1, \quad |y| < \frac{1}{2}(1-x), \\ 0 & \text{otherwise} \end{cases}$$

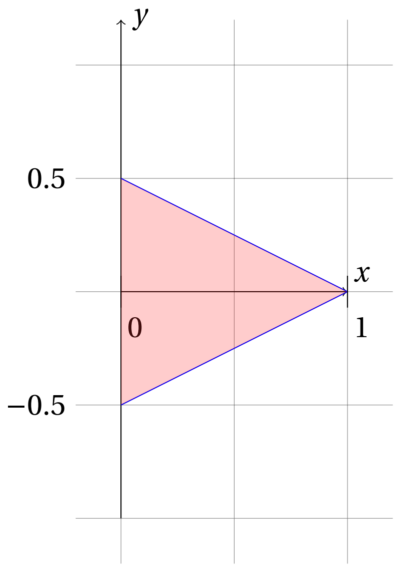
1. Calculate the marginal probability density functions $f_X(x)$ and $f_Y(y)$ and show that these are valid probability density functions.

2. Find the conditional expectation $\mathrm{E}[X|Y=y]$. Provide at least one 'sanity check' that shows that your answer makes intuitive sense. If you did not find an answer to (a), you can use that

$$f_Y(y) = \begin{cases} 2(1-2|y|) & \text{if } -\frac{1}{2} < y < \frac{1}{2}, \\ 0 & \text{otherwise.} \end{cases}$$

3. Calculate the joint cumulative distribution function $F_{X,Y}(x,y)$ for $x = \frac{1}{2}$ and $y = 3$.

s.7.35. a. The joint PDF is nonzero above the red-shaded area in the following graph. (draw, draw, draw!)



For the PDF for Y, using the graph above,

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx \\ &= \int_0^{1-2|y|} 2\,dx \\ &= 2(1-2|y|) \quad \text{for } -1/2 < y < 1/2, \end{aligned}$$

and 0 otherwise. Since $-1/2 < y < 1/2$, we have $f_Y(y) \geq 0$. Also,

$$\begin{aligned} \int_{-\infty}^{\infty} f_Y(y)\,dy &= \int_{-1/2}^{1/2} 2(1-2|y|)\,dy \\ &= 2\int_0^{1/2}(1-2y)\,dy + 2\int_{-1/2}^{0}(1+2y)\,dy \\ &= 2\left(\frac{1}{2} - \frac{1}{4} + \frac{1}{2} - \frac{1}{4}\right) = 1. \end{aligned}$$

Probability density function for X.

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}\, dy \\ &= \int_{-\frac{1}{2}(1-x)}^{\frac{1}{2}(1-x)} 2dy \\ &= 2(1-x) \quad \text{for } 0 \le x < 1, \end{aligned}$$

and 0 otherwise. We have $f_X(x) \ge 0$, and also

$$\begin{aligned} \int_{-\infty}^{\infty} f_X(x)dx &= \int_0^1 2(1-x)dx \\ &= 2\left[x - \frac{1}{2}x^2\right]_0^1 \\ &= 2\left(1-\frac{1}{2}\right) = 1. \end{aligned}$$

Since $-1/2 < y < 1/2$, we have $f_Y(y) \ge 0$. Also,

$$\begin{aligned} \int_{-\infty}^{\infty} f_Y(y)dy &= \int_{-1/2}^{1/2} 2(1-2|y|)dy \\ &= 2\int_0^{1/2}(1-2y)dy + 2\int_{-1/2}^{0}(1+2y)dy \\ &= 2\left(\frac{1}{2}-\frac{1}{4}+\frac{1}{2}-\frac{1}{4}\right) = 1. \end{aligned}$$

Probability density function for X.

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}\, dy \\ &= \int_{-\frac{1}{2}(1+x)}^{\frac{1}{2}(1+x)} 2dy \\ &= 2(1+x) \quad \text{for } -1 < x \le 0, \end{aligned}$$

and 0 otherwise. We have $f_X(x) \ge 0$, and also

$$\begin{aligned} \int_{-\infty}^{\infty} f_X(x)dx &= \int_{-1}^0 2(1+x)dx \\ &= 2\left[x + \frac{1}{2}x^2\right]_{-1}^0 \\ &= 2\left(1-\frac{1}{2}\right) = 1. \end{aligned}$$

b. Using the definition of a conditional probability density function, we have

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{1}{1-2|y|} \quad \text{for } -\frac{1}{2} < y < \frac{1}{2}, 0 < x < 1-2|y|$$

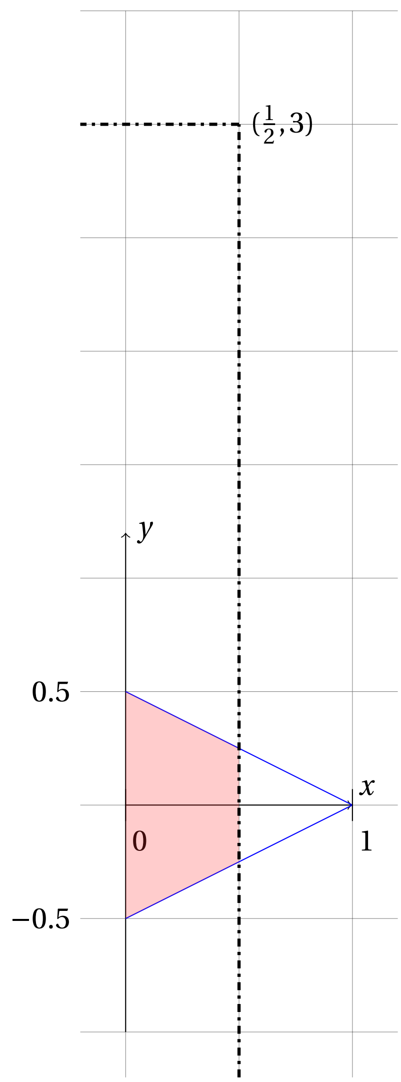
and 0 otherwise.

The required expected value is (bounds are crucial!)

$$\mathrm{E}[X|Y=y]=\int_0^{1-2|y|}\frac{x}{1-2|y|}dx=\frac{1}{2}(1-2|y|).$$

If $y=-1/2$ or $y=1/2$, we find $\mathrm{E}[X|Y=y]=0$, which makes sense based on the figure above. Similarly, if $y=0$, we find $\mathrm{E}[X|Y=y]=1/2$ since the density of x is uniform over $[0,1]$.

The point $(\frac{1}{2},3)$ is located as in the picture below. We need to integrate $f_{X,Y}(x,y)$ over the entire area on the lower left side of this point. It is crucial to realize that the joint PDF is only nonzero in the red shaded area. Moreover, it's very helpful that the PDF has a constant value of 2 above this area. So we just need to calculate the the area of the red shape in the following graph and multiply this by 2.



The dash dotted line intersects the line $y = \frac{1}{2}(1-x)$ at $x = \frac{1}{2}$, so $y = \frac{1}{4}$. The line intersects $y = -\frac{1}{2}(1-x)$ at $x = \frac{1}{2}$, so $y = -\frac{1}{4}$. The whole area of the triangle is $\frac{1}{2}$. The area *not* in red is $\frac{1}{2}\cdot(\frac{1}{4}-(-\frac{1}{4}))\cdot(1-\frac{1}{2}) = \frac{1}{8}$. So

$$F_{X,Y}\left(\frac{1}{2},3\right) = 2\left(\frac{1}{2}-\frac{1}{8}\right) = \frac{3}{4}.$$

Ex 7.36. Suppose the joint probability density function of X and Y is given by

$$f_{X,Y}(x,y) = \frac{c}{1-x}, \quad 0<x+y<1, \quad x>0, \quad y>0$$

and $f_{X,Y}(x,y) = 0$ otherwise.

1. For what value of the constant c is $f_{X,Y}(x,y)$ a joint probability density function?
2. What is the probability that $X+Y > \frac{1}{2}$?
3. What is the probability that both X and Y are smaller than $\frac{1}{2}$ given that $X+Y > \frac{1}{2}$?

s.7.36. a. $f_{X,Y}(x,y)$ is a joint probability density function if

1. If $f_{X,Y}(x,y)$ satisfies

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = 1$$

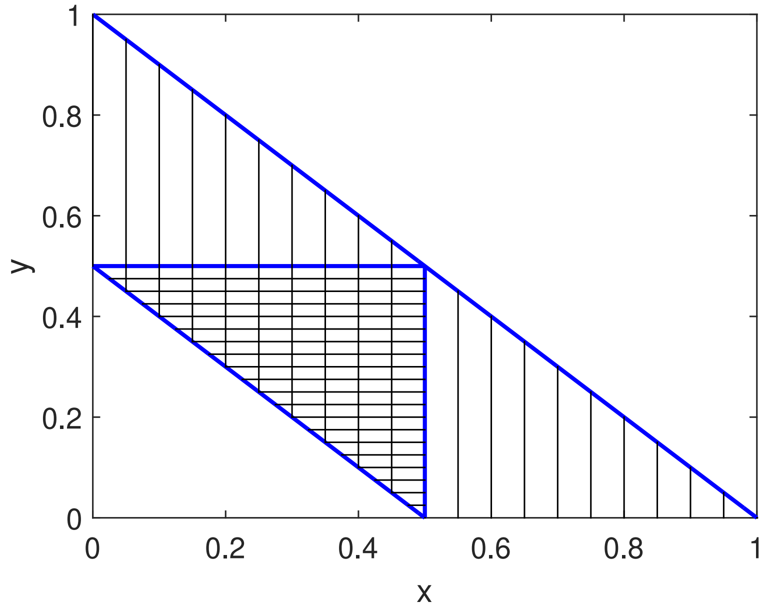
2. $f_{X,Y}(x,y) \geq 0$ for all x and y.

We have

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy\,dx &= \int_0^1\int_0^{1-x}\frac{c}{1-x}\,dy\,dx\\ &= c\int_0^1 \frac{1-x}{1-x}\,dx = 1\end{aligned}$$

So to satisfy condition (1), we need to set $c=1$.
Check Condition (2): $f_{X,Y}(x,y) \geq 0$ for all x,y.
For $0<x<1$, $f_{X,Y}(x,y) > 0$. Outside of this interval $f_{X,Y}(x,y) = 0$. So, we have that $f_{X,Y}(x,y) \geq 0$ for all x,y.

b. The following graph is used for questions b and c.



To obtain $P(X+Y>1/2)$, we need to integrate $f_{X,Y}$ over the vertically hatched area. We do this by integrating over the larger triangle defined by $x+y<1$, and then subtract the white triangle in the lower left corner defined by $x+y<\frac{1}{2}$.

$$\begin{aligned}
P\left(X+Y>\frac{1}{2}\right) &= \int_0^1 \int_0^{1-x} \frac{1}{1-x} dy dx - \int_0^{1/2} \int_0^{1/2-x} \frac{1}{1-x} dy dx \\
&= 1 - \int_0^{1/2} \frac{1/2-x}{1-x} dx \\
&= 1 - \int_0^{1/2} \left(1 - \frac{1}{2}\frac{1}{1-x}\right) dx \\
&= 1 - \frac{1}{2} + \frac{1}{2}\int_0^{1/2} \frac{1}{1-x} dx \\
&= \frac{1}{2} + \frac{1}{2}\left[-\ln(1-x)\right]_0^{1/2} \\
&= \frac{1}{2} - \frac{1}{2}\ln\left(\frac{1}{2}\right) = 0.8466
\end{aligned}$$

c. Using the definition of conditional probability

$$P\left(X<\frac{1}{2}, Y<\frac{1}{2} \middle| X+Y>\frac{1}{2}\right) = \frac{P\left(X<\frac{1}{2}, Y<\frac{1}{2}, X+Y>\frac{1}{2}\right)}{P\left(X+Y>\frac{1}{2}\right)}$$

The integral in the numerator is the integral over the horizontally hatched triangle. Easiest is to first integrate over the square $[0,1/2] \times [0,1/2]$ and then subtract the lower white triangle, which we have already calculated in the previous question to be $\frac{1}{2} + \frac{1}{2}\ln\left(\frac{1}{2}\right)$.

$$\begin{aligned}
\int_0^{1/2}\int_0^{1/2} \frac{1}{1-x} dy dx &= \frac{1}{2}\int_0^{1/2} \frac{1}{1-x} dx \\
&= \frac{1}{2}\left[-\ln(1-x)\right]_0^{1/2} \\
&= -\frac{1}{2}\ln\left(\frac{1}{2}\right)
\end{aligned}$$

Subtracting the lower triangle from the square, we see that the integral over the horizontally hatched triangle equals

$$-\frac{1}{2}\ln\left(\frac{1}{2}\right) - \frac{1}{2} - \frac{1}{2}\ln\left(\frac{1}{2}\right) = -\frac{1}{2} - \ln\left(\frac{1}{2}\right)$$

We can then calculate

$$P\left(X < \frac{1}{2}, Y < \frac{1}{2} | X + Y > \frac{1}{2}\right) = \frac{-\frac{1}{2} - \ln\left(\frac{1}{2}\right)}{\frac{1}{2} - \frac{1}{2}\ln\left(\frac{1}{2}\right)} \approx 0.23$$

Ex 7.37. X and Y be random variables with joint probability density function

$$f_{X,Y}(x,y) = \begin{cases} \frac{3}{16}xy^2, & 0 \le x \le c \text{ and } 0 \le y \le c, \\ 0, & \text{elsewhere} \end{cases}$$

where $c > 0$ is a real number.

1. Show that $c = 2$.
2. Show that $P(X + Y > 2) = \frac{9}{10}$. Start by making a clear sketch of the area in the (x,y)-plane over which you take the required integral.
3. Calculate the conditional probability $P\left(Y < X^2 \middle| X + Y < 2\right)$.

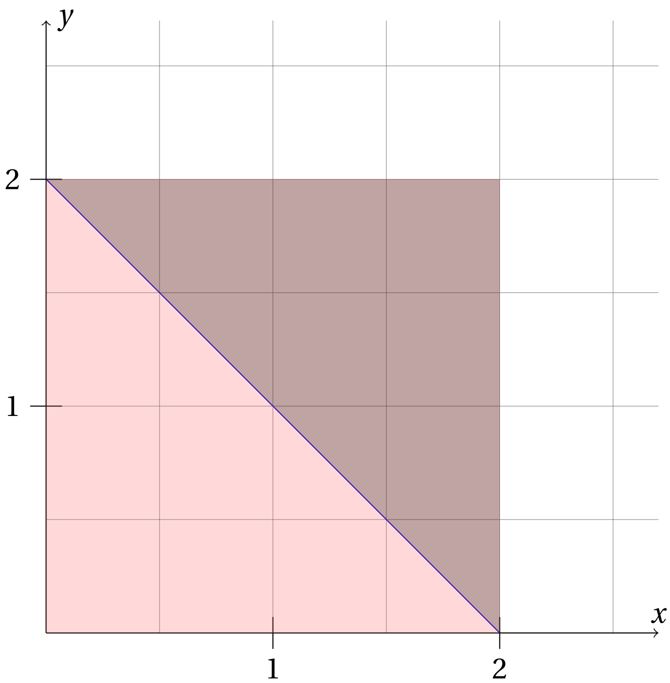
s.7.37. a. Since $f_{XY}(x,y)$ is a joint probability density function, we should have
1. $f_{X,Y}(x,y) \ge 0$. This is satisfied since $x, y \ge 0$. 2.

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\mathrm{d}x\mathrm{d}y = 1,$$

. We can calculate the integral as follows.

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\mathrm{d}x\mathrm{d}y = 1 &\iff \frac{3}{16}\int_0^c\int_0^c xy^2\mathrm{d}x\mathrm{d}y = 1 \\
&\iff \frac{3}{32}\int_0^c \left(x^2y^2\Big|_0^c\right)\mathrm{d}y = 1 \\
&\iff \frac{3c^2}{32}\int_0^c y^2\mathrm{d}y = 1 \\
&\iff \frac{3c^2}{96}y^3\Big|_0^c = 1 \\
&\iff 3c^5 = 96 \iff c = 2
\end{aligned}$$

b. First, draw the area over which the integral is taken.



We want to integrate over the darkest area. Hence, for every value of y, x varies between $2-y$ and 2. Hence, the required probability can be calculated as follows:

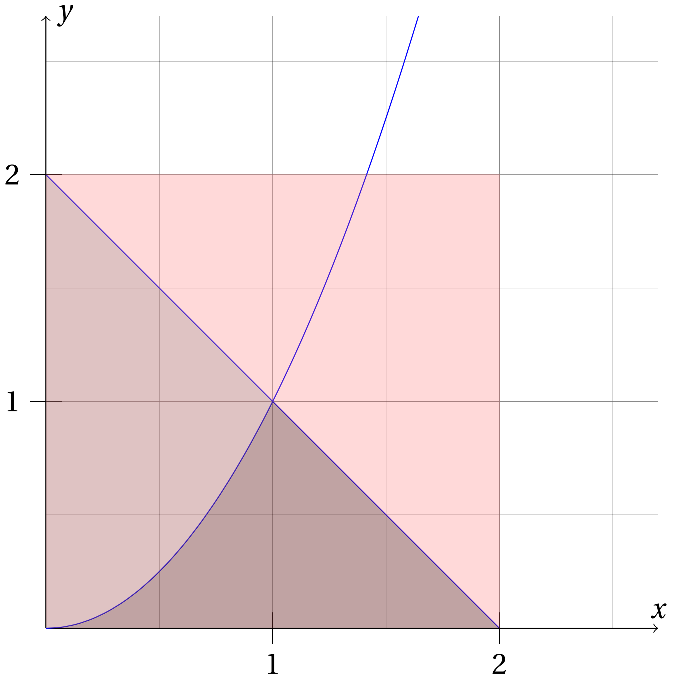
Solution 1

$$
\begin{aligned}
&P(X+Y>2)= \\
&=\int_0^2\int_{2-y}^2 f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y \\
&=\frac{3}{16}\int_0^2\int_{2-y}^2 xy^2\mathrm{d}x\mathrm{d}y \\
&=\frac{3}{32}\int_0^2\left(x^2y^2\Big|_{2-y}^2\right)\mathrm{d}y \\
&=\frac{3}{32}\int_0^2(4y^2-(2-y)^2y^2)\mathrm{d}y \\
&=\frac{3}{32}\int_0^2(4y^3-y^4)\mathrm{d}y \\
&=\frac{3}{32}\left(y^4-\frac{1}{5}y^5\Big|_0^2\right) \\
&=\frac{3}{32}\left(16-\frac{32}{5}\right) \\
&=\frac{3}{2}-\frac{3}{5}=\frac{9}{10}
\end{aligned}
$$

Solution 2

$$
\begin{aligned}
&P(X+Y>2)= \\
&=\int_0^2\int_{2-x}^2 f_{X,Y}(x,y)\mathrm{d}y\mathrm{d}x \\
&=\frac{3}{16}\int_0^2\int_{2-x}^2 xy^2\mathrm{d}y\mathrm{d}x \\
&=\frac{3}{16}\int_0^2\left(\frac{1}{3}xy^3\Big|_{2-x}^2\right)\mathrm{d}x \\
&=\frac{1}{16}\int_0^2(8x-x(2-x)^3)\mathrm{d}x \\
&=\frac{1}{16}\int_0^2(x^4+12x^2-6x^3)\mathrm{d}x \\
&=\frac{1}{16}\left(\frac{1}{5}x^5+4x^3-\frac{3}{2}x^4\Big|_0^2\right) \\
&=\frac{1}{16}\left(\frac{32}{5}+32-24\right) \\
&=\frac{2}{5}+\frac{8}{16}=\frac{9}{10}
\end{aligned}
$$

c. First, draw the area over which the integral is taken.



The conditional probability is given by

$$P(Y < X^2 | X + Y < 2) = \frac{P(X + Y < 2 \cap Y < X^2)}{P(X + Y < 2)}.$$

From part (b) we have $P(X + Y < 2) = 1 - P(X + Y > 2) = \frac{1}{10}$, which means the probability of falling into one of the two darkest areas equals $\frac{1}{10}$. The probability $P(X + Y < 2 \cap Y < X^2)$ is given by the integral over the darkest area in the plot.
Solution 1

$$\begin{aligned}
P(X + Y < 2 \cap Y < X^2) &= \int_0^1 \int_0^{x^2} f_{X,Y}(x, y)\mathrm{d}y\mathrm{d}x + \int_1^2 \int_0^{2-x} f_{X,Y}(x, y)\mathrm{d}y\mathrm{d}x \\
&= \frac{3}{16}\left[\int_0^1 \int_0^{x^2} xy^2\mathrm{d}y\mathrm{d}x + \int_1^2 \int_0^{2-x} xy^2\mathrm{d}y\mathrm{d}x\right] \\
&= \frac{1}{16}\left[\int_0^1 \left(xy^3\Big|_0^{x^2}\right)\mathrm{d}x + \int_1^2 \left(xy^3\Big|_0^{2-x}\right)\mathrm{d}x\right] \\
&= \frac{1}{16}\left[\int_0^1 x^7\mathrm{d}x + \int_1^2 x(2-x)^3\mathrm{d}x\right] \\
&= \frac{1}{16}\left[\int_0^1 x^7\mathrm{d}x + \int_1^2 (-x^4 + 6x^3 - 12x^2 + 8x)\mathrm{d}x\right] \\
&= \frac{1}{128}\left[x^8\Big|_0^1\right] + \frac{1}{16}\left(-\frac{1}{5}x^5 + \frac{3}{2}x^4 - 4x^3 + 4x^2\right)\Bigg|_1^2 \\
&= \frac{1}{128} + \frac{1}{16}\left(-\frac{32}{5} + 24 - 32 + 16 + \frac{1}{5} - \frac{3}{2} + 4 - 4\right) = \frac{17}{640}
\end{aligned}$$

Solution 2 (easier)

$$\begin{aligned}
P(X+Y<2\cap Y<X^2) &= \int_0^1 \int_{\sqrt{y}}^{2-y} f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y \\
&= \frac{3}{16}\left[\int_0^1 \int_{\sqrt{y}}^{2-y} xy^2\mathrm{d}x\mathrm{d}y\right] \\
&= \frac{3}{32}\left[\int_0^1 \left(x^2y^2\Big|_{\sqrt{y}}^{2-y}\right)\mathrm{d}y\right] \\
&= \frac{3}{32}\left[\int_0^1 (2-y)^2y^2 - y^3\mathrm{d}y\right] \\
&= \frac{3}{32}\left[\int_0^1 4y^2 - 5y^3 + y^4\mathrm{d}y\right] \\
&= \frac{3}{32}\left[\frac{4}{3}y^3 - \frac{5}{4}y^4 + \frac{1}{5}y^5\Big|_0^1\right] \\
&= \frac{3}{32}\left(\frac{4}{3} - \frac{5}{4} + \frac{1}{5}\right) = \frac{17}{640}
\end{aligned}$$

Using either Solution 1 or Solution 2, we get the final answer:

$$P(Y<X^2|X+Y<2) = \frac{P(X+Y<2\cap Y<X^2)}{P(X+Y<2)} = \frac{\frac{17}{640}}{\frac{1}{10}} = \frac{17}{64}.$$

7.7 MEMORYLESS EXCURSIONS: A CONFUSING PROBLEM WITH MEMORYLESS RVS

BH give a quick argument to compute $\mathsf{E}[M]$ and $\mathsf{E}[L]$ where $M = \max\{X, Y\}$ and $L = \min\{X, Y\}$ are the maximum and minimum of two iid exponential rvs X and Y. Since X and Y have the same distribution,

$$\mathsf{E}[L] + \mathsf{E}[M] = \mathsf{E}[L+M] = \mathsf{E}[X+Y] = 2\,\mathsf{E}[X].$$

Therefore,

$$\mathsf{E}[M] = 2\,\mathsf{E}[X] - \mathsf{E}[L]. \tag{7.7.1}$$

Next, by the fact that X and Y are memoryless,

$$\mathsf{E}[M] = \mathsf{E}[L] + \mathsf{E}[X]. \tag{7.7.2}$$

An interpretation can help to see this. There are two machines, each working on a job in parallel. Let X and Y be the production times at either machine. The time the first job finishes is evidently $L = \min\{X, Y\}$. Then, *due to memorylessness*, the service time of the remaining job 'restarts'; this takes an expected time $\mathsf{E}[X]$ to complete. Adding these two equations and noting that $\mathsf{E}[L]$ cancels we get $2\,\mathsf{E}[M] = 3\,\mathsf{E}[X]$, hence:

$$\mathsf{E}[M] = \frac{3}{2}\mathsf{E}[X], \qquad \mathsf{E}[L] = \mathsf{E}[M] - \mathsf{E}[X] = \frac{1}{2}\mathsf{E}[X]. \tag{7.7.3}$$

This argument seems general enough, so it must hold for discrete memoryless rvs too, i.e., when $X, Y \sim \text{Geo}(p)$. But that is not the case: it is only true when $X, Y \sim \text{Exp}(\lambda)$ and independent. To see what is wrong I tried as many different approaches to this problem I could think of, which resulted in this text.[1] In Section 7.7.1 we'll derive (7.7.1) for geometric rvs in multiple different ways. Hence, the culprit must be (7.7.2). Then, in Section 7.7.2 we'll show that both equations *are true* for exponential rvs. Finally, in Section 7.7.3 we find a formula that is similar to $\mathsf{E}[M] = \mathsf{E}[L] + \mathsf{E}[X]$ but that holds for both types of memoryless rvs, whether they are discrete or continuous.

THE ANALYSIS OF the above problem illustrates many general and useful probability concepts such as joint CDF, joint PMF/PDF, the fundamental bridge, integration over 2D areas, 2D LOTUS, conditional PMF/PDF, MGFs, and the change of variables formula. It pays off to do the exercises yourself and then study the hints and solutions carefully.

YOU'LL NOTICE, hopefully, that I use many different methods to the same problem, and that I take pains to see how the answers of these methods relate. There are at least two reasons for this. Often, a problem can be solved in multiple ways, and one method is not necessarily better than another; better yet, different methods may augment the understanding of the problem. The second reason is that it is easy to make a mistake in probability. If different methods give the same answer, the probability of having made a mistake becomes smaller.

[1] Part of this material was born out of annoyance. The book uses one of those typical probability arguments: slick, half complete, and wrong as soon as one tries it in other situations. In other words, the type of argument beginner books should stay clear of. I admit that I was quite irritated about the argument offered by the book.

7.7.1 *Discrete memoryless rvs*

Before embarking on a problem, it often helps to refresh our memory. This is what we do first. Let X be $\sim \text{Geo}(p)$ and write $q = 1 - p$.

Ex 7.38. What is the domain of X?

s.7.38. Of course $X \in \{0, 1, 2, \ldots\}$.

With some fun tricks with recursions it is possible to quickly derive the most important expressions for geometric rvs:

$$\begin{aligned}
\mathsf{P}\{X > 0\} &= \mathsf{P}\{\text{failure}\} = q \\
\mathsf{P}\{X > j\} &= q\,\mathsf{P}\{X > j-1\} \implies \mathsf{P}\{X > j\} = q^j\,\mathsf{P}\{X > 0\} = q^{j+1}. \\
\mathsf{P}\{X \geq j\} &= \mathsf{P}\{X > j-1\} = q^j. \\
\mathsf{P}\{X = j\} &= \mathsf{P}\{X > j-1\} - \mathsf{P}\{X > j\} = q^j - q^{j+1} = (1-q)q^j = pq^j. \\
\mathsf{E}[X] &= p \cdot 0 + q(1 + \mathsf{E}[X]) \implies \mathsf{E}[X] = q/(1-q) = q/p.
\end{aligned}$$

Mind that, even though this is neat, it only work for geometric rvs.

Ex 7.39. Explain the above.

s.7.39. For $X > 0$, the first outcome should be a failure. Then, for j failures, we need to fail $j-1$ times and then once more. For $\mathsf{E}[X]$, if there is a success, we don't need another another experiment. However, in case of a fail, we need another experiment, and we start again. Thus, $\mathsf{E}[X] = q(1 + \mathsf{E}[X]) \implies (1-q)\,\mathsf{E}[X] = q$.

Clearly, such tricks are nice and quick, but they are not general. We should also practice with the general method.

Ex 7.40. Simplify $\mathsf{P}\{X > j\} = \sum_{i=j+1}^{\infty} \mathsf{P}\{X = i\}$ to see that this is equal to q^{j+1}. Realize that from this, $\mathsf{P}\{X \geq j\} = \mathsf{P}\{X > j-1\} = q^j$.

s.7.40. With the regular method:

$$\begin{aligned}
\mathsf{P}\{X > j\} &= \sum_{i=j+1}^{\infty} \mathsf{P}\{X = i\} \\
&= p \sum_{i=j+1}^{\infty} q^i \\
&= p \sum_{i=0}^{\infty} q^{j+1+i} \\
&= pq^{j+1} \sum_{i=0}^{\infty} q^i \\
&= pq^{j+1} \frac{1}{1-q} = pq^{j+1} \frac{1}{p} = q^{j+1}.
\end{aligned}$$

Ex 7.41. Use indicator variables show that

$$\mathsf{E}[X] = \sum_{i=0}^{\infty} i\,\mathsf{P}\{X=i\} = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} I_{j<i}\,\mathsf{P}\{X=i\} = p/q.$$

s.7.41.

$$\begin{aligned}
\mathsf{E}[X] &= \sum_{i=0}^{\infty} i\,\mathsf{P}\{X=i\} \\
&= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} I_{j<i}\,\mathsf{P}\{X=i\} \\
&= \sum_{j=0}^{\infty}\sum_{i=0}^{\infty} I_{j<i}\,\mathsf{P}\{X=i\} \\
&= \sum_{j=0}^{\infty}\sum_{i=j+1}^{\infty} \mathsf{P}\{X=i\} \\
&= \sum_{j=0}^{\infty} \mathsf{P}\left\{X>j\right\} \\
&= \sum_{j=0}^{\infty} q^{j+1} \\
&= q\sum_{j=0}^{\infty} q^{j} \\
&= q/(1-q) = q/p.
\end{aligned}$$

Ex 7.42. Look up the definition of a memoryless rv, and check that X is memoryless.

s.7.42.

$$\begin{aligned}
\mathsf{P}\{X \geq n+m \,|\, X \geq m\} &= \frac{\mathsf{P}\{X \geq n+m, X \geq m\}}{\mathsf{P}\{X \geq m\}} \\
&= \frac{\mathsf{P}\{X \geq n+m\}}{\mathsf{P}\{X \geq m\}} \\
&= \frac{q^{n+m}}{q^m} \\
&= q^n = \mathsf{P}\{X \geq n\}.
\end{aligned}$$

WITH THIS REFRESHER, we can derive some useful properties of the minimum $L = \min\{X, Y\}$, where $Y \sim \text{Geo}(p)$ and independent of X. For this we use the fundamental bridge and 2D LOTUS, which in general read like

$$\mathsf{P}\left\{g(X,Y\right\} \in A\} = \mathsf{E}\left[I_{g(X,Y)\in A}\right] = \sum_i\sum_j I_{g(i,j)\in A}\,\mathsf{P}\left\{X=i, Y=j\right\}.$$

Ex 7.43. What is the domain of L? Then, show that

$$\mathsf{P}\{L \geq i\} = q^{2i} \implies L \sim \text{Geo}(1-q^2).$$

h.7.43. For 2D LOTUS, take $g(i,j) = \min\{i,j\}$.

s.7.43.

$$\begin{aligned}\mathsf{P}\{L \geq k\} &= \sum_i \sum_j I_{\min\{i,j\}\geq k} \mathsf{P}\left\{X = i, Y = j\right\} \\ &= \sum_{i\geq k} \sum_{j\geq k} \mathsf{P}\{X = i\} \mathsf{P}\left\{Y = j\right\} \\ &= \mathsf{P}\{X \geq k\} \mathsf{P}\{Y \geq k\} = q^k q^k = q^{2k}.\end{aligned}$$

$\mathsf{P}\{L > i\}$ has the same form as $\mathsf{P}\{X > i\}$, but now with q^{2i} rather than q^i.

Ex 7.44. Show that

$$\mathsf{E}[L] = q^2/(1-q^2).$$

h.7.44. $X \sim \text{Geo}(1-q) \implies \mathsf{E}[X] = q/(1-q)$. Now use that $L \sim \text{Geo}(1-q^2)$.

s.7.44. Immediate from the hint and [7.41].

Ex 7.45. Show that

$$\mathsf{E}[L] + \mathsf{E}[X] = \frac{q}{1-q}\frac{1+2q}{1+q}. \tag{7.7.4}$$

h.7.45. Use that $1-q^2 = (1-q)(1+q)$.

s.7.45.

$$\begin{aligned}\mathsf{E}[L] + \mathsf{E}[X] &= \frac{q^2}{1-q^2} + \frac{q}{1-q} \\ &= \frac{q}{1-q}\left(\frac{q}{1+q} + 1\right) \\ &= \frac{q}{1-q}\frac{1+2q}{1+q}\end{aligned}$$

NOW WE CAN combine these facts with the properties of the maximum $M = \max\{X, Y\}$.

Ex 7.46. Show that

$$2\mathsf{E}[X] - \mathsf{E}[L] = \frac{q}{1-q}\frac{2+q}{1+q}.$$

s.7.46.

$$\begin{aligned}
2\,\mathsf{E}[X] - \mathsf{E}[L] &= 2\frac{q}{1-q} - \frac{q^2}{1-q^2} \\
&= \frac{q}{1-q}\left(2 - \frac{q}{1+q}\right) \\
&= \frac{q}{1-q}\left(\frac{2+2q}{1+q} - \frac{q}{1+q}\right) \\
&= \frac{q}{1-q}\frac{2+q}{1+q}.
\end{aligned}$$

Clearly, unless $q = 0$, $\mathsf{E}[L] + \mathsf{E}[X] \neq 2\,\mathsf{E}[X] - \mathsf{E}[L]$, hence, $\mathsf{E}[M]$ can only be one of the two and (7.7.1) and (7.7.2) cannot be both true.

To convince ourselves that [7.46], hence (7.7.1), is indeed true, we pursue three ideas.

HERE IS THE FIRST idea.

Ex 7.47. Show for the PMF of M that

$$p_M(k) = \mathsf{P}\{M = k\} = 2pq^k(1-q^k) + p^2q^{2k}.$$

h.7.47. Use 2D LOTUS on $g(x, y) = I_{\max\{x,y\}=k}$.

s.7.47.

$$\begin{aligned}
\mathsf{P}\{M = k\} &= \mathsf{P}\{\max\{X, Y\} = k\} \\
&= p^2 \sum_{ij} I_{\max\{i,j\}=k} q^i q^j \\
&= 2p^2 \sum_{ij} I_{i=k} I_{j<k} q^i q^j + p^2 \sum_{ij} I_{i=j=k} q^i q^j \\
&= 2p^2 q^k \sum_{j<k} q^j + p^2 q^{2k} \\
&= 2p^2 q^k \frac{1-q^k}{1-q} + p^2 q^{2k}
\end{aligned}$$

Ex 7.48. With the previous exercise, show now that $p_M(k) = 2\,\mathsf{P}\{X = k\} - \mathsf{P}\{L = k\}$.

s.7.48. It's just algebra

$$\begin{aligned}
\mathsf{P}\{M = k\} &= 2p^2 q^k \frac{1-q^k}{1-q} + p^2 q^{2k} \\
&= 2pq^k(1-q^k) + p^2 q^{2k} \\
&= 2pq^k + (p^2 - 2p) q^{2k} \\
&= 2\,\mathsf{P}\{X = k\} - \mathsf{P}\{L = k\},
\end{aligned}$$

where I use that $p^2 - 2p = p(p-2) = (1-q)(1-q-2) = -(1-q)(1+q) = -(1-q^2)$.

Ex 7.49. Finally, show that $\mathsf{E}[M] = 2\mathsf{E}[X] - \mathsf{E}[L]$.

h.7.49.

s.7.49.

$$\begin{aligned}\mathsf{E}[M] &= \sum_k k\,\mathsf{P}\{M=k\}\\ &= \sum_k k(2\,\mathsf{P}\{X=k\} - \mathsf{P}\{L=k\})\\ &= 2\,\mathsf{E}[X] - \mathsf{E}[L].\end{aligned}$$

THE SECOND IDEA.

Ex 7.50. First show that $\mathsf{P}\{M \le k\} = (1-q^{k+1})^2$.

s.7.50.

$$\begin{aligned}\mathsf{P}\{M \le k\} &= \mathsf{P}\{X \le k, Y \le k\}\\ &= \mathsf{P}\{X \le k\}\,\mathsf{P}\{Y \le k\}\\ &= (1-\mathsf{P}\{X > k\})(1-\mathsf{P}\{Y > k\})\\ &= (1-q^{k+1})^2.\end{aligned}$$

Ex 7.51. Simplify $\mathsf{P}\{M=k\} = \mathsf{P}\{M \le k\} - \mathsf{P}\{M \le k-1\}$ to see that $p_M(k) = 2\,\mathsf{P}\{X=k\} - \mathsf{P}\{L=k\}$.

s.7.51.

$$\begin{aligned}\mathsf{P}\{M=k\} &= \mathsf{P}\{M \le k\} - \mathsf{P}\{M \le k-1\}\\ &= 1-2q^{k+1}+q^{2k+2} - (1-2q^k+q^{2k})\\ &= 2q^k(1-q) + q^{2k}(q^2-1)\\ &= 2\,\mathsf{P}\{X=k\} - q^{2k}(1-q^2)\\ &= 2\,\mathsf{P}\{X=k\} - \mathsf{P}\{L=k\}.\end{aligned}$$

AND HERE IS the third idea.

Ex 7.52. Explain that

$$\mathsf{P}\{L=i, M=k\} = 2p^2q^{i+k}\,I_{k>i} + p^2q^{2i}\,I_{i=k}).$$

s.7.52.

$$\begin{aligned}\mathsf{P}\{L=i, M=k\} &= 2\,\mathsf{P}\{X=i, Y=k\}\,I_{k>i} + \mathsf{P}\{X=Y=i\}\,I_{i=k}\\ &= 2p^2q^{i+k}\,I_{k>i} + p^2q^{2i}\,I_{i=k}.\end{aligned}$$

Ex 7.53. Use [7.52] and marginalization to compute the marginal PMF $\mathsf{P}\{M=k\}$.

h.7.53. Marginalize out L.

s.7.53.

$$\begin{aligned}
\mathsf{P}\{M=k\} &= \sum_i \mathsf{P}\{L=i, M=k\} \\
&= \sum_i (2p^2 q^{i+k} I_{k>i} + p^2 q^{2i} I_{i=k}) \\
&= 2p^2 q^k \sum_{i=0}^{k-1} q^i + p^2 q^{2k} \\
&= 2pq^k(1-q^k) + p^2 q^{2k} \\
&= 2pq^k + (p^2 - 2p) q^{2k},
\end{aligned}$$

Ex 7.54. Use [7.52] to compute $\mathsf{P}\{L=i\}$.

h.7.54. Marginalize out M.

s.7.54.

$$\begin{aligned}
\mathsf{P}\{L=i\} &= \sum_k \mathsf{P}\{L=i, M=k\} \\
&= \sum_k (2p^2 q^{i+k} I_{k>i} + p^2 q^{2i} I_{i=k}) \\
&= 2p^2 q^i \sum_{k=i+1}^{\infty} q^k + p^2 q^{2i} \\
&= 2p^2 q^{2i+1} \sum_{k=0}^{\infty} q^k + p^2 q^{2i} \\
&= 2pq^{2i+1} + p^2 q^{2i} \\
&= pq^{2i}(2q+p) \\
&= (1-q)q^{2i}(q+1), \quad p+q=1, \\
&= (1-q^2)q^{2i}.
\end{aligned}$$

IN CONCLUSION, we verified the correctness of $\mathsf{E}[M] = 2\,\mathsf{E}[X] - \mathsf{E}[L]$ in three different, and useful ways. Let us now focus on exponential rvs rather than geometric rvs.

7.7.2 *Continuous memoryless rvs*

In this section we analyze the correctness of (7.7.1) and (7.7.2) for continuous memoryless rvs, i.e., exponentially distributed rvs. I decided to analyze this in as much detail as I could think of, hoping that this would provide me with a lead to see how to generalize the equation $\mathsf{E}[M] = \mathsf{E}[L] + \mathsf{E}[X]$ such that it covers also the case with geometric rvs.

FIRST WE NEED to recall some basic facts about the exponential distribution.

Ex 7.55. Show that X is memoryless.

s.7.55. With $t \geq s$,

$$\mathsf{P}\{X > t | X > s\} = \frac{\mathsf{P}\{X > t\}}{\mathsf{P}\{X > s\}} = e^{-\lambda t} e^{\lambda s} = e^{-\lambda(t-s)} = \mathsf{P}\{X > t - s\}.$$

Ex 7.56. Show that $\mathsf{E}[X] = 1/\lambda$.

h.7.56. Use partial integration.

s.7.56. It is essential that you know both methods to solve this integral.

$$\begin{aligned}
\mathsf{E}[X] &= \int_0^\infty x f_X(x) \, \mathrm{d}x \\
&= \int_0^\infty x\lambda e^{-\lambda x} \, \mathrm{d}x \\
&= -x e^{-\lambda x}\Big|_0^\infty + \int_0^\infty e^{-\lambda x} \, \mathrm{d}x \\
&= -\frac{1}{\lambda} e^{-\lambda x}\Big|_0^\infty = \frac{1}{\lambda}.
\end{aligned}$$

Substitution is also a very important technique to solve such integrals. Here we go again:

$$\begin{aligned}
\mathsf{E}[X] &= \int_0^\infty x f_X(x) \, \mathrm{d}x \\
&= \int_0^\infty x\lambda e^{-\lambda x} \, \mathrm{d}x \\
&= \frac{1}{\lambda} \int_0^\infty u e^{-u} \, \mathrm{d}u,
\end{aligned}$$

by the substitution $u = \lambda x \implies \mathrm{d}u = \mathrm{d}(\lambda x) \implies \mathrm{d}u = \lambda \, \mathrm{d}x \implies \mathrm{d}x = \mathrm{d}u/\lambda$. With partial integration (do it!), the integral evaluates to 1.

NOW WE CAN shift our attention to the rvs L and M.

Ex 7.57. Show that $F_L(x) = 1 - e^{-2\lambda x}$.

h.7.57. Using independence and the specific property of the r.v. L that $\{L > x\} \iff \{X > x, Y > x\}$:

s.7.57. With the hint,

$$G_L(x) = \mathsf{P}\{L > x\} = \mathsf{P}\{X > x, Y > x\} = G_X(x)^2 = e^{-2\lambda x}.$$

The result follows since $F_L(x) = 1 - G_L(x)$.

Clearly, this implies that $L \sim \text{Exp}(2\lambda)$ and $\mathsf{E}[L] = 1/(2\lambda) = \mathsf{E}[X]/2$. Hence, we see that (7.7.3) holds now. Moreover, with the same trick we see that the distribution function F_M for the maximum M is given by

$$F_M(v) = \mathsf{P}\{M \leq v\} = \mathsf{P}\{X \leq v, Y \leq v\} = (F_X(v))^2 = (1 - e^{-\lambda v})^2.$$

Of course this is a nice trick, but it is not a method that allows us to compute the distribution for more general functions of X and Y. For more general cases, we have to use the fundamental bridge and LOTUS, that is, for any set[2] A in the domain of $X \times Y$

$$\begin{aligned}\mathsf{P}\{g(X,Y) \in A\} &= \mathsf{E}\left[I_{g(X,Y)\in A}\right] = \iint I_{g(x,y)\in A} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \iint_{g^{-1}(A)} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y = \iint_{\{(x,y):g(x,y)\in A\}} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y.\end{aligned}$$

The joint CDF $F_{X,Y}$ then follows because $F_{X,Y}(x,y) = \mathsf{P}\{X \leq x, Y \leq y\} = \mathsf{E}\left[I_{X\leq x, Y\leq y}\right]$. A warning is in place: conceptually this approach is easy, but doing the integration can be very challenging (or impossible). However, this expression is very important as this is the preferred way to compute distributions by numerical methods and simulation.

Ex 7.58. Use the fundamental bridge to re-derive the above expression for $F_M(v)$.

s.7.58.

$$\begin{aligned}\mathsf{P}\{M \leq v\} &= \mathsf{E}[I_{M\leq v}] \\ &= \int_0^\infty \int_0^\infty I_{x\leq v, y\leq v} f_{XY}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_0^\infty \int_0^\infty I_{x\leq v, y\leq v} f_X(x) f_Y(y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_0^v f_X(x)\,\mathrm{d}x \int_0^v f_Y(y)\,\mathrm{d}y \\ &= F_X(v)F_Y(v) = (F_X(v))^2.\end{aligned}$$

Ex 7.59. Show that the density of M has the form $f_M(v) = \partial_v F_M(v) = 2(1-e^{-\lambda v})\lambda e^{-\lambda v}$.[3]

s.7.59.

$$f_M(v) = F_M'(v) = 2F_X(v) f_X(v) = 2(1-e^{-\lambda v})\lambda e^{-\lambda v}.$$

Ex 7.60. Use the density f_M to show that $\mathsf{E}[M] = 2\mathsf{E}[X] - \mathsf{E}[L]$.

[2] If you like maths, you should be quite a bit more careful about what type of set A is acceptable. Here such matters are of no importance.

[3] We write ∂_v as a shorthand for $\mathrm{d}/\mathrm{d}v$ in the 1D case, and for ∂/∂_v the partial derivative in the 2D case.

s.7.60.

$$\begin{aligned}
\mathsf{E}[M] &= \int_0^\infty v f_M(v)\,\mathrm{d}v = \\
&= 2\lambda \int_0^\infty v(1-e^{-\lambda v})e^{-\lambda v}\,\mathrm{d}v = \\
&= 2\lambda \int_0^\infty v e^{-\lambda v}\,\mathrm{d}v - 2\lambda \int_0^\infty v e^{-2\lambda v}\,\mathrm{d}v \\
&= 2\,\mathsf{E}[X] - \mathsf{E}[L],
\end{aligned}$$

where the last equality follows from the previous exercises.

Recalling that we already obtained $\mathsf{E}[L] = \mathsf{E}[X]/2$, we see that $\mathsf{E}[M] = 2\,\mathsf{E}[X] - \mathsf{E}[L] = 3\,\mathsf{E}[X]/2$, which settles the truth of (7.7.3).

WE CAN ALSO compute the densities $f_M(y)$ (and $f_L(x)$ by marginalizing the joint density $f_{L,M}(x,y)$. However, for this, we first need the joint distribution $F_{L,M}$, and then we can get $f_{L,M}$ by differentiation, i.e., $f_{X,Y} = \partial_x\partial_y F_{X,Y}$. Let us try this approach too.

Ex 7.61. Use the fundamental bridge to show that for $u \leq v$,

$$F_{L,M}(u,v) = \mathsf{P}\{L \leq u, M \leq v\} = 2\int_0^u (F_Y(v) - F_Y(x))f_X(x)\,\mathrm{d}x.$$

s.7.61. First the joint distribution. With $u \leq v$,

$$\begin{aligned}
F_{L,M}(u,v) &= \mathsf{P}\{L \leq u, M \leq v\} \\
&= 2\iint I_{x\leq u, y\leq v, x\leq y} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= 2\int_0^u \int_x^v f_Y(y)\,\mathrm{d}y\, f_X(x)\,\mathrm{d}x && \text{independence} \\
&= 2\int_0^u (F_Y(v) - F_Y(x))f_X(x)\,\mathrm{d}x.
\end{aligned}$$

Ex 7.62. Take partial derivatives to show that

$$f_{L,M}(u,v) = 2 f_X(u) f_Y(v)\, I_{u\leq v}.$$

s.7.62. Taking partial derivatives,

$$\begin{aligned}
f_{L,M}(u,v) &= \partial_v\partial_u F_{L,M}(u,v) \\
&= 2\partial_v\partial_u \int_0^u (F_Y(v) - F_Y(x))f_X(x)\,\mathrm{d}x \\
&= 2\partial_v \left\{(F_Y(v) - F_Y(u))f_X(u)\right\} \\
&= 2 f_X(u)\partial_v F_Y(v) \\
&= 2 f_X(u) f_Y(v).
\end{aligned}$$

Ex 7.63. In [7.62] marginalize out L to find f_M, and marginalize out M to find f_L.

s.7.63.

$$\begin{aligned}
f_M(v) &= \int_0^\infty f_{L,M}(u,v)\,\mathrm{d}u \\
&= 2\int_0^v f_X(u) f_Y(v)\,\mathrm{d}u \\
&= 2 f_Y(v)\int_0^v f_X(u)\,\mathrm{d}u \\
&= 2 f_Y(v) F_X(v), \\
f_L(u) &= \int_0^\infty f_{L,M}(u,v)\,\mathrm{d}v \\
&= 2 f_X(u)\int_u^\infty f_Y(v)\,\mathrm{d}v \\
&= 2 f_X(u) G_Y(u).
\end{aligned}$$

WE DID A number of checks for the case X, Y, iid, $\sim \text{Exp}(\lambda)$, but I have a second way to check the consistency of our results. For this I use the idea that the geometric distribution is the discrete analog of the exponential distribution. Now we study how this works, and that by taking proper limits we can obtain the results for the continuous setting from the discrete setting.

First, let's try to obtain an intuitive understanding of how $X \sim \text{Geo}(\lambda/n)$ approaches $Y \sim \text{Exp}(\lambda)$ as $n \to \infty$. For this, divide the interval $[0,\infty)$ into many small intervals of length $1/n$. Let $X \sim \text{Geo}(\lambda/n)$ for some $\lambda > 0$ and $n \gg 0$. Then take some $x \geq 0$ and let i be such that $x \in [i/n, (i+1)/n)$.

Ex 7.64. Show that

$$\mathsf{P}\{X/n \approx x\} \approx \frac{\lambda}{n}\left(1-\frac{\lambda}{n}\right)^{Zion}. \tag{7.7.5}$$

s.7.64. First,

$$\begin{aligned}
\mathsf{P}\{X/n \approx x\} &= \mathsf{P}\{X/n \in [i/n,(i+1)/n]\} = \mathsf{P}\{X \in [i,i+1]\} = pq^i \\
&\approx pq^{nx} = \frac{\lambda}{n}\left(1-\frac{\lambda}{n}\right)^{xn},
\end{aligned}$$

since $p = \lambda/n$, $q = 1-p = 1-\lambda/n$, and $i = nx$.

Next, introduce the highly intuitive notation[4] $\mathrm{d}x = \lim_{n\to\infty} 1/n$, and use the standard limit[5] $(1-\lambda/n)^n \to e^{-\lambda}$ as $n \to \infty$ to see that (7.7.5) converges to

$$\mathsf{P}\{X/n \approx x\} \to \lambda e^{-\lambda x}\,\mathrm{d}x = f_X(x)\,\mathrm{d}x, \quad \text{as } n \to \infty.$$

[4] In your math classes you learned that $\lim_{n\to\infty} 1/n = 0$. Doesn't this definition therefore imply that $\mathrm{d}x = 0$? Well, no, because $\mathrm{d}x$ is not a real number but an infinitesimal. Infinitesimals allow us to consider a quantity that is so small that it cannot be distinguished from 0 within the real numbers.

[5] This is not entirely trivial to prove. If you like mathematics, check the neat proof in Rudin's Principles of mathematical analysis.

If you don't like this trick with dx, here is another method, based on with moment-generating functions.

Ex 7.65. Derive the moment-generating function $M_{X/n}(s)$ of X/n when $X \sim \text{Geo}(p)$. Then, let $p = \lambda/n$, and show that $\lim_{n\to\infty} M_{X/n}(s) = M_Y(s)$, where $Y \sim \text{Exp}(\lambda)$.

h.7.65. If you recall the Poisson distribution, you know that $e^\lambda = \sum_{i=0}^\infty \lambda^i/i!$. In fact, this is precisely Taylor's expansion of e^λ.

s.7.65.

$$M_{X/n}(s) = \mathsf{E}\left[e^{sX/n}\right] = \sum_i e^{si/n} p q^i = p \sum_i (q e^{s/n})^i = \frac{p}{1 - q e^{s/n}}.$$

With $p = \lambda/n$ this becomes

$$\begin{aligned} M_{X/n}(s) &= \frac{\lambda/n}{1 - (1-\lambda/n)(1 + s/n + 1/n^2 \times (\cdots))} \\ &= \frac{\lambda/n}{\lambda/n - s/n + 1/n^2 \times (\cdots)} \\ &= \frac{\lambda}{\lambda - s + 1/n \times (\cdots)} \\ &\to \frac{\lambda}{\lambda - s}, \quad \text{as } n \to \infty, \end{aligned}$$

where we write $1/n^2 \times (\cdots)$ for all terms that will disappear when we take the limit $n \to \infty$. This is just handy notation to hide details in which we are not interested.

With these limits in place, we can relate the minimum $L = \min\{X, Y\}$ for the discrete and the continuous settings.

Ex 7.66. Suppose that $X, Y \sim \text{Geo}(\lambda/n)$, then check that $\lim_{n\to\infty} \mathsf{E}[L/n] = 1/2\lambda$.

h.7.66. Use [7.44].

s.7.66.

$$\begin{aligned} \mathsf{E}[L/n] &= \frac{1}{n}\mathsf{E}[L] = \frac{1}{n}\frac{q^2}{1-q^2} \\ &= \frac{1}{n}\frac{(1-\lambda/n)^2}{1-(1-\lambda/n)^2} \\ &= \frac{1}{n}\frac{1 - 2\lambda/n + (\lambda/n)^2}{2\lambda/n + (\lambda/n)^2} \\ &= \frac{1 - 2\lambda/n + (\lambda/n)^2}{2\lambda + \lambda^2/n} \\ &\to \frac{1}{2\lambda}. \end{aligned}$$

Clearly, $1/2\lambda = \mathsf{E}[X]/2$ when $X \sim \text{Exp}(\lambda)$.

Here is yet another check on the correctness of $f_M(x)$.

Ex 7.67. Show that the PMF $\mathsf{P}\{M = k\}$ for the discrete M in [7.47] converges to $f_M(x)$ of [7.59] when $n \to \infty$. Take k suitable.

s.7.67. Take $p = \lambda/n$, $q = 1 - \lambda/n$, and $k \approx xn$, hence $k/n \approx x$. Then,

$$\begin{aligned}
\mathsf{P}\{M/n = k/n\} &= 2pq^{k/n}(1 - q^{k/n}) + p^2 q^{2k/n} \\
&= 2\frac{\lambda}{n}\left(1 - \frac{\lambda}{n}\right)^{k/n}\left(1 - \left(1 - \frac{\lambda}{n}\right)^{k/n}\right) + \frac{\lambda^2}{n^2}\left(1 - \frac{\lambda}{n}\right)^{2k/n} \\
&\to 2\lambda\,\mathrm{d}x e^{-\lambda x}\left(1 - e^{-\lambda x}\right) + \lambda^2\,\mathrm{d}x^2 e^{-2\lambda}.
\end{aligned}$$

Now observe that the second term, proportional to $\mathrm{d}x^2$ can be neglected.

FINALLY, I HAVE a third way to check the above results, namely by verifying (7.7.2), i.e. $\mathsf{E}[M - L] = \mathsf{E}[X]$. For this, we compute the joint CDF $f_{L,M-L}(x, y)$. With this, you'll see directly how to compute $\mathsf{E}[M - L]$.

Ex 7.68. Use the fundamental bridge to obtain

$$F_{L,M-L}(x, y) = (1 - e^{-2\lambda x})(1 - e^{-\lambda y}) = F_L(x)F_Y(y).$$

h.7.68. The idea is not difficult, but the technical details require attention, in particular the limits in the integrations.

s.7.68.

$$
\begin{aligned}
F_{L,M-L}(x,y) &= \mathsf{P}\{L \le x, M-L \le y\} \\
&= 2\,\mathsf{P}\{X \le x, Y-X \le y, X \le Y\} \\
&= 2\int_0^\infty \int_0^\infty I_{u\le x, v-u\le y, u\le v} f_{X,Y}(u,v)\,\mathrm{d}u\,\mathrm{d}v \\
&= 2\int_0^\infty \int_0^\infty I_{u\le x, v-u\le y, u\le v} \lambda^2 e^{-\lambda u} e^{-\lambda v}\,\mathrm{d}u\,\mathrm{d}v \\
&= 2\int_0^x \int_0^\infty I_{u\le v\le u+y} \lambda^2 e^{-\lambda u} e^{-\lambda v}\,\mathrm{d}v\,\mathrm{d}u \\
&= 2\int_0^x \lambda e^{-\lambda u} \int_u^{u+y} \lambda e^{-\lambda v}\,\mathrm{d}v\,\mathrm{d}u \\
&= 2\int_0^x \lambda e^{-\lambda u} (-e^{-\lambda v})\Big|_u^{u+y}\,\mathrm{d}u \\
&= 2\int_0^x \lambda e^{-\lambda u} (e^{-\lambda u} - e^{-\lambda(u+y)})\,\mathrm{d}u \\
&= 2\lambda\int_0^x e^{-2\lambda u}\,\mathrm{d}u - 2\lambda \int_0^x e^{-\lambda(2u+y)}\,\mathrm{d}u \\
&= 2\lambda\int_0^x e^{-2\lambda u}\,\mathrm{d}u - 2\lambda e^{-\lambda y} \int_0^x e^{-2\lambda u}\,\mathrm{d}u \\
&= (1-e^{-\lambda y}) 2\lambda \int_0^x e^{-2\lambda u}\,\mathrm{d}u \\
&= (1-e^{-\lambda y})(-e^{-2\lambda u})\Big|_0^x \\
&= (1-e^{-\lambda y})(1-e^{-2\lambda x}).
\end{aligned}
$$

Ex 7.69. Conclude that $M-L$ and L are independent, and $M-L \sim Y$.

s.7.69. As $F_{L,M-L}(x,y) = F_Y(y)F_L(x)$. So the CDF factors as a function of x only and a function of y only. This implies that L and $M-L$ are independent, and moreover that $F_{M-L}(y) = F_Y(y)$, so $M-L \sim Y$. We can also see this from the joint PDF:

$$f_{L,M-L}(x,y) = \partial_x \partial_y (F_Y(y)F_L(x)) = f_Y(y) f_L(x),$$

so the joint PDF (of course) also factors. The independence now follows from BH 7.1.21. Because L and $M-L$ are independent, the conditional density equals the marginal density:

$$f_{M-L|L}(y|x) = \frac{f_{L,M-L}(x,y)}{f_L(x)} = \frac{f_Y(y) f_L(x)}{f_L(x)} = f_Y(y).$$

By the above exercise, we find that $\mathsf{E}[M-L] = \mathsf{E}[Y] = \mathsf{E}[X]$, as X and Y are iid.

THIS MAKES ME wonder whether $M-L$ and L are also independent for the discrete case, i.e., when X, Y iid and $\sim \text{Geo}(p)$. Hence, we should check that for *all* i, j

$$\mathsf{P}\{L=i, M-L=j\} = \mathsf{P}\{L=i\}\,\mathsf{P}\{M-L=j\}. \tag{7.7.6}$$

Ex 7.70. Use [7.52] to see that

$$\mathsf{P}\{L = i, M - L = j\} = 2p^2 q^{2i+j} I_{j\geq 1} + p^2 q^{2i} I_{j=0}.$$

h.7.70. Realize that $\mathsf{P}\{L = i, M - L = j\} = \mathsf{P}\{L = i, M = i + j\}$. Now fill in the formula of [7.52].

Now for the RHS.

Ex 7.71. Derive that

$$\mathsf{P}\{M - L = j\} = \frac{2p^2 q^j}{1-q^2} I_{j\geq 1} + \frac{p^2 q^j}{1-q^2} I_{j=0}.$$

s.7.71. Suppose $j > 0$ (for $j = 0$ the maths is the same). Then,

$$\mathsf{P}\{M - L = j\} = 2\sum_{i=0}^{\infty} \mathsf{P}\{X = i, Y = i + j\} = \quad = 2\sum_{i=0}^{\infty} pq^i pq^{i+j} = = 2p^2 q^j \sum_{i=0}^{\infty} q^{2i} = \quad = 2p^2 q^j/(1-q^2).$$

Recalling that $\mathsf{P}\{L = i\} = (1-q^2)q^{2i}$, it follows right away from (7.7.6) that L and $M - L$ are independent. Interestingly, from [7.68] we see $M - L \sim Y$ for the continuous case. However, here, for the geometric case, $\mathsf{P}\{M - L = j\} \neq pq^j = \mathsf{P}\{Y = j\}$. This explains why $\mathsf{E}[M] \neq \mathsf{E}[L] + \mathsf{E}[X]$ for geometric rvs: we should be more careful in how to split M in terms of L and X.

ALL IN ALL, we have checked and double checked all our expressions and limits for the geometric and exponential distribution. We had success too: the solution of the last exercise provides the key to understand why (7.7.1) and (7.7.2) are true for exponentially distributed rvs, but not for geometric random variables. In fact, in the solutions we can see the term corresponding to $X = Y = i$ for $X, Y \sim \text{Geo}(p)$ becomes negligibly small when $n \to 0$. In other words, $\mathsf{P}\{X = Y\} > 0$ when X and Y are discrete, but $\mathsf{P}\{X = Y\} = 0$ when X and Y are continuous. Moreover, by [7.68], $\mathsf{E}[M] = \mathsf{E}[L + M - L] = \mathsf{E}[L] + \mathsf{E}[M - L]$, but $\mathsf{E}[M - L] \neq \mathsf{E}[X]$. So, to resolve our leading problem we should reconsider $\mathsf{E}[M - L]$.

7.7.3 *The solution*

Let us now try to repair (7.7.2), i.e., $\mathsf{E}[M] = \mathsf{E}[L] + \mathsf{E}[X]$, for the case $X, Y \sim \text{Geo}(p)$. We should be careful about the non-negligible case that $M = L$, so we move, carefully, step by step.

Ex 7.72. Why is the following true:

$$\mathsf{E}[M] = \mathsf{E}[L] + \mathsf{E}[(M - L) I_{M>L}] = \mathsf{E}[L] + 2\,\mathsf{E}[(Y - X) I_{Y>X}]. \tag{7.7.7}$$

s.7.72. Because either $M = L$ or $M > L$. Recall from earlier work that the factor 2 in the second equality follows from the fact that X, Y iid.

Ex 7.73. Show that

$$2\,\mathsf{E}\left[(Y-X)\,I_{Y>X}\right] = \frac{2q}{1-q^2}.$$

h.7.73. Reread BH.7.2.2. to realize that $\mathsf{E}[M-L] = \mathsf{E}[|X-Y|]$. Relate the latter expectation to the expression in the problem.

s.7.73.

$$\begin{aligned}
\mathsf{E}\left[(Y-X)\,I_{Y>X}\right] &= p^2\sum_{ij}(j-i)\,I_{j>i}q^i q^j \\
&= p^2\sum_i q^i \sum_{j=i+1}^{\infty}(j-i)q^j \\
&= p^2\sum_i q^i q^i \sum_{k=1}^{\infty} kq^k, \quad k=j-i \\
&= p\sum_i q^{2i}\,\mathsf{E}[X] \\
&= \frac{p}{1-q^2}\,\mathsf{E}[X] \\
&= \frac{p}{1-q^2}\frac{q}{p} \\
&= \frac{q}{1-q^2}.
\end{aligned}$$

Ex 7.74. Combine the above with the expression for $\mathsf{E}[L]$ of [7.44] to obtain [7.46] for $\mathsf{E}[M] = 2\,\mathsf{E}[X] - \mathsf{E}[L]$, thereby verifying the correctness of (7.7.7).

h.7.74.

s.7.74.

$$\mathsf{E}[L] + 2\,\mathsf{E}\left[(Y-X)\,I_{Y>X}\right] = \frac{q^2}{1-q^2} + \frac{2q}{1-q^2} = \frac{q}{1-q}\frac{q+2}{1+q},$$

where I use that $1-q^2 = (1-q)(1+q)$.

While (7.7.7) is correct, I am still not happy with the second part of (7.7.7) as I find it hard/unintuitive to interpret. Restarting again from scratch, here is another attempt to rewrite $\mathsf{E}[M]$ by using $Z \sim \mathrm{FS}(p)$, i.e., Z has the first success distribution with parameter p, in other words, $Z \sim X+1$ with $X \sim \mathrm{Geo}(p)$.

Ex 7.75. Explain that

$$\mathsf{E}[M] = \mathsf{E}[L] + \mathsf{E}\left[Z\,I_{M>L}\right], \tag{7.7.8}$$

s.7.75. To see why this might be true, I reason like this. After 'seeing' L, we want to restart. Let Z be the time from the restart to M. When $Z \sim \text{Geo}(p)$, it might happen that $Z = 0$ (with positive probability p). But if $Z = 0$, then $M = L$, and it that case, we should not restart. Hence, if $Z \sim \text{Geo}(p)$ we are 'double counting' when $Z = 0$. By including the condition $M > L$ and by taking $Z \sim \text{FS}(p)$ (so that $Z > 0$) I can prevent this.

Ex 7.76. Show that

$$\mathsf{E}[Z I_{M>L}] = \frac{2q}{1-q^2},$$

i.e., the same as [7.73], hence (7.7.8) is correct.

h.7.76. Observe that Z is independent from X and Y, hence from M and L

s.7.76. With the hint:

$$\mathsf{E}[Z I_{M>L}] = \mathsf{E}[Z]\,\mathsf{E}[I_{M>L}] = \mathsf{E}[Z]\,\mathsf{P}\{M > L\} = \frac{1}{p}\frac{2pq}{1-q^2} = \frac{2q}{1-q^2},$$

We know that $\mathsf{E}[Z] = 1 + \mathsf{E}[X] = 1 + q/p = 1/p$, while

$$\begin{aligned}\mathsf{P}\{M > L\} &= 1 - \mathsf{P}\{X = Y\} = 1 - \sum_{i=0}^{\infty} \mathsf{P}\{X = Y = i\} \\ &= 1 - \frac{p^2}{1-q^2} = \frac{1-q^2+p^2}{1-q^2} = \frac{2pq}{1-q^2}.\end{aligned}$$

I AM NEARLY happy, but I want to see that (7.7.8), which is correct for discrete rvs, has also the correct limiting behavior.

Ex 7.77. Show that $\mathsf{E}[Z/n\, I_{M>L}] \to 1/\lambda$, which is the expectation of an $\text{Exp}(\lambda)$ rv!

s.7.77. By independence,

$$\mathsf{E}[Z/n\, I_{M>L}] = \mathsf{E}[Z/n]\,\mathsf{P}\{M > L\}.$$

Then,

$$\mathsf{P}\{M > L\} = \frac{2pq}{1-q^2} = \frac{2\lambda/n(1-\lambda/n)}{1-(1-\lambda/n)^2} = \frac{2\lambda/n(1-\lambda/n)}{2\lambda/n - \lambda^2/n^2} = \frac{2(1-\lambda/n)}{2-\lambda/n} \to 1,$$

and

$$\mathsf{E}[Z/n] = \frac{1}{n}\mathsf{E}[Z] = \frac{1}{n}\frac{1}{p} = \frac{1}{n\lambda/n} = 1/\lambda.$$

Finally I understand why $\mathsf{E}[M] = \mathsf{E}[L] + \mathsf{E}[X]$ for $X, Y \sim \text{Exp}(\lambda)$ but not for when X, Y are discrete. For discrete rvs, L and M can be equal, while for continuous rvs, this is impossible[6] It took a long time, and a lot of work, to understand how to resolve the confusing problem, but I learned a lot. In particular, I find (7.7.8) a nice and revealing equation.

Finally, if you like to train with the tools you learned, you can try your hand at analyzing the same problem, but now for uniform X, Y.

[6] A bit more carefully formulated: the event $\{L = M\}$ has zero probability for continuous rvs.

8 CHAPTER 8: EXERCISES AND REMARKS

8.1 SECTION 8.1

Ex 8.1. In probability theory we often want to study properties of functions of rvs. Provide an example for such a function.

s.8.1. Recall that $\mathsf{V}[X] = \mathsf{E}\left[X^2\right] - (\mathsf{E}[X])^2$; so we have to deal with the function $g(x) = x^2$ because $\mathsf{E}\left[X^2\right] = \mathsf{E}\left[g(X)\right]$. Note that even to properly define the variance, we have to deal with a function that is not one-to-one everywhere on $\mathbb{R}$.

Ex 8.2. Let the rv X be uniform on the set $\{0,1,2,3,4,5\}$. Derive the PMF and the CDF of $Z = 3X$. Explicitly specify the domain.

s.8.2.

$$X \in \{0,\dots,5\} \implies Z \in \{0,3,6,9,12,15\}, \quad \text{and not in } \{0,1,2,\dots,14,15\}, \tag{8.1.1}$$

$$z = g(x) = 3x, \tag{8.1.2}$$

$$p_Z(z) = \sum_{x:g(x)=z} p_X(x) = \frac{1}{6} I_{z\in\{0,3,6,9,12,15\}}, \tag{8.1.3}$$

$$F_Z(z) = \frac{1}{6}\sum_{x=0}^{z} I_{x\in\{0,3,6,9,12,15\}}. \tag{8.1.4}$$

Ex 8.3. Suppose $y = g(x)$ for some differentiable function g. We like to express the PDF f_Y for $Y = g(X)$ in terms of the PDF f_X and g. This is easy when g is strictly increasing and has an inverse at y, because

$$F_Y(y) = \mathsf{P}\left\{Y \leq y\right\} = \mathsf{P}\left\{g(X) \leq y\right\} = \mathsf{P}\left\{X \leq g^{-1}(y)\right\} = F_X(g^{-1}(y)). \tag{8.1.5}$$

Now we take the derivative at the LHS and RHS to get with the chain rule

$$f_Y(y) = F'_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_X(g^{-1}(y)) = f_X(g^{-1}(y)) \frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y) = f_X(x)\frac{1}{g'(x)},$$

where we write $x = g^{-1}(y)$ in the last step. But why is the derivative of $g^{-1}(y)$ at y equal to $1/g'(x)$, with $x = g^{-1}(y)$?

s.8.3. To get the derivative of g^{-1}, consider the equality $g(g^{-1}(y)) = y$. Then, taking derivatives with respect to y at both sides, and applying the chain rule,

$$g(g^{-1}(y)) = y \implies \frac{\mathrm{d}}{\mathrm{d}y} g(g^{-1}(y)) = 1 \iff g'(g^{-1}(y))\frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y) = 1 \implies \frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y) = 1/g'(x),$$

where we use that $g^{-1}(y) = x$. Notice why g is assumed increasing: now we know that $g'(x) \neq 0$.

Ex 8.4. When g is not strictly increasing everywhere, there can be no or multiple points x such that $g(x) = y$. Explain that in such cases it is much more difficult to express F_Y in terms of F_X than directly use the densities (assuming that g is differentiable). Extend your reasoning to 2D.

s.8.4. When working the CDFs, we need to solve the problem $\{x : g(x) \leq y\}$. If we take $g(x) = \sin x + x/100$ then this is really messy. In fact, to solve this, we first solve for the set $x : g(x) = y$, which might still be hard, but requires less work than check each and every interval.

With PDFs we only have to require *locally* that g is one-to-one, and we don't have to work with inequalities, but can direcly focus on the set $\{x : g(x) = y\}$.

In 2D, functions can have saddle points, i.e., points in which the function increases in one direction and decreases in another. Then finding the set of points x such that[1] $g(x, y) \leq (u, v)$ (which we need if we want to express $\mathsf{P}\left\{g(X, Y) \leq (u, v)\right\}$ in terms of the distribution $F_{X,Y}$) is not a particularly attractive task, to say the least.

See also the inverse function theorem, which will be covered in more detail next block.

Ex 8.5. The general 1D change of variables formula is like this,

$$f_Y(y) = \sum_{x_i : g(x_i) = y} f_X(x_i) \frac{1}{|g'(x_i)|},$$

with some natural conditions on g. Apply this formula to the case $g(x) = x^2$.

s.8.5. Note that $g(x) = x^2$ is not monotone increasing, moreover, $g^{-1}(y)$ does not exist (in $\mathbb{R}$) for $y < 0$. We split the line into disjoint intervals in which g is either strictly increasing or decreasing, and then we apply the above rule in each of the intervals. Since $g'(x) = 2x$ and $x = \pm\sqrt{y}$,

$$f_Y(y) = f_X(\sqrt{y}) \frac{1}{2\sqrt{y}} + f_X(-\sqrt{y}) \frac{1}{2\sqrt{y}}.$$

Ex 8.6. If $X \sim \text{Exp}(1)$, use the change-of-variables theorem to obtain the density of $Y = g(X) = \lambda X$. What is $\mathsf{E}[Y]$?

s.8.6. Take $y = g(x) = \lambda x$. Then,

$$f_Y(y) = f_X(x) \frac{\mathrm{d}x}{\mathrm{d}y} = f_X(x) \frac{1}{g'(x)} = e^{-y/\lambda} \frac{1}{\lambda}.$$

With this,

$$\mathsf{E}[Y] = \int_0^\infty y f_Y(y) \,\mathrm{d}y = \int_0^\infty y e^{-y/\lambda} \frac{1}{\lambda} \,\mathrm{d}y.$$

To solve this integral, I recognize y/λ in the exponent, and I want to get rid of the $1/\lambda$ factor. Hence, I write $u = y/\lambda$, and use this to see that

$$u = y/\lambda \implies \mathrm{d}u = \mathrm{d}y/\lambda \implies \mathrm{d}y = \lambda \,\mathrm{d}u.$$

[1]

Then, including a and b for the boundaries to show explicitly what is going on when changing the variables

$$\int_a^b y/\lambda e^{-y/\lambda}\,\mathrm{d}y = \int_{a/\lambda}^{b/\lambda} ue^{-u}\lambda\,\mathrm{d}u = \lambda\int_{a/\lambda}^{b/\lambda} ue^{-u}\,\mathrm{d}u.$$

Applying this to our case so that $a = 0/\lambda = 0$ and $b = \infty/\lambda = \infty$,

$$\mathsf{E}[Y] = \lambda\int_0^\infty ue^{-u}\,\mathrm{d}u = \lambda\,\mathsf{E}[X].$$

Ex 8.7. Show that the 1D change-of-variables formula relates directly to the substitution rule of integration theory to solve 1D integrals.

s.8.7. When we have the density f_Y and the function g, then the substitution rule says that,

$$\int_a^b f_Y(g(x))g'(x)\,\mathrm{d}x = \int_{g(a)}^{g(b)} f_Y(y)\,\mathrm{d}y.$$

We also want that the transformation from X to Y does not affect the probability of the set (event) $A = [a, b]$, hence,

$$\int_{g(a)}^{g(b)} f_Y(y)\,\mathrm{d}y = \int_a^b f_X(x)\,\mathrm{d}x.$$

Combining the above two equations gives that

$$\int_a^b f_Y(g(x)))g'(x)\,\mathrm{d}x = \int_a^b f_X(x)\,\mathrm{d}x.$$

Since this holds for any a and b, it follows that

$$f_Y(g(x))g'(x) = f_X(x).$$

Ex 8.8. Use the change of variable formula to relate the Geo(p) and the FS(p) distributions.

s.8.8. If X follows a geometric distribution and we have $Y = X+1$, then Y follows a first success distribution. This is just re-indexing, as you will also find with the change of variables formula.

Ex 8.9. BH.8.1.1 write that 'The support of Y is all $g(x)$ with x in the support of X.' Do they say that $\operatorname{supp}(Y) = \{x : g(x) \in \operatorname{supp}(X)$? BTW, what is the difference between $\operatorname{supp}(X)$ and $\sup X$?

h.8.9. $\sup$ stands for supremum, $\operatorname{supp}()$ stands for support.

s.8.9. They say that $\operatorname{supp}(Y) = \{g(x) : x \in \operatorname{supp}(X)\}$. The support of a function f is defined as $\operatorname{supp}(f) = \{\omega : f(\omega) \neq 0\}$. The supremum of a set is the least upper bound of that set, and the supremum of a function would be the supremum of its range.

Ex 8.10. BH.8.1.3. Check how all moments were found.

s.8.10. See page 284 of BH.

Ex 8.11. Let $X \sim \text{Unif}(0,5)$. Using the one dimensional change of variables theorem (BH.8.1.1), derive the PDF and the CDF of $Z = 3X$. Explicitly specify the domain.

s.8.11.

$$X \in [0,5] \implies Z \in [0,15], \tag{8.1.6}$$

$$z = 3x = g(x) \implies x = z/3, \tag{8.1.7}$$

$$f_Z(z) = f_X(x)\frac{\mathrm{d}x}{\mathrm{d}z}, \tag{8.1.8}$$

$$\frac{\mathrm{d}z}{\mathrm{d}x} = 3, \tag{8.1.9}$$

$$f_Z(z) = f_X(z/3)\frac{1}{3}. \tag{8.1.10}$$

$F_Z(u) = 1$ for $u \geq 15$ and $F_Z(u) = 0$ for $u \leq 0$. When $0 \leq u \leq 15$,

$$F_Z(u) = \int_0^u f_X(z/3)\frac{1}{3}\,\mathrm{d}z = \frac{1}{5}\int_0^u I_{0\leq z/3 \leq 5}\frac{1}{3}\,\mathrm{d}z \tag{8.1.11}$$

$$= \frac{1}{5}\int_0^u I_{0\leq z\leq 15}\frac{1}{3}\,\mathrm{d}z = \frac{u}{15}. \tag{8.1.12}$$

Ex 8.12. When $Z = X^3$ and $X \sim \text{Unif}(0,5)$, using the one dimensional change of variables theorem to derive the PDF and the CDF of Z. Specify the domain of Z.

s.8.12.

$$X \in [0,5] \implies Z \in [0,125], \tag{8.1.13}$$

$$z = x^3 = g(x) \implies x = z^{1/3}, \tag{8.1.14}$$

$$f_Z(z) = f_X(x)\frac{\mathrm{d}x}{\mathrm{d}z}, \tag{8.1.15}$$

$$\frac{\mathrm{d}z}{\mathrm{d}x} = 3x^2 = 3z^{2/3}, \tag{8.1.16}$$

$$f_Z(z) = f_X(z^{1/3})\frac{1}{3z^{2/3}}. \tag{8.1.17}$$

When $F_Z(u) = 1$ for $u \geq 125$ and $F_Z(u) = 0$ for $u \leq 0$. When $0 \leq u \leq 125$,

$$F_Z(u) = \int_0^u f_X(z^{1/3})\frac{1}{3z^{2/3}}\,\mathrm{d}z = \frac{1}{5}\int_0^u I_{0\leq z^{1/3}\leq 5}\frac{1}{3z^{2/3}}\,\mathrm{d}z \tag{8.1.18}$$

$$= \frac{1}{5}\int_0^u I_{0\leq z\leq 125}\frac{1}{3z^{2/3}}\,\mathrm{d}z \tag{8.1.19}$$

$$= \frac{1}{5}\int_0^u \frac{1}{3z^{2/3}}\,\mathrm{d}z = \frac{1}{5}z^{1/3}\Big|_0^u = u^{1/3}/5. \tag{8.1.20}$$

Ex 8.13. Let $X \sim \text{Norm}(\mu, \sigma^2)$. Using the one dimensional change of variables theorem BH.8.1.1, show that $Z = \frac{X-\mu}{\sigma} \sim \text{Norm}(0,1)$.

s.8.13.

$$
\begin{aligned}
z = g(x) = (x-\mu)/\sigma, &\implies x = \sigma z + \mu && (8.1.21)\\
f_Z(z) &= f_X(x)\frac{\mathrm{d}x}{\mathrm{d}z}, && (8.1.22)\\
\frac{\mathrm{d}z}{\mathrm{d}x} &= \frac{1}{\sigma}, && (8.1.23)\\
f_Z(z) &= f_X(x)\sigma = \sigma f_X(\sigma z + \mu) && (8.1.24)
\end{aligned}
$$

and now using the density of $X \sim \mathrm{Norm}(\mu, \sigma)$,

$$
f_Z(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(\sigma z + \mu - \mu)^2/2\sigma^2}\sigma = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}. \tag{8.1.25}
$$

Ex 8.14. Let $X \sim \mathrm{Exp}(1)$. Derive the PDF of e^{-X}.

s.8.14.

$$
\begin{aligned}
z = g(x) = e^{-x} &\implies x = -\log z, && (8.1.26)\\
x \in (0, \infty) &\implies z \in (0,1), && (8.1.27)\\
f_Z(z) &= f_X(x)\frac{\mathrm{d}x}{\mathrm{d}z}, && (8.1.28)\\
\frac{\mathrm{d}z}{\mathrm{d}x} &= -e^{-x}, \quad \text{Don't forget to take the abs value next,} && (8.1.29)\\
f_Z(z) &= f_X(x)e^x = e^{-x}e^x = 1\, I_{0<z<1}, && (8.1.30)
\end{aligned}
$$

where we include the domain of Z in the last equality.

Ex 8.15. Let X, Y be iid standard normal. Using the n-dimensional change of variables theorem, derive the joint PDF of $(X+Y, X-Y)$.

Check your final answer using BH.7.5.8.

s.8.15.

$$
\begin{aligned}
(u, v) = (x+y, x-y) = g(x,y) &\implies (x,y) = ((u+v)/2, (u-v)/2), && (8.1.31)\\
\frac{\partial(u,v)}{\partial(x,y)} &= \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2 \implies |-2| = 2, && (8.1.32)\\
f_{U,V}(u,v) &= f_{X,Y}(x,y)\frac{\partial(x,y)}{\partial(u,v)} = f_{X,Y}((u+v)/2, (u-v)/2)/2 && (8.1.33)\\
&= \frac{1}{4\pi} e^{-((u+v)/2)^2/2} e^{-((u-v)/2)^2/2} && (8.1.34)\\
&= \frac{1}{4\pi} e^{-u^2/4 - v^2/4}, && (8.1.35)
\end{aligned}
$$

where we work out the squares and simplify. Hence, U and V are independent and normally distributed with mean 0 and $\sigma = \sqrt{2}$. This is in line with our earlier definition of a multi-variate normal distribution.

Ex 8.16. Specify the domain of the new random variable for the following transformations; this important aspect of the change of variables is often overlooked. Let U, V, W, X, X_1, X_2, Y and Z be rvs and let a, b and c be arbitrary constants.

1. $Z = Y^4$ for $Y \in (-\infty, \infty)$;
2. $Y = X^3 + a$ for $X \in (0,1)$;
3. $U = |V| + b$ for $V \in (-\infty, \infty)$;
4. $Y = e^{X^3}$ for $X \in (-\infty, \infty)$;
5. $V = U\, I_{U \le c}$ for $U \in (-\infty, \infty)$;
6. $Y = \sin(X)$ for $X \in (-\infty, \infty)$;
7. $Y = \frac{X_1}{X_1 + X_2}$ for $X_1 \in (0, \infty)$ and $X_2 \in (0, \infty)$;
8. $Z = \log(UV)$ for $U \in (0, \infty)$ and $V \in (0, \infty)$.

s.8.16.

1. $Z = Y^4 \in [0, \infty)$ for $Y \in (-\infty, \infty)$;
2. $Y = X^3 + a \in (a, a+1)$ for $X \in (0,1)$;
3. $U = |V| + b \in [b, \infty)$ for $V \in (-\infty, \infty)$;
4. $Y = e^{X^3} \in (0, \infty)$ for $X \in (-\infty, \infty)$;
5. $V = U\, I_{U \le c} \in (-\infty, c]$ for $U \in (-\infty, \infty)$;
6. $Y = \sin(X) \in [-1, 1]$ for $X \in (-\infty, \infty)$;
7. $Y = \frac{X_1}{X_1 + X_2} \in (0,1)$ for $X_1 \in (0, \infty)$ and $X_2 \in (0, \infty)$;
8. $Z = \log(UV) \in (-\infty, \infty)$ for $U \in (0, \infty)$ and $V \in (0, \infty)$.

Ex 8.17. When adding a different equality, we need to be careful to not create a functional relationship between our two new variables U, V, for example $U = X + Y$ and $V = \sin(X + Y)$, or $U = \frac{X}{Y}$ and $V = \frac{Y}{X}$ for conforming X, Y. What would happen to the determinant of the Jacobian matrix if we did? Why would this happen? Explain in your own words.

s.8.17. When the variables become dependent, the Jacobian becomes zero. For instance, in the latter case,

$$\frac{\partial(u,v)}{\partial(x,y)} = \begin{vmatrix} 1/y & -x/y^2 \\ -y/x^2 & 1/x \end{vmatrix} = \frac{1}{xy} - \frac{x}{y^2}\frac{y}{x^2} = 0. \tag{8.1.36}$$

Moreover, the function g is not locally one-to-one.

8.2 SECTION 8.2

Ex 8.18. To find the distribution of a convolution through the change of variables formula, we seem to need to add a 'redundant' equality? But why is that? What would be the problem if we do not add this? Explain in your own words.

s.8.18. If we would not add this extra variable, we cannot use the change of variables theorem. We also need a function to deal with the scaling. In the change of variables theorem, this is the Jacobian.

There is also another problem. Consider the function $g(x, y)$ that maps $\mathbb{R}^2$ to $\mathbb{R}$. The inverse set $\{(x, y) : g(x, y) = z\}$ can be quite complicated, while the set $\{y : g(x, y) = z\}$ for a fixed x is hopefully just one point. Hence, the mapping $(x, y) \to (x, g(x, y))$ is, at least locally, one-to-one.

It is possible to deal with the more general problem, but this requires much more theory than we need for this course.

Ex 8.19. In this exercise, we combine what we learned in BH.8.1.4 and BH.8.1.9. Let S be the sum of two iid chi-square distributed variables (with one degree of freedom). Using just these two examples, show that $S \sim \text{Exp}(1/2)$.

h.8.19. Let X, Y be iid standard normal. Since the square of a standard normal r.v. is chi-square distributed, we can write S as $S = X^2 + Y^2$ (here we use BH.8.1.4).

s.8.19. From BH.8.1.4: Z chi-square $\implies X = \sqrt{Z} \sim \text{Norm}(0, 1)$. Then, from BH.8.1.9,

$$X^2 + Y^2 = (\sqrt{2T}\cos U)^2 + (\sqrt{2T}\sin U)^2 = 2T\left(\cos^2 U + \sin^2 U\right) = 2T \sim \text{Exp}(1/2), \tag{8.2.1}$$

when $X, Y \sim \text{Norm}(0, 1)$.

Ex 8.20. A student has obtained an iid random sample of size 2 from a Cauchy distribution. Let the rvs X and Y model the values of the first and second sample. Since s/he does not know what the mean of a Cauchy distribution is, s/he wants to average the sample to obtain what she thinks is a good estimate of the true mean.

To find the distribution of this sample mean, we need to find an expression for $f_W(w)$, where $W = \frac{X+Y}{2}$.

1. Find an expression for $f_W(w)$ in the form of an integral, but do not solve it.

2. It turns out that if we solve the integral, we get that $f_W(w) = f_X(w)$. The distribution of our sample mean is still Cauchy; we did not obtain a better estimate of the Cauchy mean by calculating the sample mean!

   Explain (in your own words) why this makes sense.

s.8.20. Take $g(x, y) = (x, w) = (x, (x + y)/2)$. Then, $y = 2w - x$.

$$\frac{\partial(x,w)}{\partial(x,y)} = \begin{vmatrix} 1 & 0 \\ 1/2 & 1/2 \end{vmatrix} = 1/2, \tag{8.2.2}$$

$$f_{X,W}(x,w) = f_{X,Y}(x,y)\frac{\partial(x,y)}{\partial(x,w)} = \frac{1}{\pi(1+x^2)}\frac{1}{\pi(1+(2w-x)^2)}2, \tag{8.2.3}$$

$$f_W(w) = \int_{-\infty}^{\infty} f_{X,W}(x,w)\,\mathrm{d}x = \frac{2}{\pi^2}\int_{-\infty}^{\infty}\frac{1}{1+x^2}\frac{1}{1+(2w-x)^2}\,\mathrm{d}x. \tag{8.2.4}$$

The expectation of a Cauchy distributed r.v. X is not well-defined because $\mathsf{E}[|X|] = \infty$. As a consequence, taking the average of some outcomes (i.e. a sample) will also not give a sensible answer.

8.3 SECTION 8.3

Ex 8.21. If $a = b = 1$, why is Beta(a, b) = Unif([0, 1]])?

s.8.21. When $a = b = 1$, the probability density functions are the same, hence the distributions are the same.

Ex 8.22. If $a = b$, why is Beta(a, b) symmetric?

s.8.22. Because $\beta(a, b)$ and $x^a(1-x)^b$ are (around $\frac{1}{2}$).

Ex 8.23. If $a > b$ and $X \sim$ Beta(a, b), is $\mathsf{E}[X] > 1/2$?

s.8.23. $\mathsf{E}[X] = \frac{a}{a+b}$.

Ex 8.24. BH.8.3.2, last equation. How do the authors get to the equation

$$\beta(a, b) = \frac{1}{(a+b-1)\binom{a+b-2}{a-1}}?$$

s.8.24. They notice that $\int_0^1 \binom{n}{k} p^k(1-p)^{n-k} \, \mathrm{d}p = \frac{1}{n+1}$ for all $0 \le k \le n$. Then filling in $k = a-1, n = b-1$, the claim follows.

Ex 8.25. BH.8.3.3. The authors write that X is not marginally Binomial, but is conditionally Binomial. What is the difference?

s.8.25. A marginal distribution is simply a distribution, called marginal to emphasize is was marginalized out from another, joint, distribution. A conditional distribution is a distribution given some restrictions from another random variable.

Ex 8.26. BH.8.3.3. The authors use a smart trick to find an expression for the posterior distribution $f_{p|X=k}$ of p. Use this posterior to derive an expression for $\mathsf{P}\{X = k\}$ by using the fact that

$$\mathsf{P}\{X = k\} = \frac{\binom{n}{k} p^k(1-p)^{n-k}\text{Beta}(p; a, b)}{\text{Beta}(p; a+k, b+n-k)},$$

and simplyfing the RHS.

s.8.26. This is just some algebra, look up the beta-binomial distribution to see if you got it right.

Ex 8.27. BH.8.3.3. Why does $a-1$ correspond to the number of prior successes, in other words, why is it not a, but $a-1$?

s.8.27. If we have no prior successes, it makes most sense for our prior to be uniform. Similarly for failures. Since the uniform distribution corresponds to $a = b = 1$, we need $a-1$ to represent the number of successes.

Ex 8.28. BH.8.3.4.b. Given that the first patient is cured, what is the probability that the rest of the patients, i.e., the other $n-1$, will also be cured?

s.8.28. Let $X_{1,\dots,n}$ denote the random variable counting how many of the first n people are cured, with similar notation for other subsets. Then, we calculate:

$$\begin{aligned}
\mathsf{P}\left\{X_{2,\dots,n} = n-1 | X_1 = 1\right\} &= \frac{\mathsf{P}\left\{X_{2\dots,n} = n-1, X_1 = 1\right\}}{\mathsf{P}\left\{X_1 = 1\right\}} \\
&= \frac{\int_0^1 \mathsf{P}\left\{X_{2\dots,n} = n-1, X_1 = 1 | p\right\} \mathrm{d}p}{\int_0^1 \mathsf{P}\left\{X_1 = 1 | p\right\} \mathrm{d}p} \\
&= \frac{\int_0^1 \mathsf{P}\left\{X_{2\dots,n} = n-1 | p\right\} \mathsf{P}\left\{X_1 = 1 | p\right\} \mathrm{d}p}{\int_0^1 \mathsf{P}\left\{X_1 = 1 | p\right\} \mathrm{d}p} \\
&= \frac{\int_0^1 p^n \mathrm{d}p}{\int_0^1 p \mathrm{d}p} \\
&= \frac{2}{n+1},
\end{aligned}$$

where we use the definition of conditional probability, the law of total probability, independence given p, and some easy integration, in that order.

Ex 8.29. Is this claim correct? Let T be the sum of two iid Unif$(0,1)$ rvs. Then there exist a, b such that $T \sim \text{Beta}(a,b)$. (You don't need to derive the distribution of T.)

s.8.29. Incorrect: The support of T is $(0,2)$ whereas the support of any beta distribution is $(0,1)$. Hence, T does not have a beta distribution for some a, b.
Also see page 378 of the book for the distribution of the sum of two uniform distributions. This might help your intuition for this solution.

Ex 8.30. Show that $\beta(1,b) = 1/b$ by integrating the PDF of the beta distribution for $a = 1$. (Do not use the results of BH 8.5 for this exercise.)

s.8.30. We use that the PDF integrates to 1:

$$1 = \int_0^1 \frac{1}{\beta(1,b)} (1-x)^{b-1} \mathrm{d}x = \frac{1}{\beta(1,b)} \left[-\frac{1}{b} (1-x)^b \right]_0^1 = \frac{1}{\beta(1,b) b}.$$

Hence, $\beta(1,b) = \frac{1}{b}$.

Ex 8.31. Let $a, b > 1$. Show that the PDF of the beta distribution attains a maximum at $x = \frac{a-1}{a+b-2}$. Explicitly indicate where the assumption that $a, b > 1$ is used.

s.8.31. The scaling factor $\beta(a,b)$ is a positive constant, so we may as well leave it out and maximize $x^{a-1}(1-x)^{b-1}$. Note that its derivative (to x) is given by

$$\begin{aligned}
\frac{d}{dx} x^{a-1}(1-x)^{b-1} &= ((a-1)(1-x) - (b-1)x) x^{a-2} (1-x)^{b-2} \\
&= ((a-1) - (a+b-2)x) x^{a-2} (1-x)^{b-2}.
\end{aligned}$$

Setting this to zero yields $x = \frac{a-1}{a+b-2}$ as the only candidate for an interior optimum. Since $a, b > 1$, we have $0 < x < 1$. If $a, b > 1$, then the PDF converges to 0 as $x \to 0$ or $x \to 1$, so then we conclude that $x = \frac{a-1}{a+b-2}$ indeed yields a maximum. (Think about this last sentence; most students do not use the information that $a, b > 1$ correctly.)

Ex 8.32. Explain in your own words:

1. What is a prior?

2. What is a conjugate prior?

s.8.32. A prior is a distribution reflecting one's information or belief about a parameter before updating it with information.

It is harder than you might think, hardly any student gives a completely satisfactory answer here. Compare your solution to the definition above. If they are different, try to understand how exactly your solution was different and determine which definition is better.

A conjugate prior is a prior distribution such that the posterior distribution is in the same family of distributions.

Ex 8.33.

1. Look up on the web: what is the conjugate prior of the multinomial distribution? Give a name and a formula.

2. Explain why the Beta distribution is a special case of this distribution.

s.8.33. Dirichlet distribution. The Beta distribution is a special case of the Dirichlet distribution, because binomial is a special case of multinomial. Of course, this can also be shown directly using the formula.

Ex 8.34. You make a test with n multiple choice questions and you give the correct answer to each question independently with probability p. The teacher's prior belief about p is reflected by a uniform distribution: $p \sim \text{Unif}(0, 1)$. Let X be the number of correct answers you give. What is the teacher's posterior distribution $p|X = k$? (You don't have to do a lot of math here; simply use a result from the book.)

s.8.34. The prior is $p \sim \text{Beta}(1, 1)$. The posterior is $p|X = k \sim \text{Beta}(1 + k, 1 + n - k)$.

Ex 8.35. You find a coin on the street. Initially, you are rather confident that this should be (approximately) a fair coin. This is reflected in your prior belief of the probability p of heads: $p \sim \text{Beta}(10, 10)$. Your friend is a bit more skeptical and assumes a uniform prior: $p \sim \text{Unif}(0, 1)$. You toss the coin 1000 times, and it comes up heads 900 times.

1. Determine your posterior distribution. (Again, use a result from the book)

2. Determine your friend's posterior distribution.

3. Compare the means of your posterior distribution and your friend's posterior distribution. Comment on the effect of the prior distribution if you have a lot of data.

s.8.35. Let X denote the number of heads.

1. Your posterior is $p|X = 900 \sim \text{Beta}(910, 110)$.

2. Your friend's posterior is $p|X = 900 \sim \text{Beta}(901, 101)$.

3. The mean of your posterior is $\frac{910}{910+110} = \frac{91}{102} \approx 0.892$; the mean of your friend's posterior is $\frac{901}{901+101} = \frac{901}{1002} \approx 0.899$. The difference is small, so the effect of the prior distribution is small if you have a lot of data. This effect is known as *washing out the prior.*

Ex 8.36. We have an urn with 1000 coins. One of those is biased such that $\mathsf{P}\{H\} = 99/100 = 1 - \mathsf{P}\{T\}$, all others are fair. You select at random a coin, i.e., with probability $1/1000$ you select the biased one, and start throwing. You see 10 heads in row. What is the probability you picked the biased coin?

s.8.36. Of course, this should depend on when we see this string of outcomes: if it happens for the first time after a thousand throws then we can still confidently claim the coin is fair. Assuming this happens in the first 10 throws however, we use Bayes' theorem to find the following:

$$\begin{aligned}\mathsf{P}\{B|H_{10}\} &= \frac{\mathsf{P}\{H_{10}|B\}\,\mathsf{P}\{B\}}{\mathsf{P}\{H_{10}|B\}\,\mathsf{P}\{B\} + \mathsf{P}\{H_{10}|B^c\}\,\mathsf{P}\{B^c\}} \\ &= \frac{(99/100)^{10}/1000}{(99/100)^{10}/1000 + (1/2)^{10}\cdot 999/1000} \\ &\approx 0.48.\end{aligned}$$

Ex 8.37. BH.8.3.5. write that X_j is an indicator of the jth throw beind made. Can this be the formal definition: $X_j = I_{N\geq j}$?

s.8.37. No, of course not. She gets N free throws, but her getting e.g. two tries in a game is not equivalent to her scoring the first one.

Ex 8.38. Use the pmf of the Beta-Binomial distribution to prove the following identity:

$$\sum_{k=0}^{n} \frac{\binom{n}{k}\binom{a+b-2}{a-1}(a+b-1)}{\binom{a+b+n-2}{a+k-1}(a+b+n-1)} = 1.$$

for all positive integers a, b, n.

s.8.38. This states that the PMF of the Beta-Binomial distribution,

$$P(X = k) = \binom{n}{k}\frac{\beta(a+k, b+n-k)}{\beta(a,b)},$$

sums to 1. To see this, we have to rewrite the beta functions in terms of binomial coefficients:

$$\frac{1}{\beta(a,b)} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} = \frac{(a+b-1)!}{(a-1)!(b-1)!} = (a+b-1)\binom{a+b-2}{a-1},$$
$$\frac{1}{\beta(a+k,b+n-k)} = (a+b+n-1)\binom{a+b+n-2}{a+k-1}.$$

Plugging this in gives the result.

8.4 SECTION 8.4

Ex 8.39. What is the SCV of Gamma(n,λ) distributed rv X?

s.8.39. $\mathsf{V}[X] = n/\lambda^2$, $\mathsf{E}[X] = n/\lambda$, SCV $= 1/n$.

Ex 8.40. We have a machine that has temperature $x_0 e^{-\alpha t}$ after a time $t \geq 0$. We switch the machine on after the arrival of q jobs. Job interarrival times are iid $\sim$ Exp(λ). The temperature at the moment the qth job arrives has the distribution $x_0 \exp\{-\alpha Y\}$, with $Y \sim$ Gamma(q,λ). Explain why this is so.

s.8.40. We have to wait for q exponentially distributed interarrival times. The sum of these interarrival times is Gamma$(1,\lambda)$.

Ex 8.41. Consider the chi-square distribution (with one degreee of freedom) from BH.8.1.4.

Starting from the expression $f_Y(y) = \varphi\left(\sqrt{y}\right) y^{-1/2}$ in this example, show that this chi-square distribution is a special case of the Gamma distribution and specify the corresponding values of the parameters a and λ.

s.8.41. We fill in $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ to find

$$f_Y(y) = \varphi\left(\sqrt{y}\right) y^{-1/2} = \frac{1}{\sqrt{2\pi}} y^{-1/2} e^{-(\sqrt{y})^2/2} = \frac{1}{\sqrt{2\pi}} y^{-1/2} e^{-y/2},$$

so $a = \frac{1}{2}$ and $\lambda = \frac{1}{2}$.

Ex 8.42. Is the sum of any two Gamma distributions again Gamma?

s.8.42. Incorrect: the scale parameters λ need to be the same *and* both random variables need to be independent.

Ex 8.43. Prove by induction that $\Gamma(n) = (n-1)!$ if n is a positive integer.

s.8.43. The base case is $n = 1$. We have $\Gamma(1) = \int_0^\infty e^{-x}\,\mathrm{d}x = 1 = 0!$, so the statement holds for $n = 1$. Now let $k \in \mathbb{N}$ be arbitrary and assume that the statement holds for $n = k$, i.e. that $\Gamma(k) = (k-1)!$. Then

$$\Gamma(k+1) = k\Gamma(k) = k(k-1)! = k! = ((k+1)-1)!, \tag{8.4.1}$$

so the statement also holds for $n = k+1$. By mathematical induction, we conclude that $\Gamma(n) = (n-1)!$ for all positive integers n.

Ex 8.44. Is the Poisson distribution the conjugate prior of the Gamma distribution?

s.8.44. Incorrect: It is the other way around, the Gamma distribution is the conjugate prior of the Poisson distribution. This statement doesn't make much sense, for example one would need to say for which parameter of the Gamma distribution it is the prior. In addition, the parameters of the Gamma distribution can be any positive real number, so the conjugate prior of (either parameter) of the Gamma distribution is a continuous distribution, so in particular not the Poisson distribution.

Ex 8.45. Let $X \sim \text{Gamma}(4,2)$ and $Y \sim \text{Gamma}(7,2)$ be independent rvs. What is the distribution of $X + Y$? What is the distribution of $\frac{X}{X+Y}$?

s.8.45. $X + Y \sim \text{Gamm}(11,2)$ and $\frac{X}{X+Y} \sim \text{Beta}(4,7)$.

8.5 SECTION 8.6

Read the definition of an order statistic. Skip the rest of BH.8.6.

Ex 8.46. If you can answer this question, then you basically know everything you need to know about order statistics for the purpose of this course.)
Let $X_1, X_2, \ldots, X_9$ be a collection of random variables. Fill in the gaps (with just one word each time):

1. $X_{(1)}$ denotes the ... of $X_1, X_2, \ldots, X_9$.
2. $X_{(9)}$ denotes the ... of $X_1, X_2, \ldots, X_9$.
3. $X_{(5)}$ denotes the ... of $X_1, X_2, \ldots, X_9$.

s.8.46.

1. Minimum
2. Maximum
3. Median

9 CHAPTER 9: EXERCISES AND REMARKS

9.1 SECTION 9.1

Remark 9.1. Skip BH.9.1.6.

Ex 9.2. BH.9.1.7. We will also simulate this in an assignment.

I like this example as it shows how to make optimal decisions under uncertainty, but I have to admit that I don't understand the reasoning, or the use of conditional probability to solve this problem. Here is how I would solve the problem.

1. Explain that $W = (V - b)\, I_{b \geq \alpha V}$, with $\alpha = 2/3$, is the rv that models our payoff.
2. Why is this wrong: $\mathsf{E}[W] = \mathsf{E}[(V-b)\, I_{b\geq\alpha V}] = V\,\mathsf{E}[I_{b\geq\alpha V}] - b\,\mathsf{E}[I_{b\geq\alpha V}]$?
3. Compute $\mathsf{E}[W]$, and provide a bound on α to ensure that $\mathsf{E}[W] > 0$.

s.9.2.

1. The indicator function is zero when the bid is rejected, hence there is a zero payoff. When the bid is accepted the indicator function is one and the resulting payoff is $V - b$, i.e., the prize value minus the offer.
2. V is a rv, hence cannot be taken out the expectation.
3. Since W is a function of V we can apply LOTUS and find the expectation by integration. First we assume $b/\alpha < 1$.

$$
\begin{aligned}
\mathsf{E}[W] &= \int_0^1 (v-b)\, I_{b\geq\alpha v}\, \mathrm{d}v = \int_0^{b/\alpha} v\,\mathrm{d}v - b\int_0^{b/\alpha} \mathrm{d}v \\
&= b^2/2\alpha^2 - b^2/\alpha = \frac{b^2}{\alpha^2}\frac{1-2\alpha}{2}.
\end{aligned}
$$

Clearly, we need to assume $\alpha < 1/2$ to ensure that $\mathsf{E}[W] > 0$.

Next, assume that $b > \alpha$. Then the expected payout is $\mathsf{E}[W] = 1/2 - b$, as any bid is accepted (why?). Thus, if $b < 1/2$ the expected payout is positive. But we assumed that $b > \alpha$, hence $\alpha < b < 1/2$, so again $\alpha < 1/2$ for a positive payout.

Ex 9.3. BH.9.1.7. continued. By the previous exercise, we know that only when $\alpha < 1/2$ we should participate in the game. In other words, when our bid b should be larger than αV to be accepted. If we know α, what would be our optimal bid b?

s.9.3. When $\alpha < 1/2$ the expected payoff is positive so it becomes interesting. We can write the total payout as:

$$\mathsf{E}[W] = (1/2 - b)\, I_{b>\alpha} + \frac{1-2\alpha}{2\alpha^2} b^2\, I_{b\leq\alpha}$$

Assume $0 < \alpha < 1/2$. The expected payoff is a piecewise continuous function since $\lim_{b\uparrow\alpha} \mathsf{E}[W] = 1/2 - \alpha = \lim_{b\downarrow\alpha} \mathsf{E}[W]$. Now, the left term in the RHS of the equation above, i.e., $(1/2 - b)\, I_{b>\alpha}$, decreases strictly for $b > \alpha$, while in the right term b^2 increases strictly for $b < \alpha$. Hence, $\mathsf{E}[W]$ achieves its maximum for $b = \alpha$. Therefore, the optimal betting amount is α and results in an expected payoff of $1/2 - \alpha$.

Ex 9.4. BH.9.1.8. Apply the same type of argumentation to find $\mathsf{E}[X]$ when $X \sim \mathrm{FS}(p)$.

s.9.4. $\mathsf{E}[X] = 1 + q\,\mathsf{E}[X]$, because we have to throw at least once, and with probability q, we start again. Hence, $\mathsf{E}[X] = 1/(1-q) = 1/p$.

Ex 9.5. BH.9.1.8. Use first step analysis to find $N_r := \mathsf{E}[X]$ when $X \sim \mathrm{NBin}(r, p)$.

s.9.5. Suppose the first throw is a success, then we need $r - 1$ more successes, if the first throw is a failure, we are back at 'hole one'. Thus, $N_r = pN_{r-1} + q(1 + N_r)$. Simplifying (and using that $p/(1-q) = 1$) gives $N_r = N_{r-1} + q/p$, which implies $N_r = rq/p$.

Ex 9.6. BH.9.1.9. I reason slightly differently here. Write N_r for the number of throws required to reach r heads in row. Then I need N_{r-1} throws in expectation to reach the state in which there are $r-1$ heads in row. Suppose now that we are in this state, i.e., there are $r-1$ heads in row. Then, if I throw heads, with probability p, I reach the state with r heads in row, and I am done. However, if I throw tails, with probability q, I have to start all over again. Use this argument to derive the recursion $N_r = N_{r-1} + p \cdot 1 + q(1 + N_r)$. Solve this to obtain

$$N_r = \sum_{i=1}^{r} 1/p^i. \tag{9.1.1}$$

h.9.6. If this is new to you, check out appendix A.4

s.9.6.

$$N_r = N_{r-1} + p \cdot 1 + q(1 + N_r) \implies N_r = N_{r-1}/p + 1/p \implies N_r = \sum_{i=1}^{r} 1/p^i. \tag{9.1.2}$$

Ex 9.7. Compute the expected outcome of a die throw (with a 6-sided die), given that the outcome is even. Introduce proper notation for random variables and events.

s.9.7. Let X be the outcome of the die throw (note that X is a random variable) and let A be the event that the outcome is even. Then

$$\mathsf{E}[X \mid A] = 2\,\mathsf{P}\{X = 2|A\} + 4\,\mathsf{P}\{X = 4|A\} + 6\,\mathsf{P}\{X = 6|A\} = \tfrac{1}{3} \cdot (2 + 4 + 6) = 4.$$

We conclude that $\mathsf{E}[X \mid A] = 4$.

Ex 9.8. BH.9.1.10. Let p_i, $0 \leq i \leq b$, be the probability to hit b before 0, with i being the current position of the drunkard.

1. Why is $p_0 = p_1/2$?
2. Why is $p_1 = p_2/2$?
3. Why is $p_b = 1$?
4. Explain the recursion $p_i = p_{i-1}/2 + p_{i+1}/2$ for $1 < i < b$?
5. Show that the previous points imply that $p_i = \alpha i$ for $0 < i < b$ for any α we like.
6. Combine the fact that $p_b = 1$ with $p_i = \alpha i$ for all $0 < i < b$ to see that $\alpha = 1/b$.
7. Conclude that $p_0 = 1/2b$.

h.9.8. Read the gambler's ruin problem BH.2.7.3

s.9.8.

1. As in BH, condition on the first step. If the drunkard starts at zero, and makes a step to the left, he is at position -1. To get to b, with $b > 0$, requires to pass 0 again. Hence, in this case, the drunkard did not reach b before 0.

   When the drunkard makes a step to the right, which occurs with probability $1/2$, then the drunkard moves to position 1. Writing p_1 for the probability to reach b before 0, it must be that $p_0 = p_1/2$.

   Written in another way: $p_0 = p_{-1} \cdot 1/2 + p_1 \cdot 1/2$, but $p_{-1} = 0$, hence $p_0 = p_1/2$.

2. The reasoning is the same as in the previous step. When the drunkard is in state 1 and moves to the left, he hits 0 before b so the process stops. If he moves to the right, then the process continues. Therefore $p_1 = p_2/2$.

3. If the drunkard starts in b, the probability to hit b before 0 is 1. (In other words, what is the probability to be in b when the drunkard starts in b?)

4. In some state i, $2 \leq i < b$, condition again on whether the inebriate moves to the left or to the right. The probability p_i of reaching b before 0 is p_{i+1} when he makes a step to the right and p_{i-1} when he moves to the left. Since a step in both directions is equally likely, the equation of the problem follows.

5. Plug in $p_i = \alpha i$ in the RHS:

   $$\frac{\alpha(i-1)}{2} + \frac{\alpha(i+1)}{2} = \alpha i - \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha i = p_i$$

   This shows $p_i = \alpha i$ is a solution to this recursive equation for any choice of α.

6. Use the boundary condition $p_b = 1$ in the expression $p_i \alpha = i$ for $i = b - 1$. Solving for α in $\alpha(b-1) = \alpha(b-2)/2 + 1/2$, gives $\alpha = 1/b$.

7. Since $p_1 = p - 2/2 = 2/b \cdot 1/2 = 1/b$ and $p_0 = p_1/2$, we get that $p_0 = 1/2b$.

Another way you might have found p_0 is by noticing the following pattern when plugging in the previous values in the equation in 3.

$$
\begin{aligned}
\text{As noted before} \quad & p_1 = p_2/2 \\
\text{Plugging the above result} \quad & p_2 = p_1/2 + p_3/2 = p_2/4 + p_3/2 \\
\text{This implies} \quad & p_2 = 2/3 p_3 \\
\text{Continuing substituting} \quad & p_3 = 1/3 p_3 + p_4/2 \\
& p_3 = 3/4 p_4 \\
\text{Notice a pattern?} \quad & \dots \\
& p_{b-1} = \frac{b-1}{b} p_b = \frac{b-1}{b} \\
\text{As we already knew} \quad & p_b = 1 \\
\text{Now we can recursively substitute what we know to find} \quad & p_1 = \frac{1}{2}\frac{2}{3}\frac{3}{4}\cdots\frac{b-1}{b} = \frac{1}{b} \\
\text{Then using the result from 1} \quad & p_0 = p_1/2 = \frac{1}{2b}
\end{aligned}
$$

9.2 SECTION 9.2

Remark 9.9. About BH. 9.2.1. This definition is subtle, and it takes time to understand. Here is a slightly different explanation; perhaps it's useful for you.

Take some random variable X, say. Then, as in Chapter 7, we can be interested in $\mathsf{E}\left[g(X)\right]$, i.e., the expectation of the rv $g(X)$.

When Y is continuous we can compute $\mathsf{E}[Y \mid X = x]$ with the conditional CDF

$$\mathsf{E}[Y \mid X = x] = \int f_{Y|X}(y|x)\,\mathrm{d}y.$$

(For discrete rv., replace the integral by the PMF.) Observe that this is just a function of x; define this function as $g(x) = \int f_{Y|X}(y|x)\,\mathrm{d}y$. And now, as before, we consider the random variable $g(X)$, and we *call this rv the conditional expectation* of Y given X.

It is true that X plays some sort of double role—first we use it in the conditioning in the definition of the function g, and then we plug it into g again—and this is perhaps confusing. But I finally 'got it', when I understood that g can be interpreted as just some function of x. And then we compute $\mathsf{E}\left[g(X)\right]$, and so on.

Ex 9.10. BH.9.2.2. Is $\mathsf{E}[Y \mid I_A]$ a number or a rv?

h.9.10. Is I_A a rv or an event?

s.9.10. It is a rv as the indicator function isn't crystallized, so we are conditioning on a rv ($I_A \sim \text{Bern}(\mathsf{P}\{A\})$) We could write $\mathsf{E}[Y \mid I_A] = \mathsf{E}[Y \mid A]\, I_A + \mathsf{E}\left[Y \,\middle|\, A^C\right](1 - I_A)$.

9.3 SECTION 9.3

Remark 9.11. On BH.9.3.2 (Taking out what is known.) Perhaps it is easier to crystallize X into x. Then $g(x) = \mathsf{E}[h(x)Y|X] = h(x)\mathsf{E}[Y|X]$, because $h(x)$ is just a function. The rvs $\mathsf{E}[H(X)Y|X]$ and $h(X)\mathsf{E}[Y|X]$ are then both equal to $g(X)$.

Ex 9.12. On BH.9.3.9. Show that $\mathsf{Cov}[Y - \mathsf{E}[Y|X], \mathsf{E}[Y|X]] = 0$.

s.9.12. We have that $\mathsf{E}[Y - \mathsf{E}[Y|X]] = 0$. Hence, $\mathsf{E}[Y - \mathsf{E}[Y|X]]\,\mathsf{E}[X] = 0$. Then define $h(X) = \mathsf{E}[Y|X]$ and apply BH.9.3.9 to see that $\mathsf{E}[(Y - \mathsf{E}[Y|X])h(X)] = 0$. From the definition of the covariance, $\mathsf{Cov}[W, Z] = \mathsf{E}[WZ] - \mathsf{E}[W]\,\mathsf{E}[Z]$, we have shown that both terms are zero.

9.4 SECTION 9.4

Ex 9.13. Consider a casino where, for any $a > 0$, it is possible to pay a euro and get a chance of $\frac{1}{5}$ on receiving $4a$ euro and a chance of $\frac{4}{5}$ of receiving nothing. Adam enters the casino with b euros, and bets half of his money on this gamble. Let X be the amount of money he has after the gamble. After that, he again bets half of the money he then has (i.e. half of X) on this gamble. Let Y be the amount of money he has after the second gamble.

1. Compute $\mathsf{E}[X]$.
2. Compute $\mathsf{E}[Y|X]$.
3. Compute $\mathsf{E}[Y]$.

Explicitly mention the laws/rules you use.

s.9.13. 1. Since Adam keeps $b/2$ and does the gamble with $a = b/2$, we have

$$\mathsf{E}[X] = b/2 + \tfrac{1}{5}\cdot 4(b/2) + \tfrac{4}{5}\cdot 0 = 0.9b.$$

2. The computation is the same as in part 1., but with X instead of b:

$$\mathsf{E}[Y|X] = X/2 + \tfrac{1}{5}\cdot 4(X/2) + \tfrac{4}{5}\cdot 0 = 0.9X.$$

Note that the result is a random variable.
3. Using Adam's law (and linearity of expectation), we conclude that:

$$\mathsf{E}[Y] = \mathsf{E}[\mathsf{E}[Y|X]] = \mathsf{E}[0.9X] = 0.9\,\mathsf{E}[X] = 0.81b.$$

In general, if Adam would do this n times, the expected amount of money he has after n such gambles would be $0.9^n b$. This would be very difficult to show without Adam's law!

Ex 9.14. Let $N \sim \mathrm{Pois}(\lambda)$, and let $X|N \sim \mathrm{Bin}(N, p)$, where $p \in (0,1)$ and $\lambda > 0$ are known constants. Compute $\mathsf{E}[X]$ using Adam's law. Check your answer using the chicken-egg story; with this story you can also obtain the distribution of X.

s.9.14. We have $\mathsf{E}[X|N] = Np$, so using Adam's law (and linearity of expectation), we conclude that $\mathsf{E}[X] = \mathsf{E}[\mathsf{E}[X|N]] = \mathsf{E}\left[Np\right] = \mathsf{E}[N]\,p = \lambda p$.

This is in accordance with $X \sim \mathrm{Pois}(\lambda p)$, which was shown in the chicken-egg story.

Some students reported answers like $\lambda^2 p$. This is wrong, and can be immediately seen by checking units: the unit of λ being 1 per time.

Others wrote $\mathsf{E}[X \mid N = n]\, np$, hence $\mathsf{E}[X] = \mathsf{E}[\mathsf{E}[X \mid N]] = \mathsf{E}\left[np\right] = np$.

Apparently, such students are not aware of the idea that $\mathsf{E}[X \mid N]$ is a random variable. When this happens during the exam, you will score 0 points for that particular part of a question.

Ex 9.15. Correct? If A is an event and I_A is its indicator, then for all random variables X we have $\mathsf{E}[X \mid A] = \mathsf{E}[X \mid I_A]$.

s.9.15. Incorrect: $\mathsf{E}[X \mid A]$ is a number since A is an event, whereas $\mathsf{E}[X \mid I_A]$ is a random variable since I_A is a random variable. A correct statement is $\mathsf{E}[X \mid A] = \mathsf{E}[X \mid I_A = 1]$.

Ex 9.16. Correct? If X and Y are independent, then $\mathsf{V}[\mathsf{E}[Y \mid X]] = 0$.

s.9.16. Correct, if X and Y are independent, then $\mathsf{E}[Y \mid X] = \mathsf{E}[Y]$ which is a constant (formally, a degenerate random variable). Since the variance of a constant is 0, we conclude that $\mathsf{V}[\mathsf{E}[Y \mid X]] = 0$.

Ex 9.17. Let $X \sim \mathrm{Exp}(\lambda)$, and let a be a constant.

1. Compute $\mathsf{E}[X \mid X \geq a]$ using an integral and an indicator.
2. Explain the answer using a property of the exponential distribution.

s.9.17. 1. We compute $\mathsf{E}[X \mid X \geq a]$ as follows:

$$\begin{aligned}
\mathsf{E}[X \mid X \geq a] &= \int_0^\infty y f(y|A)\,\mathrm{d}y \\
&= \int_0^\infty y \frac{\lambda e^{-\lambda y} I_{y \geq a}}{e^{-\lambda a}}\,\mathrm{d}y \\
&= \lambda \int_a^\infty y e^{-\lambda(y-a)}\,\mathrm{d}y \\
&= -y e^{-\lambda(y-a)}\Big|_a^\infty + \int_a^\infty e^{-\lambda(y-a)}\,\mathrm{d}y \\
&= a - \frac{1}{\lambda} e^{-\lambda(y-a)}\Big|_a^\infty = a + \frac{1}{\lambda}.
\end{aligned}$$

2. The result also follows from the memoryless property, which states that conditional on the event that $X \geq a$, we have that $X - a | X \geq a \sim \mathrm{Exp}(\lambda)$.

Ex 9.18. A hat contains 9 fair coins and one coin that lands heads with probability 0.8. You pick a coin from the hat uniformly at random and toss it 10 times. Let A be the event that you pick a fair coin, and let X be the number of heads. Let B be the event that the first four tosses all show heads.

1. Compute $\mathsf{E}[X \mid A]$.
2. Compute $\mathsf{E}[X \mid A^c]$.
3. Compute $\mathsf{E}[X]$.
4. Compute $\mathsf{P}\{B\}$.
5. Compute $\mathsf{P}\{A \mid B\}$.
6. Compute $\mathsf{E}[X \mid B]$.
7. Compute $\mathsf{E}[X \mid B^c]$. *Hint*: it is not necessary to compute $\mathsf{P}\left\{A \mid B^c\right\}$.

s.9.18. 1. Note that $X \mid A \sim \text{Bin}(10, 0.5)$, so $\mathsf{E}[X \mid A] = 10 \cdot 0.5 = 5$.
2. Note that $X \mid A^c \sim \text{Bin}(10, 0.8)$, so $\mathsf{E}[X \mid A^c] = 10 \cdot 0.8 = 8$.
3. By LOTE we have $\mathsf{E}[X] = \mathsf{P}\{A\}\,\mathsf{E}[X \mid A] + \mathsf{P}\{A^c\}\,\mathsf{E}[X \mid A^c] = 0.9 \cdot 5 + 0.1 \cdot 8 = 5.3$.
4. Note that $\mathsf{P}\{B \mid A\} = 0.5^4$ and $\mathsf{P}\{B \mid A^c\} = 0.8^4$. By LOTP we have

$$\mathsf{P}\{B\} = \mathsf{P}\{A\}\,\mathsf{P}\{B \mid A\} + \mathsf{P}\{A^c\}\,\mathsf{P}\{B \mid A^c\} = 0.9 \cdot 5 + 0.1 \cdot 8 = 0.09721.$$

5. By Bayes' rule $\mathsf{P}\{A \mid B\} = \frac{\mathsf{P}\{B \mid A\}\mathsf{P}\{A\}}{\mathsf{P}\{B\}} \approx 0.57864$.
6. Note that $\mathsf{E}[X \mid A, B] = 4 + 6 \cdot 0.5 = 7$ and $\mathsf{E}[X \mid A^c, B] = 4 + 6 \cdot 0.8 = 8.8$. By LOTP with extra conditioning we have

$$\mathsf{P}\{X \mid B\} = \mathsf{P}\{A \mid B\}\,\mathsf{E}[X \mid A, B] + \mathsf{P}\{A^c \mid B\}\,\mathsf{E}[X \mid A^c, B] \approx 7.75844.$$

7. By LOTE we have $\mathsf{P}\{B\}\,\mathsf{E}[X \mid B] + \mathsf{P}\{B^c\}\,\mathsf{E}[X \mid B^c] = \mathsf{E}[X] = 5.3$. We know $\mathsf{P}\{B\}$ and $\mathsf{E}[X \mid B]$, so solving this for $\mathsf{E}[X \mid B^c]$ yields $\mathsf{E}[X \mid B^c] \approx 5.035$.

One or more students wrote the LOTE as $\mathsf{E}[X] = \sum_Y \mathsf{E}[X \mid Y]\,\mathsf{P}\{Y\}$. This is wrong, as you cannot sum over a rv. This is correct: $\mathsf{E}[X] = \sum_y \mathsf{E}[X \mid Y = y]\,\mathsf{P}\{Y = y\}$, so sum over the *outcomes* of a rv.

Ex 9.19. Consider random variables $X, Y \in [0,1]^2$ with joint PDF $f_{X,Y}(x,y) = 2\,I_{x \le y}$.
Determine $\mathsf{E}[Y \mid X]$ and $\mathsf{E}[X \mid Y]$.

s.9.19. The marginal density of X is given by $f_X(x) = 2(1-x)$.
So the conditional density is given by $f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{I_{x \le y}}{1-x}$. Hence,

$$\mathsf{E}[Y \mid X = x] = \int_0^1 y \frac{I_{x \le y}}{1-x}\,\mathrm{d}y = \frac{1}{1-x}\int_x^1 y\,\mathrm{d}y = \frac{1}{1-x}\left[\tfrac{1}{2}y^2\right]_x^1 = \frac{\frac{1}{2}\left(1-x^2\right)}{1-x} = \tfrac{1}{2}(1+x).$$

We conclude that $\mathsf{E}[Y \mid X] = \frac{1}{2}(1+X)$.
The marginal density of Y is given by $f_Y(y) = 2y$.
So the conditional density is given by $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{I_{x \le y}}{y}$. So

$$\mathsf{E}[X \mid Y = y] = \int_0^1 x \frac{I_{x \le y}}{y}\,\mathrm{d}x = \frac{1}{y}\int_0^y x\,\mathrm{d}x = \tfrac{1}{2}y.$$

We conclude that $\mathsf{E}[X \mid Y] = \frac{1}{2}Y$.

Some students wrote for instance $\mathsf{E}[X \mid Y] = y/2$. Apparently, such students are not aware of the idea that $\mathsf{E}[X \mid N]$ is a random variable. When this happens during the exam, you will score 0 points for that particular part of a question.

Ex 9.20. Prove that $\mathsf{E}[X \mid X \ge a] > \mathsf{E}[X]$ for any a with $0 < \mathsf{P}\{X \ge a\} < 1$.

s.9.20. Note that $\mathsf{E}[X \mid X \ge a] \ge a > \mathsf{E}[X \mid X < a]$. By LOTE:

$$\begin{aligned}
\mathsf{E}[X] &= \mathsf{P}\{X \ge a\}\,\mathsf{E}[X \mid X \ge a] + \mathsf{P}\{X < a\}\,\mathsf{E}[X \mid X < a] \\
&< \mathsf{P}\{X \ge a\}\,\mathsf{E}[X \mid X \ge a] + \mathsf{P}\{X < a\}\,\mathsf{E}[X \mid X \ge a] \\
&= \mathsf{E}[X \mid X \ge a],
\end{aligned}$$

where the inequality is strict since $\mathsf{P}\{X < a\} > 0$.

Ex 9.21. Let $N \sim \text{Pois}(\lambda)$ and let $X|N \sim \text{Bin}(N, p)$, where $p \in (0,1)$ and $\lambda > 0$ are known constants. Find $\mathsf{E}[N \mid X]$.

h.9.21. For a smart argument, use the chicken-egg story. Recall that the number of hatched eggs and the number of unhatched eggs are independent (since $N \sim \text{Pois}(\lambda)$); i.e. $N - X$ and X are independent.

s.9.21. With the hint,

$$\mathsf{E}[N \mid X] = \mathsf{E}[N - X \mid X] + \mathsf{E}[X \mid X] = \mathsf{E}[N - X] + X = \lambda(1-p) + X.$$

As a check, $\mathsf{E}[\mathsf{E}[N \mid X]] = \mathsf{E}\left[\lambda(1-p) + X\right] = \lambda(1-p) + \lambda p = \lambda = \mathsf{E}[N]$.

Here is straightforward computation. You should check each and every step as they are based on pattern recognition.

$$\begin{aligned}
\mathsf{E}[N \mid X = k] &= \sum_{n=k}^{\infty} n \mathsf{P}\{N = n \mid X = k\} && (9.4.1)\\
&= \frac{1}{\mathsf{P}\{X = k\}} \sum_{n=k}^{\infty} n e^{-\lambda} \frac{\lambda^n}{n!} \binom{n}{k} p^k (1-p)^{n-k} && (9.4.2)\\
&= \frac{1}{\mathsf{P}\{X = k\}} \sum_{n=k}^{\infty} n e^{-\lambda} \frac{1}{n!} \frac{n!}{k!(n-k)!} (\lambda p)^k (\lambda(1-p))^{n-k} && (9.4.3)\\
&= \frac{e^{-\lambda p} (\lambda p)^k / k!}{\mathsf{P}\{X = k\}} \sum_{n=k}^{\infty} n e^{-\lambda(1-p)} \frac{1}{(n-k)!} (\lambda(1-p))^{n-k} && (9.4.4)\\
&= \sum_{n=k}^{\infty} n e^{-\lambda(1-p)} \frac{1}{(n-k)!} (\lambda(1-p))^{n-k} && (9.4.5)\\
&= \sum_{n=0}^{\infty} (n+k) e^{-\lambda(1-p)} \frac{1}{n!} (\lambda(1-p))^{n} && (9.4.6)\\
&= k + \sum_{n=0}^{\infty} n e^{-\lambda(1-p)} \frac{1}{n!} (\lambda(1-p))^{n} && (9.4.7)\\
&= k + \lambda(1-p). && (9.4.8)
\end{aligned}$$

Hence, $\mathsf{E}[N \mid X] = \lambda(1-p) + X$. Since $\mathsf{E}[X] = \lambda p$, we get $\mathsf{E}[N] = \lambda$ with Adam's law, as above.

9.5 SECTION 9.5

Ex 9.22. BH.9.5.1. Is $\mathsf{V}[Y|X] = \mathsf{V}[\mathsf{E}[Y|X]]$?

h.9.22.

s.9.22.

Ex 9.23. Use Eve's law to show that $\mathsf{V}[Y] \geq \mathsf{V}[\mathsf{E}[Y\,|\,X]]$.

s.9.23. By Eve's law,

$$\mathsf{V}[Y] = \mathsf{E}[\mathsf{V}[Y\,|\,X]] + \mathsf{V}[\mathsf{E}[Y\,|\,X]] \geq \mathsf{V}[\mathsf{E}[Y\,|\,X]], \tag{9.5.1}$$

since $\mathsf{V}[Y\,|\,X] \geq 0$ for all X, which implies that $\mathsf{E}[\mathsf{V}[Y\,|\,X]] \geq 0$.

Ex 9.24. Let $Z \sim \mathcal{N}(\mu, \sigma^2)$ and $Y = \sqrt{Z} + Z^2$. Find $\mathsf{V}[Y|Z]$.

s.9.24. Conditional on Z, Y is a constant, and the variance of a constant is 0. Hence, $\mathsf{V}[Y|Z] = 0$.

Ex 9.25. Correct? $\mathsf{V}[Y] = \mathsf{V}[Y|A]\,\mathsf{P}\{A\} + \mathsf{V}[Y|A^c]\,\mathsf{P}\left\{A^c\right\}$ for any random variable Y and event A.

s.9.25. Incorrect. Counterexample: Let $Y \sim \text{Bern}(1/2)$ and A be the event $Y = 0$. then $\text{Var}(Y|A)$ and $\text{Var}(Y|A^c)$ are both 0, but $\text{Var}(Y)$=1/4.

Ex 9.26. Let X, Y be random variables. Explain the difference between $\mathsf{V}[Y|X]$ and $\mathsf{V}[Y|X = x]$.

s.9.26. $\mathsf{V}[Y|X]$ is a random variable, but $\mathsf{V}[Y|X = x]$ is a constant.

Ex 9.27. Show that $\mathsf{E}\left[(Y - \mathsf{E}[Y|X])^2|X\right] = \mathsf{E}\left[Y^2|X\right] - (\mathsf{E}[Y|X])^2$.

s.9.27. Define $g(X) = \mathsf{E}[Y|X]$. Then,

$$\begin{aligned}
\mathsf{E}\left[(Y - \mathsf{E}[Y|X])^2|X\right] &= \mathsf{E}\left[(Y - g(X))^2|X\right] && (9.5.2)\\
&= \mathsf{E}\left[Y^2 - 2Yg(X) + g(X)^2|X\right] && (9.5.3)\\
&= \mathsf{E}\left[Y^2|X\right] - 2\,\mathsf{E}\left[Yg(X)|X\right] + \mathsf{E}\left[g(X)^2|X\right] && (9.5.4)\\
&= \mathsf{E}\left[Y^2|X\right] - 2g(X)\,\mathsf{E}[Y|X] + g(X)^2 && (9.5.5)\\
&= \mathsf{E}\left[Y^2|X\right] - 2g(X)^2 + g(X)^2 && (9.5.6)\\
&= \mathsf{E}\left[Y^2|X\right] - (\mathsf{E}[Y|X])^2 && (9.5.7)
\end{aligned}$$

Ex 9.28. Let $X \sim \mathcal{N}(\mu, \sigma^2)$ and $W|X \sim \mathcal{N}(0, X^2)$. Find $\mathsf{V}[W]$.

s.9.28. Using Eve's Law we have

$$\mathsf{V}[W] = \mathsf{V}[\mathsf{E}[W|X]] + \mathsf{E}[\mathsf{V}[W|X]] = \mathsf{V}[0] + \mathsf{E}\left[X^2\right] = 0 + \mu^2 + \sigma^2 = \mu^2 + \sigma^2. \tag{9.5.8}$$

10 CHAPTER 10: EXERCISES AND REMARKS

10.1 SECTION 10.1

Ex 10.1. On BH.10.1.1: here is perhaps simpler proof of the Cauchy-Schwarz inequality. Define $f(t) = \mathsf{E}\left[(Y - tX)^2\right]$.

1. Explain that $f(t) \geq 0$.
2. Write $f(t) = \mathsf{E}\left[(Y - tX)^2\right]$ as a polynomial of the second degree, i.e., in the form $f(t) = at^2 + bt + c$ (Hint, see the proof of BH.10.1.1).
3. Since $f(t) \geq 0$, how many (real) roots can it have at most?
4. What are the implications of this for the discriminant $D = b^2 - 4ac$?
5. Show that the Cauchy-Schwarz inequality directly follows from this restriction on D.

s.10.1. f is the expectation of something non-negative. Then, work out the square and apply linearity of the expectation. As $f \geq 0$, it can have most one root, hence $D \leq 0$. But $D = 4\,\mathsf{E}[XY] - 4\,\mathsf{E}\left[X^2\right]\mathsf{E}\left[Y^2\right]$.

Remark 10.2. I find it easier to remember the Cauchy-Schwarz inequality in the form $(\mathsf{E}[XY])^2 \leq \mathsf{E}\left[X^2\right]\mathsf{E}\left[Y^2\right]$; like this there are squares on both sides.

Ex 10.3. On BH.10.1.3. How do they get from $\mathsf{P}\{X > 0\}$ to the inequality for $\mathsf{P}\{X = 0\}$? (Provide the details.)

h.10.3.

s.10.3.

Ex 10.4. On BH.10.1.3. Do the algebra to show that $\mathsf{P}\{X = 0\} = 1/(\mu + 1)$.

h.10.4.

s.10.4.

Ex 10.5. On BH.10.1.3. Explain that we actually use Markov's inequality.

h.10.5.

s.10.5.

Ex 10.6. On BH.10.1.3. What is the probability of two people with birthdays 2 days apart?

h.10.6.

s.10.6.

Remark 10.7. I often forget the direction in Jensen's inequality. To check, the following reasoning works for me: I know that $\mathsf{V}[X] \geq 0$, but $\mathsf{V}[X] = \mathsf{E}\left[X^2\right] - (\mathsf{E}[X])^2 = \mathsf{E}\left[g(X)\right] - g(\mathsf{E}[X])$ with $g(x) = x^2$. Then, from the graph of the parabola, i.e., the graph of g, I know that g is convex.

Ex 10.8. In Jensen's inequality, when does equality hold? Can you explain (in terms of convexity and concavity) why equality holds for only this type of functions?

s.10.8. Equality holds for functions that are both convex and concave. The only functions that are both convex and concave are affine functions, i.e., functions of the type $g(x) = ax + b$. Assuming that g is twice differentiable, we can show this as follows. Convexity is equivalent to $g''(x) \geq 0$ and concavity is equivalent to $g''(x) \leq 0$. This means $g''(x) = 0$ and the only functions for which this holds are affine functions.

If you like maths, consider generalizing the condition. Is it necessary to assume that g is twice differentiable? For instance, it is not hard to prove that a convex function is continuous. Consider now a point at which g is convex and concave at the same time, does it follow that g is twice differentiable at such a point?

Remark 10.9. Skip BH.10.1.7, 10.1.8, 10.1.9

Ex 10.10. When X is a non-negative rv, prove the simplest form of Markov's inequality: $\mathsf{P}\{X \geq a\} \leq \mathsf{E}[X]/a$ for $a \geq 0$. Then show that BH.10.1.10 follows from this.

h.10.10. Use that $X \geq X I_{X \geq a} \geq a I_{X \geq a}$.

s.10.10. In the equation of the hint, take expectations at both sides. Realize that $\mathsf{E}[I_{X \geq a}] = \mathsf{P}\{X \geq a\}$. Next, for any rv, $|X| \geq 0$. Hence, we can apply the simple form of Markov's inequality to get the result of the book.

Ex 10.11. Which of the following are equivalent to Chebyshev's inequality? Show why or why not.

1. $P(|X - \mathsf{E}[X]| \geq a) \leq \frac{\mathsf{V}[X]}{a^2}$ for all $a > 0$
2. $P(|X - \mathsf{E}[X]| < a) > \frac{\mathsf{V}[X]}{a^2}$ for all $a > 0$
3. $P(|X - \mathsf{E}[X]| < a) \geq 1 - \frac{\mathsf{V}[X]}{a^2}$ for all $a > 0$
4. $P(|X - \mathsf{E}[X]| \geq c\sigma) \leq \frac{1}{c^2}$ for all $c > 0$ and $\sigma^2 = \mathsf{V}[X]$.

s.10.11. By definition, equation (1) is Chebyshev's inequality. Letting $a = c\sigma_X$ we get (4). Equation (3) follows from multiplying (1) by -1, adding 1 and using the complement rule. Equation (2) is not equivalent to any of the others, as this is not how reversing inequalities works.

Ex 10.12. On BH.10.1.11. Why is Chebyshev's inequality of no use if we try to plug in values for $0 < a \leq \sqrt{\mathsf{V}[X]}$?

s.10.12. This will result in the trivial bound $\mathsf{P}\{|X-\mu| \geq a\} \leq B$, for some $B \geq 1$. But we already know that every probability is at most one. So the bound does not tell us anything interesting.

Ex 10.13. BH.10.1.13 shows that Chernoff's inequality is a very strict bound. Is Chernoff's inequality always the tightest bound (out of the ones you know)? What about the case where X is defined as follows

$$\mathsf{P}\{X = 0\} = \frac{3}{4}, \qquad \mathsf{P}\{X = 2\} = \frac{1}{4}.$$

h.10.13. Consider $\mathsf{P}\{X \geq 2\}$ and compare Chernoff's inequality to Markov's inequality.

s.10.13. In this (pathological) example we get from Markov's inequality that $P(X \geq 2) \leq \frac{E(X)}{2} = \frac{1}{4}$. This means the Markov bound is tight, as it is equal to the probability that X exceeds 2. From Chernoff's bound we get

$$P(X \geq 2) \leq \frac{E(e^{tX})}{e^{2t}} = \frac{3+e^{2t}}{4e^{2t}} = \frac{1}{4}\left(1 + \frac{3}{e^{2t}}\right) > \frac{1}{4} \quad \forall t > 0.$$

Hence here the Markov bound is tighter. We use the facts from probability theory that $E(X) = \frac{1}{2}$ and that $E(e^{tX}) = \frac{3}{4} + e^{2t}\frac{1}{4}$ in this example.

Ex 10.14. Here is inequality from which all inequalities in BH 10.1.3 immediately follow. It's worth memorizing. Take any rv X and a function f that is non-negative and non-decreasing.

1. Why is this true for any a: $f(a)\, I_{X\geq a} \leq f(X)\, I_{X \geq a} \leq f(X)$?
2. Take expectations in the inequality of the previous step and use the fundamental bridge to show that $\mathsf{P}\{X \geq a\} \leq \mathsf{E}\left[f(X)\right]/f(a)$.
3. What part of the proof goes wrong if f can also be negative?
4. Show that Markov's inequality follows by taking $Y = |X|$ and $f(x) = x$. Why don't we take $f(x) = |x|$?
5. Show that Chebyshev's inequality follows by taking $Y = |X-\mu|$ and $f(x) = x^2$.
6. Show that Chernoff's inequality follows by taking $f(x) = e^x$.

10.2 SECTION 10.2

Ex 10.15. Is the following statement equivalent to the strong or the weak law of large numbers? Fix $\epsilon > 0$. For all $\delta > 0$, there is an n so large that $\mathsf{P}\left\{|\bar{X}_n - \mu| > \epsilon\right\} \leq \delta$.

h.10.15.

s.10.15.

Ex 10.16. Which of the two following statements correctly represents the strong law:

- $\lim_{n\to\infty} \mathsf{P}\left\{|X_n - \mu| > \epsilon\right\} = 1$ for all $\epsilon > 0$,
- $\mathsf{P}\left\{\lim_{n\to\infty} |X_n - \mu| = 0\right\} = 1$.

s.10.16. Check the book. It's the second option.

Ex 10.17. On BH.10.2.5. Where have we applied this idea earlier?

10.3 SECTION 10.3

Ex 10.18. On BH.10.3.1. I prefer to write $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$. Why could that be?

s.10.18. Divide by the std corresponds to the standard transformation $(X-\mu)/\sigma$. Like this, I don't have to remember anything new. Algebra gives the formula of the book.

Ex 10.19. On BH.10.3.7. $\mathsf{E}\left[\log Y_n\right] = \log 100 - 0.081n$. Explain the 0.081. Think also about the paradoxical outcome. (Once again, probability theory *is* hard.)

h.10.19.

s.10.19.

Ex 10.20. On BH.10.3.7. The stock rises α % and decreases β %. Find a relation between α, β such that $\lim Y_n \geq 1$.

h.10.20.

s.10.20.

Ex 10.21. On BH.10.3.7. Note that $\mathsf{E}\left[\log Y_n\right] \sim -0.081n$, i.e., has has negative drift, but $\log \mathsf{E}\left[Y_n\right] \sim n \log 1.1$. Check that this is not in conflict with Jensen's inequality.

h.10.21.

s.10.21.

10.4 SECTION 10.4

Ex 10.22. On BH.10.4.3. Why is $n\bar{Z}_n^2 \sim \chi_1^2$?

h.10.22.

s.10.22. Here is the reason. $\bar{Z}_n$ is the sum of n normal rvs Z_j, hence normal itself. As each of these Z_j is standard normal, $\mathsf{E}\left[\bar{Z}_n\right] = 0$, and $\mathsf{V}\left[\bar{Z}_n\right] = n^{-2}\sum_j \mathsf{V}\left[Z_j\right] = 1/n$, by independence. Therefore, $\sqrt{n}\bar{Z}_n \sim N(0,1) \implies (\sqrt{n}\bar{Z}_n)^2 \sim \chi_1^2$, where we use Definition 10.4.1 and Theorem 10.4.2 in the last step.

Ex 10.23. On BH.10.4.3. Show that $\sum_j^n (Z_j - \bar{Z}_n)^2$ and $\bar{Z}_n^2$ are independent.

h.10.23.

s.10.23.

Ex 10.24. A fair coin is tossed 100 times. We are interested in the probability that the number of heads that turn up is at most 40. What is the tightest upper bound on this probability that we can find by using Chebyshev's inequality? Hint: use a symmetry argument.

s.10.24. Let X be the r.v. corresponding to the number of heads. Then $X \sim \text{Binomial}\left(100, \frac{1}{2}\right)$, which has moments $\mathsf{E}[X] = 100 \cdot \frac{1}{2} = 50$ and $\mathsf{V}[X] = 100 \cdot \frac{1}{2} \cdot \frac{1}{2} = 25$. By symmetry of the $\text{Binomial}\left(100, \frac{1}{2}\right)$ distribution,

$$\mathsf{P}\{X \leq 40\} = \mathsf{P}\{X \geq 60\}. \tag{10.4.1}$$

Hence, using Chebyshev's inequality,

$$\begin{aligned} \mathsf{P}\{X \leq 40\} &= \frac{1}{2}\mathsf{P}\{|X - 50| \geq 10\} && (10.4.2) \\ &= \frac{1}{2}\mathsf{P}\{|X - 50| \geq 10\} && (10.4.3) \\ &\leq \frac{1}{2}\frac{\mathsf{V}[X]}{10^2} && (10.4.4) \\ &= \frac{1}{2}\frac{25}{100} = \frac{1}{8}. && (10.4.5) \end{aligned}$$

Hence,

$$\mathsf{P}\{X \leq 40\} \leq \frac{1}{8}. \tag{10.4.6}$$

Ex 10.25. Let the set of r.v.s $\{X_k, k \geq 1\}$ be the outcomes of throws of a biased coin. We take $X_j = 1$ if the outcome is heads, and $X_j = 0$ if tails. Suppose $\mathsf{E}[X_k] = \mu$ and $\mathsf{V}[X_k] = \sigma^2$. Let $Y_j = \sum_{i=nj+1}^{(n+1)j} X_i/n$, i.e., Y_j is the sample mean of the j batch of throws. Since $\{Y_j, j \geq 1\}$ are iid, take Y as the common r.v., i.e., $Y_j \sim Y$. What is a (frequentist) explanation of the statement $\mathsf{P}\left\{|Y - \mu| > \epsilon\right\} \leq \sigma^2/n\epsilon$?

s.10.25. We assemble m observations of Y_j (hence, we throw the coin nm times). Suppose we see M times that $|Y_j - \mu| > \epsilon$. Then we expect that $M/m < \sigma^2/n\epsilon$.

Thus, Chebyshev's inequality makes a statement about sample means of size n, say.

Ex 10.26. Interpret the WLLN in terms of the previous exercise.

s.10.26. First fix some $\epsilon > 0$. Now take some n and determine the fraction of outliers, that is, count how many of the sample means $Y_1 = \sum_{i=1}^{n} X_i/n$, $Y_2 = \sum_{i=n+1}^{2n} X_i/n, \ldots$ lie outside the interval $[\mu-\epsilon, \mu+\epsilon]$ and divide by the number of samples taken. The WLLN says this: If the sample averages Y_1, Y_2 are taken over larger sets of the X_j, i.e., n is larger so that we put more throws in a batch, then the fraction of outliers become smaller.

Ex 10.27. In the setting of [10.25], the probability of a sequence of outcomes like this: $H, T, H, T, H, T, \ldots$, i.e., a sequence in which the heads and tails alternate, has probability zero. However, $\sum_{i=1}^{n} I_{X_i=H}/n \to 1/2$. So, we have a sequence that occurs with probability zero, but still the average along the sequence has the proper limit. Doesn't this violate the SLLN?

s.10.27. The SLLN says nothing about individual sample paths, i.e., strings of outcomes like $H, T, H, T, \ldots$. In fact, the probability of obtaining any particular sample path has zero probability. Instead, the SLLN makes a statement about sets of sample paths. For the coin it says that it is virtually impossible to pick a path from the set of paths whose long-run fraction of heads is not equal to $1/2$.

11 OLD EXAM QUESTIONS

QUESTION

Ex 11.1. Inspired by BH.7.3.6, compute $\mathsf{V}[M]$ and $\mathsf{V}[L]$ by first computing $\mathsf{E}\left[M^2\right]$ and $\mathsf{E}\left[L^2\right]$.

h.11.1. By the previous exercise, realize that $L \sim \text{Exp}(2\lambda)$. Use these properties.

s.11.1. I *have remembered* that $\mathsf{V}[X] = \mathsf{E}[X]$ when $X \sim \text{Exp}(\lambda)$. Since $\mathsf{V}[X] = \mathsf{E}\left[X^2\right] - (\mathsf{E}[X])^2$, $\mathsf{E}\left[X^2\right] = 2/\lambda^2$. Applying this to L, we see that $\mathsf{E}\left[L^2\right] = 2/(2\lambda)^2 = 1/2\lambda^2$. Moreover, $\mathsf{V}[L] = 1/4\lambda^2$.

Next, $f_M(x) = 2f_x(x)F_Y(x)$. Hence,

$$\mathsf{E}\left[M^2\right] = \int x^2 2\lambda e^{-\lambda x}(1 - e^{-\lambda x} \,\mathrm{d}x = 2\int x^2\lambda e^{-\lambda x}\,\mathrm{d}x - \int x^2 2\lambda e^{-2\lambda x}\,\mathrm{d}x.$$

The first integral is just 2 times $\mathsf{E}\left[X^2\right]$, the second is $\mathsf{E}\left[L^2\right]$. Hence, $\mathsf{E}\left[M^2\right] = 4/\lambda^2 - 1/2\lambda^2 = 7/2\lambda^2$. Finally, $\mathsf{V}[M] = 7/2\lambda^2 - 9/4\lambda^2 = 5/4\lambda^2$.

QUESTION

Ex 11.2. Let (X, Y) follow a Bivariate Normal distribution, with X and Y marginally following $\mathcal{N}(\mu, \sigma^2)$ and with correlation ρ between X and Y.

1. Use the definition of a Multivariate Normal Distribution to show that $(X+Y, X-Y)$ is also Bivariate Normal.
2. Find the marginal distributions of $X+Y$ and $X-Y$.
3. Compute $\mathsf{Cov}[X+Y, X-Y]$. Then, write down the expression for the joint PDF of $(X+Y, X-Y)$.

s.11.2.

1. Since (X, Y) are bivariate normally distributed, every linear combination of X and Y is normally distributed. Note that every linear combination of $(X+Y)$ and $(X-Y)$ can be written as a linear combination of X and Y. Hence, every linear combination of $(X+Y)$ and $(X-Y)$ is normally distributed. Hence, $(X+Y, X-Y)$ is bivariate normally distributed.
2. By the story above, both X and Y are normally distributed. We have
$$\mathsf{E}[X+Y] = \mathsf{E}[X] + \mathsf{E}[Y] = \mu + \mu = 2\mu, \tag{11.0.1}$$
and
$$\mathsf{E}[X-Y] = \mathsf{E}[X] - \mathsf{E}[Y] = \mu - \mu = 0. \tag{11.0.2}$$
Moreover,
$$\mathsf{V}[X+Y] = \mathsf{V}[X] + \mathsf{V}[Y] + 2\,\mathsf{Cov}[X, Y] = 2\sigma^2 + 2\rho\sigma^2 = 2(1+\rho)\sigma^2. \tag{11.0.3}$$
Simlarly,
$$\begin{aligned} \mathsf{V}[X-Y] = \mathsf{V}[X] + \mathsf{V}[-Y] + 2\,\mathsf{Cov}[X, -Y] &= \mathsf{V}[X] + \mathsf{V}[Y] - 2\,\mathsf{Cov}[X, Y] && (11.0.4)\\ = 2\sigma^2 - 2\rho\sigma^2 = 2(1-\rho)\sigma^2. & && (11.0.5) \end{aligned}$$
So we have found that $X+Y \sim N(2\mu, 2(1+\rho)\sigma^2)$ and $X-Y \sim N(0, 2(1-\rho)\sigma^2$.
3. We have
$$\begin{aligned} \mathsf{Cov}[X+Y, X-Y] &= \mathsf{Cov}[X, X] - \mathsf{Cov}[X, Y] + \mathsf{Cov}[Y, X] - \mathsf{Cov}[Y, Y] && (11.0.6)\\ &= \mathsf{V}[X] - \mathsf{V}[Y] = \sigma^2 - \sigma^2 = 0. && (11.0.7) \end{aligned}$$
Write $U = X+Y$, $V = X-Y$. Plugging all the parameters into the formula for the joint pdf of a bivariate normal distribution (see https://en.wikipedia.org/wiki/Multivariate_normal_distribution#Bivariate_case), we obtain
$$f_{U,V}(u, v) = \frac{1}{2\pi\sqrt{2(1+\rho)\sigma^2 2(1-\rho)\sigma^2}} \exp\left(-\frac{1}{2}\left[\frac{(u-2\mu)^2}{2(1+\rho)\sigma^2} + \frac{v^2}{2(1-\rho)\sigma^2}\right]\right). \tag{11.0.8}$$

QUESTION

Ex 11.3. This is about the simplest model for an insurance company that I can think of. We start with an initial capital $I_0 = 2$. The company receives claims and contributions every period, a week say. In the ith period, we receive a contribution X_i uniform on the set $\{1, 2, \ldots, 10\}$ and a claim C_i uniform on $\{0, 1, \ldots 8\}$.

1. What is the meaning of $I_1 = I_0 + X_1 - C_1$?
2. What is the meaning of $I_2 = I_1 + X_2 - C_2$?
3. What is the interpretation of $I'_1 = \max\{I_0 - C_1, 0\} + X_1$?
4. What is the interpretation of $I'_2 = \max\{I'_1 - C_2, 0\} + X_2$?
5. What is the interpretation of $\bar{I}_n = \min\{I_i : 0 \leq i \leq n\}$?
6. What is $\mathsf{P}\{I_1 < 0\}$?
7. What is $\mathsf{P}\left\{I'_1 < 0\right\}$?
8. What is $\mathsf{P}\{I_2 < 0\}$?
9. What is $\mathsf{P}\left\{I'_2 < 0\right\}$?
10. Provide an interpretation in terms of the inventory of rice, say, at a supermarket for I_1 and I'_1.

s.11.3. This question tests your modeling skills too.

In hindsight, the questions have to reorganized a bit. The capital at the end of the ith week is $I_i = I_{i-1} + X_i - C_i$.

Suppose claims arrive at the beginning of the week, and contributions arrive at the end of the week (people prefer to send in their claims early, but they prefer to pay their contribution as late as possible). If we don't have sufficient money in cash, then we cannot pay a claim. Thus, $\max\{I_0 - C_1\}$ is our capital just before the contribution arrives. Hence, I'_1 is our capital at the end of week 1 under the assumption that we never pay out more than we have in cash. Likewise for I'_2

$\bar{I}_n$ is the lowest capital we have seen for the first n weeks.

In the supermarket setting, I_i is our inventory is we can be temporarily out of stock, but as soon as new deliveries—so called replenishments—arrive then we serve the waiting customers immediately. The model with I' corresponds to a setting is which we consider unmet demand as lost.

$$\mathsf{P}\{I_0 <= 0\} = \mathsf{P}\{2 + X_1 - C_1 < 0\} = \frac{1}{10}\sum_{i=1}^{10} \mathsf{P}\{C_1 > 2 + i\} = \frac{1}{10}\sum_{i=1}^{5} \mathsf{P}\{C_1 > 2 + i\} \tag{11.0.9}$$

$$= \frac{1}{10}\sum_{i=1}^{5} \frac{6-i}{9}. \tag{11.0.10}$$

When grading, I realized that questions 8 was not quite reasonable to ask as an exam question. We graded this leniently. As I find it too boring to compute these probabilities by hand, here is the python code. The ideas in the code are highly interesting and useful. The main data structure here is a dictionary, one of the most used data structure in python. I don't have the R code yet, so if you take the (unwise) decision to stick to only R, you have to wait a bit until somebody sends me the R code for this problem.

Python Code

```
C = {}
for i in range(0, 9):
    C[i] = 1 / 9

X = {}
for i in range(1, 11):
    X[i] = 1 / 10


I0 = 2

I1 = {}
for k, p in X.items():
    for l, q in C.items():
        i = I0 + k - l
        I1[i] = I1.get(i, 0) + p * q

print("I1, ", sum(I1.values()))  # check


# compute P(I1<0):
P = sum(r for i, r in I1.items() if i < 0)
print(P)


I2 = {}
for i, r in I1.items():
    for k, p in X.items():
        for l, q in C.items():
            j = i + k - l
            I2[j] = I2.get(j, 0) + r * p * q

print("I2 ", sum(I2.values()))  # just a check

# compute P(I2<0):
P = sum(r for i, r in I2.items() if i < 0)
```

Interestingly, $I'_i \geq 1$. (This is so simple to see that I first did it wrong.)

Mistake: note that X_i and C_i are discrete rvs, not continuous. The sum of two uniform random variables is not uniform. For example, think of the sum of two die throws. Is getting 2 just as likely as getting 7?

QUESTION

Ex 11.4. Take $X \sim \text{Unif}(\{-2,-1,1,2\})$ and $Y = X^2$. What is the correlation coefficient of X and Y? If we would consider another distribution for X, would that change the correlation?

s.11.4. We have

$$\mathsf{Cov}[X,Y] = \mathsf{Cov}\left[X,X^2\right] = \mathsf{E}\left[XX^2\right] - \mathsf{E}[X]\,\mathsf{E}\left[X^2\right] = 0 - 0\cdot 2.5 = 0. \tag{11.0.11}$$

Hence, $\text{Corr}(X,Y) = 0$.
Yes, for instance, take $X \sim \text{Unif}(\{0,1\})$. Then,

$$\mathsf{Cov}[X,Y] = \mathsf{E}\left[XX^2\right] - \mathsf{E}[X]\,\mathsf{E}\left[X^2\right] = 0.5 - 0.5\cdot 0.5 = 0.25. \tag{11.0.12}$$

QUESTION

Ex 11.5. We have a machine that consists of two components. The machine works as long as both components have not failed (in other words, the machine fails when one of the two components fails). Let X_i be the lifetime of component i.

1. What is the interpretation of $\min\{X_1, X_2\}$?
2. If X_1, X_2 iid $\sim$ Exp(10) (in hours), what is the probability that the machine is still 'up' (i.e., not failed) at time T?
3. Use the previous result to determine the distribution of $\min\{X_1, X_2\}$.
4. What is the expected time until the machine fails?

s.11.5.

1. The interpretation is: the time until the first component fails. That is, the time until the machine stops working.
2. Let $\lambda = 10$. We have

$$\begin{aligned}
\mathsf{P}\{\text{machine not failed at time } T\} &= \mathsf{P}\{\min\{X_1, X_2\} > T\} && (11.0.13)\\
&= \mathsf{P}\{X_1 > T, X_2 > T\} && (11.0.14)\\
&= \mathsf{P}\{X_1 > T\}\,\mathsf{P}\{X_2 > T\} && (11.0.15)\\
&= e^{-\lambda T} \cdot e^{-\lambda T} && (11.0.16)\\
&= e^{-(2\lambda)T} && (11.0.17)\\
&= e^{-20T} && (11.0.18)\\
& && (11.0.19)
\end{aligned}$$

3. Note that

$$\mathsf{P}\{\min\{X_1, X_2\} \le T\} = 1 - \mathsf{P}\{\min\{X_1, X_2\} > T\} = 1 - e^{-20T}. \tag{11.0.20}$$

Note that this is the cdf of an exponential distribution with parameter 20. Hence, $\min\{X_1, X_2\} \sim \exp(20)$.

4. The expected time until the machine fails is

$$\mathsf{E}[\min\{X_1, X_2\}] = 1/20, \tag{11.0.21}$$

i.e., 3 minutes. Apparently, the machine is not very robust.

QUESTION

Ex 11.6. We have two rvs X and Y with the joint PDF $f_{X,Y}(x,y) = \frac{6}{7}(x+y)^2$ for $x, y \in (0,1)$ and 0 else. Also we consider the two rvs U and V with the joint PDF $f_{U,V}(u,v) = 2$ for $u, v \in [0,1], u+v \leq 1$ and 0 else.

1. Compute $\mathsf{P}\{X+Y>1\}$.

2. Compute $\mathsf{Cov}[U,V]$.

(Hint: first draw the area over which you want to integrate, if this does not help check out the discussion board post on exercise 7.13a from the first Tutorial)

s.11.6. 1. We have

$$\begin{aligned}
\mathsf{P}\{X+Y>1\} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_{X+Y>1} f_{X,Y}(x,y)\,\mathrm{d}y\,\mathrm{d}x && (11.0.22)\\
&= \int_0^1 \int_{1-x}^1 \frac{6}{7}(x+y)^2\,\mathrm{d}y\,\mathrm{d}x && (11.0.23)\\
&= \frac{6}{7}\int_0^1 \left[\frac{1}{3}(x+y)^3\right]_{y=1-x}^1 \mathrm{d}x && (11.0.24)\\
&= \frac{2}{7}\int_0^1 \Big((x+1)^3-(x+1-x)^3\Big)\,\mathrm{d}x && (11.0.25)\\
&= \frac{2}{7}\int_0^1 \Big((x+1)^3-1\Big)\,\mathrm{d}x && (11.0.26)\\
&= \frac{2}{7}\left[\frac{1}{4}(x+1)^4-x\right]_{x=0}^1 && (11.0.27)\\
&= \frac{1}{14}\left[(x+1)^4-4x\right]_{x=0}^1 && (11.0.28)\\
&= \frac{1}{14}\Big(\big((1+1)^4-4\big)-\big((0+1)^4-0\big)\Big) && (11.0.29)\\
&= \frac{1}{14}\Big(16-4-1\Big) && (11.0.30)\\
&= \frac{11}{14}. && (11.0.31)
\end{aligned}$$

2. We have

$$\mathsf{Cov}[U,V] = \mathsf{E}[UV] - \mathsf{E}[U]\,\mathsf{E}[V]. \qquad (11.0.32)$$

First, we compute

$$\begin{aligned}
\mathsf{E}[UV] &= \int_0^1 \int_0^{1-u} 2uv \, \mathrm{d}v \, \mathrm{d}u && (11.0.33)\\
&= \int_0^1 [uv^2]_{v=0}^{1-u} \, \mathrm{d}u && (11.0.34)\\
&= \int_0^1 \left(u(1-u)^2 - 0\right) \mathrm{d}u && (11.0.35)\\
&= \int_0^1 u(1-2u+u^2) \, \mathrm{d}u && (11.0.36)\\
&= \int_0^1 (u-2u^2+u^3) \, \mathrm{d}u && (11.0.37)\\
&= \left[\frac{1}{2}u^2 - \frac{2}{3}u^3 + \frac{1}{4}u^4\right]_{u=0}^1 && (11.0.38)\\
&= \frac{1}{2} - \frac{2}{3} + \frac{1}{4} && (11.0.39)\\
&= \frac{1}{12}. && (11.0.40)
\end{aligned}$$

Next,

$$\begin{aligned}
\mathsf{E}[U] &= \int_0^1 \int_0^{1-u} 2u \, \mathrm{d}v \, \mathrm{d}u && (11.0.41)\\
&= \int_0^1 2u \int_0^{1-u} 1 \, \mathrm{d}v \, \mathrm{d}u && (11.0.42)\\
&= \int_0^1 2u(1-u) \, \mathrm{d}u && (11.0.43)\\
&= 2\int_0^1 (u-u^2) \, \mathrm{d}u && (11.0.44)\\
&= 2\left[\frac{1}{2}u^2 - \frac{1}{3}u^3\right]_{u=0}^1 && (11.0.45)\\
&= 2\left(\frac{1}{2} - \frac{1}{3}\right) && (11.0.46)\\
&= \frac{1}{3} && (11.0.47)
\end{aligned}$$

By symmetry, $\mathsf{E}[V] = \frac{1}{3}$. Hence,

$$\begin{aligned}
\mathsf{Cov}[U,V] &= \mathsf{E}[UV] - \mathsf{E}[U]\,\mathsf{E}[V] && (11.0.48)\\
&= \frac{1}{12} - \frac{1}{3}\frac{1}{3} && (11.0.49)\\
&= \frac{1}{12} - \frac{1}{9} && (11.0.50)\\
&= -\frac{1}{36}. && (11.0.51)
\end{aligned}$$

QUESTION

Ex 11.7. Let $U = X + Y$ and $V = X - Y$ where $X, Y \sim U[0,1]$ and independent. Show that

$$f_{U,V}(u,v) = \frac{1}{2} I_{|v| \le u \le 2-|v|}.$$

s.11.7. Since $(u, v) = g(x, y) = (x + y, x - y)$,

$$\frac{\partial(u,v)}{\partial(x,y)} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = |-2| = 2.$$

Moreover, $x = (u + v)/2$, $y = (u - v)/2$, so that

$$f_{U,V}(u,v) = f_{X,Y}(x,y)\frac{\partial(x,y)}{\partial(u,v)} = f_{X,Y}(x,y)/2 = f_X(x)f_Y(y)/2 = \frac{1}{2}.$$

The difficulty is in the domain, however. Note that x and y satisfy $0 \le x \le 1$, $0 \le y \le 1$. So $0 \le (u + v)/2 \le 1$ and $0 \le (u - v)/2 \le 1$, which simplifies to $-v \le u \le 2 - v$ and $v \le u \le 2 + v$, which can also be written as $|v| \le u \le 2 - |v|$.

QUESTION

Ex 11.8. Let X and Y have PDF $f_{X,Y}$. Take $g(x, y) = (\min\{x, y\}, \max\{x, y\})$. Why is

$$f_{U,V}(u, v) = f_{X,Y}(u, v) + f_{X,Y}(v, u)?$$

Simplify to $f_{U,V}(u, v) = 2f(u)f(v)$ for the case X, Y iid with common PDF f.

s.11.8. Note that $u(x, y) = \min\{x, y\} = x\, I_{x\leq y} + y\, I_{x>y}$. With a similar expression for v we find for the Jacobian:

$$\frac{\partial(u, v)}{\partial(x, y)} = \begin{bmatrix} I_{x\leq y} & I_{y<x} \\ I_{y<x} & I_{x\leq y} \end{bmatrix} = |\, I_{x\leq y} - I_{x>y}| = 1.$$

If $(U, V) = g(X, Y)$, then $g^{-1}(u, v) = \{(u, v), (v, u)\}$, i.e., a set of two points.

If X, Y iid with PDF, then $f_{X,Y}(x, y) = f(x)f(y)$.

QUESTION

Ex 11.9. Let X, Y be continuous rvs with CDF $F_{X,Y}(x, y) = (x-1)^2(y-2)/8$ for $x \in (1,3)$.

a. Explain that $y \in (2,4)$ for F to be a proper CDF.

b. What is $F(3,7)$?

c. Determine the PDF.

d. Compute $\mathsf{P}\{2 < X < 3\}$

e. Compute $\mathsf{P}\{2 < X < 3, 2 < Y < 3\}$.

f. Compute $\mathsf{P}\{Y < 2X\}$.

g. Compute $\mathsf{P}\{Y \leq 2X\}$.

h. Compute $\mathsf{P}\{Y < 2X, Y + 2X > 6\}$.

s.11.9. a.

$$F \geq 0 \implies 2 < y \tag{11.0.52}$$
$$F \leq 1 \implies F(3, y) \leq 1 \implies F(3,4) = 1 \tag{11.0.53}$$

b. $F(3,7) = 1$.
c. $f(x, y) = \partial_x \partial_y F(x, y) = (x-1)/4$ for $x \in (1,3), y \in (2,4)$ and 0 elsewhere.
d.

$$\mathsf{P}\{2 < X < 3\} = F_X(3) - F_X(2) \tag{11.0.54}$$
$$= F_{X,Y}(3,4) - F_{X,Y}(2,4) = 1 - 1 \cdot 2/8 = 3/4. \tag{11.0.55}$$

e. Make a drawing of the rectangle $[2,3] \times [2,4]$. Then check what parts of this are covered by $F_{X,Y}$.

$$\mathsf{P}\{2 < X < 3, 2 < Y < 3\} = F_{X,Y}(3,3) - F_{X,Y}(2,3) - F_{X,Y}(3,2) + F_{X,Y}(2,2). \tag{11.0.56}$$

The rest is just number plugging.
f. Use the fundamental bridge and c.

$$\mathsf{P}\{Y < 2X\} = \mathsf{E}[I_{Y<2X}] \tag{11.0.57}$$
$$= \iint I_{y<2x} f_{X,Y}(x, y) \,\mathrm{d}x \,\mathrm{d}y \tag{11.0.58}$$
$$= \frac{1}{4} \iint I_{y<2x} I_{2<y<4} I_{1<x<3}(x-1) \,\mathrm{d}x \,\mathrm{d}y \tag{11.0.59}$$
$$= \frac{1}{4} \int_1^3 (x-1) \int I_{2<y<\min\{2x,4\}} \,\mathrm{d}y \,\mathrm{d}x \tag{11.0.60}$$
$$= \frac{1}{4} \int_1^3 (x-1)(\min\{2x,4\} - 2) \,\mathrm{d}x \tag{11.0.61}$$
$$= \frac{1}{4} \int_1^2 (x-1)(2x-2) \,\mathrm{d}x + \frac{1}{4} \int_2^3 (x-1)(4-2) \,\mathrm{d}x. \tag{11.0.62}$$

Finishing the computation must be easy for you now (and if not, practice real hard).

g. As X, Y continuous, the answer is equal to that of f.

h. Similar to f. but a bit more involved.

$$\begin{aligned}
\mathsf{P}\{Y < 2X, Y + 2X > 6\} &= \mathsf{E}\left[I_{Y<2X, Y>6-2X} \right] &(11.0.63)\\
&= \iint I_{y<2x, y>6-2x} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y &(11.0.64)\\
&= \frac{1}{4} \iint I_{y<2x, y>6-2x}\, I_{2<y<4}\, I_{1<x<3} (x-1)\,\mathrm{d}x\,\mathrm{d}y &(11.0.65)\\
&= \frac{1}{4} \int_1^3 (x-1) \int I_{\max\{2,6-2x\}<y<\min\{2x,4\}}\,\mathrm{d}y\,\mathrm{d}x &(11.0.66)\\
&= \frac{1}{4} \int_1^3 (x-1) [\min\{2x,4\} - \max\{2, 6-2x\}]^+\,\mathrm{d}x, &(11.0.67)
\end{aligned}$$

where we need the $[\cdot]^+$ to ensure the positivity of $\min\{2x,4\} - \max\{2,6-2x\}$. To see this, make a graph of the function $\min\{2x,4\} - \max\{2,6-2x\}$. Also, from this graph,

$$= \frac{1}{4} \int_{3/2}^2 (x-1)(2x-6+2x)\,\mathrm{d}x + \frac{1}{4} \int_2^3 (x-1)(4-2)\,\mathrm{d}x. \qquad (11.0.68)$$

The rest is for you.

QUESTION

Ex 11.10. Consider the general case where we are given the relationship $U = V^4$ between the random variables U and V for $V \in (-3,2)$. Explain why we cannot simply invoke the change of variables theorem.

Now imagine V following a uniform distribution on the given interval. Consider the given transformation on the intervals $(-3,0)$ and $(0,2)$ separately. Explain why this allows you to employ the change of variables theorem and find the distribution of U on these intervals. Finally combine these results (using indicator functions) and state the PDF of U (remember to adjust the domain for the indicator functions according to the transformations).

h.11.10. What is the domain of V on each of the intervals $(-3,0)$ and $[0,2)$? For the final part, combining the results into one PDF: Use LOTP, conditioning on $U \geq 0$.

s.11.10. The function $g(x) = x^4$ is not one-to-one on $\mathbb{R}$. It is, however, locally, one-to-one, around the roots of U. (In this course we don't deal with complex numbers, for your interest, it can be proven that the equation $x^4 - y$ has, in general, four roots in the complex plane.)

We need to be bit careful with applying the change of variables formula, but we are OK if we apply it locally around the roots $U^{1/4}$ and $-U^{1/4}$. However, mind that we also should take care of the domain of V, so it might be that these roots don't lie in the domain of V.

With all this, let's first tackle the Jacobian, and then get the domain right with indicators.

$$u = g(v) = v^4 \implies v = \pm u^{1/4}, \tag{11.0.69}$$

$$f_U(u)\,\mathrm{d}u = f_V(v)\,\mathrm{d}v \implies f_U(u) = f_V(v)\frac{\mathrm{d}v}{\mathrm{d}u}, \tag{11.0.70}$$

$$\frac{\mathrm{d}u}{\mathrm{d}v} = 4v^3 = 4u^{3/4}\,I_{v\geq 0} - 4u^{3/4}\,I_{v<0}, \tag{11.0.71}$$

$$f_U(u) = \frac{f_V(-u^{1/4})}{4(-u)^{3/4}}\,I_{-u^{1/4}\in(-3,0)} + \frac{f_V(u^{1/4})}{4(u)^{3/4}}\,I_{u^{1/4}\in[0,2)} \tag{11.0.72}$$

$$= \frac{f_V(-u^{1/4})}{4(-u)^{3/4}}\,I_{u\in(0,81)} + \frac{f_V(u^{1/4})}{4(u)^{3/4}}\,I_{u\in[0,16)}. \tag{11.0.73}$$

If V has the uniform distribution, then $f_V(v) = \frac{1}{5}$ for $v \in (-3,2)$, so

$$f_U(u) = \frac{1}{20(-u)^{3/4}}\,I_{u\in(0,81)} + \frac{1}{20(u)^{3/4}}\,I_{u\in[0,16)}. \tag{11.0.74}$$

QUESTION

Ex 11.11. Let $U \sim \text{Unif}(0,\pi)$. Use BH.8.1.9 to show that $X = \tan(U)$ has the Cauchy distribution. Compare this exercise to BH.8.1.5.

s.11.11. Here is a direct approach.

$$x = \tan u = g(u) \implies u = \arctan x \tag{11.0.75}$$

$$\frac{\mathrm{d}x}{\mathrm{d}u} = \frac{1}{\cos^2 u} = \frac{\sin^2 u + \cos^2 u}{\cos^2 u} = \tan^2 u + 1 = x^2 + 1, \tag{11.0.76}$$

$$f_X(x) = f_U(u)\frac{\mathrm{d}u}{\mathrm{d}x} = \frac{1}{\pi} I_{u\in(0,\pi)} \frac{1}{1+x^2} \tag{11.0.77}$$

$$= \frac{1}{\pi(1+x^2)} I_{\arctan x\in(0,\pi)} = \frac{1}{\pi(1+x^2)}. \tag{11.0.78}$$

In the last equation we just shifted the tan from $(-\pi/2,\pi/2]$ to the interval $(0,\pi)$. The tan has also a proper inverse in $(0,\pi)$ (make a drawing of tan to see this), hence all is well-defined.

QUESTION

You walk into a bar and you find two people, Amy and Bob, playing a game of darts. Their game consists of several rounds, called *legs*, and the first person to win 7 legs wins the match. You have never seen Amy or Bob play before, so you don't know their strength. Denoting by p the probability that Amy wins a leg, your prior belief can be modeled by a uniform distribution: $p \sim \text{Unif}(0,1)$. (Note: we assume that p remains constant during the entire match; even though your *beliefs* about p might change.)

Denoting by A a leg won by Amy and by B a leg won by Bob, the result of the first 10 legs is: $AAABBAABAB$. Your friend Charles is very confident that Amy will win the match and he offers you a bet: if Amy wins the match, you must pay €10 to Charles; if Bob wins the match, he must pay you €300. You are tempted to take the bet, but you want to do some calculations first.

Ex 11.12. Is the order in which Amy and Bob won their respective legs relevant for your posterior probability that Bob will win the match?

s.11.12. No. The only relevant information is the amount of legs won by each player.

Ex 11.13. Let A_n denote the number of legs that Amy won out of a total of n legs. Express the result of the first 10 legs in terms of A_n

s.11.13. Our current information can be represented as: $A_{10} = 6$.

Ex 11.14. What is the distribution of $A_n|p$ (i.e., the distribution of A_n given the value of p)?

s.11.14. We have $A_n \sim \text{Bin}(n,p)$.

Ex 11.15. Find the posterior distribution of p after observing $A_n = k$.

s.11.15. Let f_0 denote the prior distribution of p. Then for the posterior pdf we find by Bayes' theorem:

$$f_1(p|A_n = k) = \frac{\mathsf{P}\{A_n = k\,|p\}\,f_0(p)}{\mathsf{P}\{A_n = k\}} \tag{11.0.79}$$

$$= \frac{\binom{n}{k}p^k(1-p)^{n-k}\cdot 1}{\mathsf{P}\{A_n = k\}} \tag{11.0.80}$$

$$\propto p^k(1-p)^{n-k}, \tag{11.0.81}$$

in which we recognize the pdf of a $\text{Beta}(k+1, n-k+1)$ distribution (up to a normalizing constant). Hence, $p|A_n = k \sim \text{Beta}(k+1, n-k+1)$.

Ex 11.16. According to your posterior belief about p, what is the probability that Bob wins the match? Express your answer in terms of beta functions. (Hint: Use the law of total probability.)

s.11.16. Important: we have already observed 10 legs with an outcome, with which we have updated our belief. Hence, we should use the *posterior* distribution given $A_{10} = 6$ in this exercise! (It's easy to make a mistake here.) Think about it. Suppose instead we had observed

1000 legs and Amy had won 990 of them (i.e., $A_{1000} = 990$). Wouldn't we use this information if someone offered us a bet?

Note that Bob should wins the match if and only if he wins the next three legs. Let W_k be short-hand notation for the event "Bob wins the kth leg". Then, observing that W_{11}, W_{12}, W_{13} are independent, and using the LOTP in the fourth step, we obtain

$$\begin{aligned}
\mathsf{P}\{\text{Bob wins the match} \mid A_{10} = 6\} &= \mathsf{P}\{W_{11} \cap W_{12} \cap W_{13} \mid A_{10} = 6\} && (11.0.82)\\
&= \mathsf{P}\{W_{11} \mid A_{10} = 6\}\,\mathsf{P}\{W_{12} \mid A_{10} = 6\}\,\mathsf{P}\{W_{13} \mid A_{10} = 6\} && (11.0.83)\\
&= \mathsf{P}\{W_{11} \mid A_{10} = 6\}^3 && (11.0.84)\\
&= \int_0^1 \mathsf{P}\left\{I_{11} \mid p, A_{10} = 6\right\}^3 f_1(p|A_{10} = 6)\,dp && (11.0.85)\\
&= \int_0^1 (1-p)^3 \cdot \frac{p^6(1-p)^4}{\beta(7,5)}\,dp && (11.0.86)\\
&= \frac{\beta(7,8)}{\beta(7,5)} \int_0^1 \frac{p^6(1-p)^7}{\beta(7,8)}\,dp && (11.0.87)\\
&= \frac{\beta(7,8)}{\beta(7,5)}. && (11.0.88)
\end{aligned}$$

(Note that we very explicitly do all the steps here. It might be more intuitively clear if you skip the first few steps and write

$$\mathsf{P}\{\text{Bob wins the match} \mid A_{10} = 6\} = \int_0^1 (1-p)^3 f_1(p|A_{10} = 6)\,dp, \tag{11.0.89}$$

and work from there).

Ex 11.17. Using the expression

$$\beta(a,b) = \frac{(a-1)!(b-1)!}{(a+b-1)!} \tag{11.0.90}$$

for every positive integers a, b, compute the probabilty from the previous question as a number.

s.11.17. We have

$$\begin{aligned}
\mathsf{P}\{\text{Bob wins the match} \mid A_{10} = 6\} &= \frac{\beta(7,8)}{\beta(7,5)} && (11.0.91)\\
&= \left(\frac{6!7!}{14!}\right) / \left(\frac{6!4!}{11!}\right) && (11.0.92)\\
&= \frac{7!/4!}{14!/11!} && (11.0.93)\\
&= \frac{7\cdot 6\cdot 5}{14\cdot 13\cdot 12} && (11.0.94)\\
&= 5/52 && (11.0.95)\\
&= 0.0962. && (11.0.96)
\end{aligned}$$

Ex 11.18. Assuming that you want to maximize your expected outcome, should you take the bet?

s.11.18. Our expected profit when taking the bet is

$$
\begin{aligned}
&300 \cdot \mathsf{P}\left\{\text{Bob wins the match} \mid A_{10} = 6\right\} - 10 \cdot \mathsf{P}\left\{\text{Amy wins the match} \mid A_{10} = 6\right\} && (11.0.97)\\
&= 300 \cdot \frac{5}{52} - 10 \cdot \left(1 - \frac{5}{52}\right) && (11.0.98)\\
&= 19.808. && (11.0.99)
\end{aligned}
$$

So we expect to make a profit of €19.81. Hence, you should take the bet.

QUESTION

Ex 11.19. On BH.9.5.4. With Z and W as given in the example, show that $\mathsf{E}[Z|W] = \rho$ and $\mathsf{V}[Z|W] = 1-\rho^2$.

h.11.19.

s.11.19.

Ex 11.20. Prove that

$$\mathsf{E}\left[(Y - \mathsf{E}[Y|X] - h(X))^2\right] = \mathsf{E}\left[(Y - \mathsf{E}[Y|X])^2\right] + \mathsf{E}\left[(h(X))^2\right] \tag{11.0.100}$$

for all random variables X, Y and all functions h. Explain why this result implies that $\mathsf{E}[Y|X]$ is the best predictor of Y based on X.

s.11.20. By the linearity of expectation and BH Theorem 9.3.9:

$$\begin{aligned}
\mathsf{E}\left[(Y - \mathsf{E}[Y|X] - h(X))^2\right] &= \mathsf{E}\left[(Y - \mathsf{E}[Y|X])^2 - 2(Y - \mathsf{E}[Y|X])h(X) + (h(X))^2\right] \\
&= \mathsf{E}\left[(Y - \mathsf{E}[Y|X])^2\right] - \mathsf{E}[2(Y - \mathsf{E}[Y|X])h(X)] + \mathsf{E}\left[(h(X))^2\right] \\
&= \mathsf{E}\left[(Y - \mathsf{E}[Y|X])^2\right] + \mathsf{E}\left[(h(X))^2\right].
\end{aligned}$$

Since $\mathsf{E}\left[(h(X))^2\right] \geq 0$, we conclude that $\mathsf{E}\left[(Y - \mathsf{E}[Y|X] - h(X))^2\right] \geq \mathsf{E}\left[(Y - \mathsf{E}[Y|X])^2\right]$ for any function h, so $\mathsf{E}[Y|X]$ is the predictor of Y based on X with the lowest mean squared error, i.e. the best predictor of Y based on X.

QUESTION

In the next couple of problems we derive Eve's law in a slightly different way then BH.

Define $\hat{X} = \mathsf{E}[X \mid Y]$ as an *estimator* of X and $\tilde{X} = X - \hat{X}$ as the estimation error.

Ex 11.21. Show that $\mathsf{E}\left[\tilde{X} \,\middle|\, Y\right] = 0$.

s.11.21.

$$\begin{aligned}
\mathsf{E}\left[\tilde{X} \,\middle|\, Y\right] = \mathsf{E}\left[X - \hat{X} \,\middle|\, Y\right] &= \mathsf{E}[X \mid Y] - \mathsf{E}[\mathsf{E}[X \mid Y] \mid Y] && (11.0.101)\\
&= \mathsf{E}[X \mid Y] - \mathsf{E}[X \mid Y]\,\mathsf{E}[1 \mid Y] && (11.0.102)\\
&= \mathsf{E}[X \mid Y] - \mathsf{E}[X \mid Y]\,1 = 0 && (11.0.103)
\end{aligned}$$

Ex 11.22. Prove that $\mathsf{E}\left[\tilde{X}\right] = 0$. What does it mean that $\mathsf{E}\left[\tilde{X}\right] = 0$?

h.11.22. Use [11.21]

s.11.22.

$$\mathsf{E}\left[\tilde{X}\right] = \mathsf{E}\left[\mathsf{E}\left[\tilde{X} \,\middle|\, Y\right]\right] = \mathsf{E}[0 \mid Y] = 0. \qquad (11.0.104)$$

This means that the estimation error $\tilde{X}$ does not have bias.

Ex 11.23. Prove that $\mathsf{E}\left[\tilde{X}\hat{X}\right] = 0$.

h.11.23. Use [11.22] and the definitions.

s.11.23.

$$\begin{aligned}
\mathsf{E}\left[\tilde{X}\hat{X}\right] &= \mathsf{E}\left[\mathsf{E}\left[\tilde{X}\hat{X} \,\middle|\, Y\right]\right] && (11.0.105)\\
&= \mathsf{E}\left[\mathsf{E}\left[\tilde{X}\,\mathsf{E}[X \mid Y] \,\middle|\, Y\right]\right] && (11.0.106)\\
&= \mathsf{E}\left[\mathsf{E}[X \mid Y]\,\mathsf{E}\left[\tilde{X} \,\middle|\, Y\right]\right] && (11.0.107)\\
&= \mathsf{E}[\mathsf{E}[X \mid Y]\,0 \mid Y] = 0 && (11.0.108)
\end{aligned}$$

Here, in the rest of the exercises about this topic, we have seen the most terrible mistakes during grading. Hence, study the reasonging applied very carefully, and ensure you know the motivation behind each and every step. There will be questions in the exam about this, and you have to be able to use the arguments. If not, you fail the exam; simple as that. So, you are warned!

Ex 11.24. Show that $\mathsf{Cov}\left[\hat{X}, \tilde{X}\right] = 0$. Conclude that

$$\mathsf{V}[X] = \mathsf{V}\left[\hat{X} + \tilde{X}\right] = \mathsf{V}\left[\hat{X}\right] + \mathsf{V}\left[\tilde{X}\right]. \qquad (11.0.109)$$

s.11.24. Using the previous exercises,

$$\mathsf{Cov}\left[\hat{X},\tilde{X}\right] = \mathsf{E}\left[\hat{X}\tilde{X}\right] - \mathsf{E}\left[\hat{X}\right]\mathsf{E}\left[\tilde{X}\right] = 0 - \mathsf{E}\left[\hat{X}\right]0 = 0. \tag{11.0.110}$$

Next, from the definition of $\tilde{X} = X - \hat{X} \implies X = \tilde{X} + \hat{X}$. The variance of the sum is the sum of the variances since $\hat{X}$ and $\tilde{X}$ are uncorrelated.

Ex 11.25. Prove that $\mathsf{V}\left[\tilde{X}\right] = \mathsf{E}\left[\mathsf{V}\left[X \mid Y\right]\right]$. Conclude Eve's law.

s.11.25. Since $\mathsf{E}\left[\tilde{X}\right] = 0$,

$$\begin{aligned}
\mathsf{V}\left[\tilde{X}\right] &= \mathsf{E}\left[\tilde{X}^2\right] &(11.0.111)\\
&= \mathsf{E}\left[\mathsf{E}\left[\tilde{X}^2 \,\middle|\, Y\right]\right] &(11.0.112)\\
&= \mathsf{E}\left[\mathsf{E}\left[(X-\hat{X})^2 \,\middle|\, Y\right]\right] &(11.0.113)\\
&= \mathsf{E}\left[\mathsf{E}\left[(X-\mathsf{E}\left[X \mid Y\right])^2 \,\middle|\, Y\right]\right] &(11.0.114)\\
&= \mathsf{E}\left[\mathsf{V}\left[X \mid Y\right]\right], &(11.0.115)
\end{aligned}$$

where the last equation follow from the definition of $\mathsf{V}\left[X \mid Y\right]$.

For Eve's law, use the above and the previous exercise to see that

$$\mathsf{V}\left[X\right] = \mathsf{V}\left[\hat{X}\right] + \mathsf{V}\left[\tilde{X}\right] = \mathsf{V}\left[\mathsf{E}\left[X \mid Y\right]\right] + \mathsf{E}\left[\mathsf{V}\left[X \mid Y\right]\right]. \tag{11.0.116}$$

CHEBYSHEV'S INEQUALITY IS useful in proving notions of convergence in probability, which you will see repeatedly in later courses. We say X_n converges in probability to the random variable Z if

$$\lim_{n\to\infty} P(|X_n - Z| \geq \varepsilon) = 0 \quad \forall \varepsilon > 0$$

Note that in the above definition setting $Z = a$ for some constant a would still be valid, as technically the constant a is a random variable.

Ex 11.26. Let Y_n denote the number of heads obtained from throwing a fair coin n times. Then $\frac{Y_n}{n}$ clearly is the proportion of heads in the sample. Find the expectation of this proportion, and show that it converges in probability to its mean. This is denoted as $\frac{Y_n}{n} \xrightarrow{p} \mathsf{E}\left[\frac{Y_n}{n}\right]$ and is known as the weak law of large numbers.

h.11.26. Use Chebyshev's inequality; then take the limit on both sides.

s.11.26. From Probability Theory we know $E\left(\frac{Y_n}{n}\right) = \frac{1}{2}$ and $V\left(\frac{Y_n}{n}\right) = \frac{1}{4n}$. Then by Chebyshev's inequality,

$$\lim_{n\to\infty} P\left(\left|\frac{Y_n}{n} - \frac{1}{2}\right| \geq \varepsilon\right) \leq \lim_{n\to\infty} \frac{V\left(\frac{Y_n}{n}\right)}{\varepsilon^2} = \lim_{n\to\infty} \frac{1}{4n\varepsilon^2} = 0 \quad \forall \varepsilon > 0.$$

Ex 11.27. Where would this proof break down if we try to apply it to e.g. the Cauchy distribution?

s.11.27. The Cauchy distribution has no mean to converge to.

QUESTION

Ex 11.28. Let $Z \sim \text{Norm}(0,1)$. In this exercise, we try to find yet another bound for $\mathsf{P}\{|Z|>3\}$ not yet presented in BH Example 10.1.3.

1. Prove that $\mathsf{P}\{|Z|>3\} \leq e^{-9t}\,\mathsf{E}\left[e^{tZ^2}\right]$.
2. If $X \sim \text{Gamma}(n,\lambda)$ then the MGF of X is given by $M_X(t) = (1-t/\lambda)^{-n}$.
   Use this to derive the MGF of the chi-square distribution with 1 degree of freedom.
3. Determine which t yields the best upper bound of $\mathsf{P}\{|Z|>3\}$, and give this bound.
4. Argue that $\mathsf{P}\{|Z|>3\} = \mathsf{P}\{Z^4>81\}$, and use this to prove that $\mathsf{P}\{|Z|>3\} \leq \frac{1}{27}$.
5. Recall from BH Example 6.5.2. that $\mathsf{E}\left[Z^{2n}\right] = \frac{(2n)!}{2^n n!}$ for all positive integers n.
   Use this (in a way similar to the previous part) to give yet another bound for $\mathsf{P}\{|Z|>3\}$.
   For what value(s) of n do we obtain the strongest bound for $\mathsf{P}\{|Z|>3\}$?

s.11.28. 1. This follows from the inequality $\mathsf{P}\{X \geq a\} \leq \mathsf{E}\left[f(X)\right]/f(a)$ with $f(x) = e^{tX^2}$.

2. The chi-square distribution with 1 degree of freedom is the Gamma distribution with $n=1/2$ and $\lambda=1/2$, so $M_{Z^2}(t) = (1-2t)^{-1/2}$.

3. We find $\mathsf{P}\{|Z|>3\} \leq e^{-9t}(1-2t)^{-1/2}$. So we want to minimize $e^{-9t}(1-2t)^{-1/2}$. It is easier if we take the logarithm first and minimize $-9t-1/2\log(1-2t)$. Its derivative to t is $-9+\frac{1}{1-2t}$, so setting the derivative to 0 yields $t=4/9$. This gives us the bound $\mathsf{P}\{|Z|>3\} \leq e^{-4}(1-2t)^{-1/2} \approx 0.055$.

4. We have $\mathsf{P}\{|Z|>3\} = \mathsf{P}\{Z^4>81\}$ since $|Z|>3$ if and only if $Z^4>81$. The inequality now directly follows from Markov's inequality.

5. By Markov's inequality,

$$\mathsf{P}\{|Z|>3\} = \mathsf{P}\{Z^{2n}>9^n\} \leq \frac{\mathsf{E}\left[Z^{2n}\right]}{9^n} = \frac{1}{9^n}\frac{(2n)!}{2^n n!}.$$

From the formula for $\mathsf{E}\left[Z^{2n}\right]$ we see that $\mathsf{E}\left[Z^{2(n+1)}\right] = (2n+1)\,\mathsf{E}\left[Z^{2n}\right]$. We now consider what happens when incrementing n. If $n<4$ then $2n+1<9$, so then incrementing n improves the bound, for $n=4$ incrementing n doesn't change the bound and for $n>4$ the bound becomes weaker. So we get the best possible bound for $n=4$ and $n=5$, which yields the bound

$$\mathsf{P}\{|Z|>3\} \leq \frac{1\cdot 3\cdot 5\cdot 7}{9^4} \approx 0.016,$$

which is the best bound obtained so far.

QUESTION

An insurance company offers a theft insurance for electric bikes. When a claim is filed, the insurer pays out the size of the claim, with a maximum of 1000 euros. So a claim of 500 euros is paid out completely, while a claim of 1500 euros yields a payout of 1000 euros.

Let X denote the size of the claim in thousands of euros. We assume that X has the following pdf:

$$f_X(x) = \frac{3}{4}x(2-x), \quad 0 \leq x \leq 2. \tag{11.0.117}$$

Let Y denote the size of the payout. Note that $Y = \min\{X, 1\}$.

Ex 11.29 (1)**.** What is the probability that the claim is at most 1000 euros?

s.11.29. We have

$$\begin{aligned}
\mathsf{P}\{X \leq 1\} &= \int_0^1 \frac{3}{4}x(2-x)\,dx &(11.0.118)\\
&= \frac{3}{4}\left[x^2 - \frac{1}{3}x^3\right]_0^1 dx &(11.0.119)\\
&= \frac{3}{4}\left(1 - \frac{1}{3}\right) &(11.0.120)\\
&= 1/2. &(11.0.121)
\end{aligned}$$

An argument based on the fact that f_X is a parabola centered around 1 is also fine.

Correct initial integral: 0.5 point.

Correct solution: 0.5 point.

Ex 11.30 (1.5)**.** What is the expected payout given the information that the claim is at most 1000 euros?

s.11.30. Let Y denote the payout in thousands of euros. Then, $Y = \min\{X, 1\}$. We find

$$\begin{aligned}
\mathsf{E}[Y|X \leq 1] &= \mathsf{E}[X|X \leq 1] &(11.0.122)\\
&= \int_0^1 x f_X(x|X \leq 1)\,dx &(11.0.123)\\
&= \int_0^1 x \frac{f_X(x)}{\mathsf{P}\{X \leq 1\}}\,dx &(11.0.124)\\
&= 2\int_0^1 x\frac{3}{4}x(2-x)\,dx &(11.0.125)\\
&= \frac{3}{2}\int_0^1 (2x^2 - x^3)\,dx &(11.0.126)\\
&= \frac{3}{2}\left[\frac{2}{3}x^3 - \frac{1}{4}x^4\right]_0^1 &(11.0.127)\\
&= \frac{3}{2}\left(\frac{2}{3} - \frac{1}{4}\right) &(11.0.128)\\
&= \frac{5}{8}. &(11.0.129)
\end{aligned}$$

Writing down $\mathsf{E}[X|X \leq 1]$: 0.5 point.
Correct derivation and solution: 1 point.

Ex 11.31 (1.5)**.** What is the (unconditional) expected payout?

s.11.31. By the law of total expectation,

$$\begin{aligned} \mathsf{E}[Y] &= \mathsf{P}\{X \leq 1\}\,\mathsf{E}[Y|X \leq 1] + \mathsf{P}\{X > 1\}\,\mathsf{E}[Y|X > 1] && (11.0.130)\\ &= \frac{1}{2}\frac{5}{8} + \frac{1}{2} \cdot 1 && (11.0.131)\\ &= \frac{13}{16}. && (11.0.132) \end{aligned}$$

Mentioning law of total expectation: 0.5 point.
Correctly using law of total expectation: 0.5 point.
Correct solution: 0.5 point.

The time T (in hours) it takes for the company to process a payout of size $Y = y$ is uniformly distributed on the interval $[y, 2y]$.

Ex 11.32 (1)**.** Compute the (unconditional) expected value of T.

s.11.32. By Adam's law,

$$\begin{aligned} \mathsf{E}[T] &= \mathsf{E}[\mathsf{E}[T|Y]] && (11.0.133)\\ &= \mathsf{E}\left[\frac{3}{2}Y\right] && (11.0.134)\\ &= \frac{3}{2}\mathsf{E}[Y] && (11.0.135)\\ &= \frac{3}{2} \cdot \frac{13}{16} = \frac{39}{32}. && (11.0.136) \end{aligned}$$

Mentioning and using Adam's law: 0.5 point.
Correct solution: 0.5 point.
Note: if in your interpretation, $Y = y$ is actually given (I realized that the question is ambiguous), then $\frac{3}{2}y$ (lowercase!) will also be regarded as correct.

QUESTION

Suppose it is now Sinterklaas and everyone in your family writes and reads poems for each other for celebration. You are in a family of 5 (including you) and you have already heard 3 poems, which means there are still 2 poems left. Let X_1, X_2, X_3 be the time (in minutes) spent on each of the first 3 poems, and X_4, X_5 be that of the remaining poems. Assume that $X_i \sim \text{Norm}(3,1)$ for $i = 1, \dots, 5$.

Ex 11.33 (1.5)**.** First assume that the times spent on each poem are all independent. What is the expected number of remaining poems that take more time to read than each of the 3 poems you have already heard?

s.11.33. Let I_i be the indicator r.v. for the ith poem taking more time to read than each of poem 1, 2 and 3. Then:

$$\begin{aligned}
\mathsf{P}\{I_i = 1\} &= \mathsf{P}\{X_i > X_1, X_i > X_2, X_i > X_3\} \\
&= \mathsf{P}\{X_i = \max\{X_1, X_2, X_3, X_i\}\} \\
&= \frac{1}{4},
\end{aligned}$$

by symmetry. Then $\mathsf{E}\left[\Sigma_{i=4}^5 I_i\right] = \frac{1}{4} \cdot 2 = \frac{1}{2}$.

For the next two exercises, suppose that $(X_1, \dots, X_5)$ is now Multivariate Normal distributed with $\mathsf{Corr}\left[X_1, X_j\right] = \frac{1}{2}$ for $j = 4, 5$

Ex 11.34 (2.5)**.** On average, how many of the remaining poems take at least 1 minutes more to read compared to the 1st poem?

s.11.34. In order to answer this question, we want to know $P(X_i - X_1 > 1)$, for $i = 4, 5$. We first consider $i = 4$. Since X_4 and X_1 are jointly normal distributed, $X_4 - X_1$ is also normally distributed, with $\mathsf{E}[X_4 - X_1] = \mathsf{E}[X_4] - \mathsf{E}[X_1] = 0$ and

$$\begin{aligned}
\mathsf{V}[X_4 - X_1] &= \mathsf{V}[X_4] + \mathsf{V}[-X_1] + 2\,\mathsf{Cov}[X_4, -X_1] \\
&= 1 + 1 - 2 \cdot \frac{1}{2} \cdot 1 \cdot 1 \\
&= 1
\end{aligned}$$

Then we know $X_4 - X_1$ follows a standard normal distribution and $P(X_4 - X_1 > 1) = 0.16$ Similarly, $P(X_5 - X_1 > 1) = 0.16$. So the average number of the remaining poems that take 1 minute more to read than the first poem is $0.16 \cdot 2 = 0.32$.

Ex 11.35 (1)**.** Show that there exists a constant c such that $X_1 - cX_4$ and X_4 are independent, and determine the value of c.

s.11.35. $\mathsf{Cov}[X_1 - cX_4, X_4] = \mathsf{Cov}[X_1 - cX_4, X_4] = \mathsf{Cov}[X_1, X_4] - c\mathsf{V}[X_4] = \frac{1}{2} - c$, so for $c = \frac{1}{2}$, we have that $X_1 - cX_4$ and X_4 are uncorrelated. Since $(X_1, \dots, X_5)$ has the multivariate normal distribution, it follows that $X_1 - cX_4$ and X_4 are independent.

Remarks and grading scheme:

1. Ex 3.1: Many students assume that $X_i > X_1$ and $X_i > X_2$ is independent. This is not the case. In fact, If $X_i > X_1$, then it's more likely that X_i is large. In consequence, it is also more likely that $X_i > X_2$.
2. Ex 3.1: 0.5 point for multiply your probability with 2 (even if it is calculated wrongly). Full point(1.5) for correct answer.
3. Ex 3.2: 0.5 point for mentioning that $X_4 - X_1$ is normally distributed ,0.5 point for correctly calculated $\mathsf{E}[X_4 - X_1]$ and 0.5 point for correctly calculated $Var(X_4 - X_1)$.
4. Ex 3.3: 0.5 point for writing out the formula for covariance.

QUESTION

Let X and Y be independent and exponentially distributed with rates $\lambda_1 = 1$ and $\lambda_2 = 2$ respectively.

Ex 11.36 (0.5)**.** Find the joint PDF $f(x,y)$ of X and Y.

s.11.36. Since X and Y are independent, their joint PDF is the product of their marginal PDFs. This gives:

$$f(x,y) = (1e^{-1x})(2e^{-2y}) = 2e^{-x-2y}$$

One mistake, zero points

Ex 11.37 (0.5)**.** Show that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x\,\mathrm{d}y = 1$.

s.11.37. Calculating this integral gives:

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x\,\mathrm{d}y =& \int_0^{\infty}\int_0^{\infty} 2e^{-x-2y}\,\mathrm{d}x\,\mathrm{d}y \\ =& \int_0^{\infty} 2e^{-2y)}[-e^{-x}]_0^{\infty}\,\mathrm{d}y \\ =& \int_0^{\infty} 2e^{-2y}\,\mathrm{d}y \\ =& \int_0^{\infty} h(y)\,\mathrm{d}y \\ =& 1\end{aligned}$$

Where $h(y)$ is the PDF of Y.
One mistake, zero points

Ex 11.38 (3)**.** Find $\mathsf{E}[|X-Y|]$, i.e., the expected distance between X and Y.

s.11.38. Similar to example 7.2.2., we get that by 2D LOTUS:

$$\begin{aligned}\mathsf{E}[|X-Y|] =& \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |x-y| f(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ =& \int_0^{\infty}\int_y^{\infty}(x-y)(2e^{-x-2y})\,\mathrm{d}x\,\mathrm{d}y + \int_0^{\infty}\int_0^{y}(y-x)(2e^{-x-2y})\,\mathrm{d}x\,\mathrm{d}y \\ =& \int_0^{\infty}\left([(x-y)\cdot -2e^{-x-2y}]_y^{\infty} - \int_y^{\infty} -2e^{-x-2y}\,\mathrm{d}x\right)\mathrm{d}y \\ &+ \int_0^{\infty}\left([(y-x)\cdot -2e^{-x-2y}]_0^{y} - \int_0^{y} 2e^{-x-2y}\,\mathrm{d}x\right)\mathrm{d}y \\ =& \int_0^{\infty}[-2e^{-x-2y}]_y^{\infty}\,\mathrm{d}y + \int_0^{\infty} 2ye^{-2y} - [-2e^{-x-2y}]_0^{y}\,\mathrm{d}y \\ =& \int_0^{\infty} 2e^{-3y}\,\mathrm{d}y + \int_0^{\infty} 2ye^{-2y} + 2e^{-3y} - 2e^{-2y}\,\mathrm{d}y \\ =& \left[-\tfrac{2}{3}e^{-3y}\right]_0^{\infty} + \left[-ye^{-2y} + \tfrac{1}{2}e^{-2y} - \tfrac{2}{3}e^{-3y} + e^{-2y}\right]_0^{\infty} \\ =& \tfrac{2}{3} + \tfrac{1}{2} + \tfrac{2}{3} - 1 = \tfrac{5}{6}.\end{aligned}$$

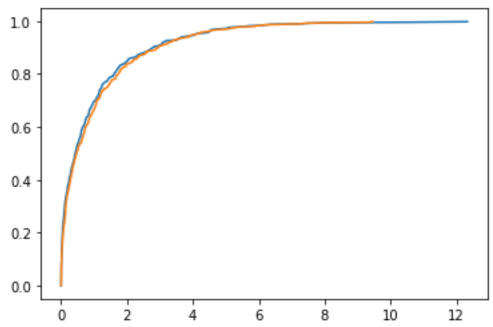
One point for writing down the integral correctly using LOTUS and splitting it up correctly. Two points for the computations.

Consider the following code:

Python Code

```python
import numpy as np
np.random.seed(3)

num = 500

x = np.random.normal(loc = 50, scale = 200, size = num)

result2 = np.zeros(num)
for i in range(0,num):
    result2[i] = ((x[i]-50)/200)**2

probs = np.arange(0,num)/num
result2 = np.sort(result2)
y = np.random.chisquare(df = 1, size = num)
y = np.sort(y)
plt.plot(result2, probs)
plt.plot(y, probs)
plt.show()
```

Ex 11.39 (1)**.** What does the code above do and why would you expect to get the graph below as output?



s.11.39. It loads the required packages and creates one sample with 500 observations from a $\mathcal{N}(50,200)$-distribution. Then for all observations it standardizes and takes the square. The empirical CDF of the standardized values is plotted against the PDF of a sample from a chi-square distribution with 1 degree of freedom. It can be seen they look very much alike. This is expected as for $Z \sim \mathcal{N}(0,1)$ we have $Z^2 \sim \chi^2_1$.
0.5 points for mentioning data is standardized, 0.5 points for mentioning a squared standard normal r.v. is chi-square.

QUESTION

Ex 11.40 (1)**.** Let $U_1, U_2 \sim \text{Unif}(0,1)$. Find the PDF of $X_1 = (U_1)^{1/a}$ and then immediately give the PDF of $X_2 = (U_2)^{1/b}$ for $a, b > 0$.

s.11.40. For $y \in (0,1)$:

$$F_{X_1}(y) = \mathsf{P}\left\{X_1 \le y\right\} = \mathsf{P}\left\{(U_1)^{1/a} \le y\right\} = \mathsf{P}\left\{U_1 \le y^a\right\} = F_{U_1}(y^a).$$

We know that $F_{U_1}(y) = y$ for $y \in (0,1)$. Hence $F_{X_1}(y) = F_{U_1}(y^a) = y^a$. Then

$$f_{X_1}(y) = \frac{\partial y^a}{\partial y} = a y^{a-1} \quad \forall y \in (0,1).$$

Now we can say $f_{X_2}(y) = b y^{b-1}$ for all $y \in (0,1)$. Both PDFs are 0 outside of this region.

Grading scheme:

- Correct application of transformation theorem or CDF technique 0.5pt.
- Most of: the correct bounds, the verification the theorem is applicable, no mistakes in calculation 0.5pt.

Ex 11.41 (0.5)**.** What distributions do X_1 and X_2 have? Also give the corresponding parameters.

s.11.41. $X_1 \sim \text{Beta}(a,1)$ and $X_2 \sim \text{Beta}(b,1)$.

Grading scheme:

- Correct 0.5pt.

Ex 11.42 (2)**.** Let $B \sim \text{Beta}(p,q)$ for some $p, q > 0$. Show that $1 - B \sim \text{Beta}(q,p)$.

s.11.42. For $y \in (0,1)$ we have that

$$F_{1-B}(y) = \mathsf{P}\left\{1 - B \le y\right\} = \mathsf{P}\left\{B \ge 1 - y\right\} = 1 - \mathsf{P}\left\{B \le 1 - y\right\} = 1 - F_B(1-y).$$

We can write this as

$$\begin{aligned}
1 - F_B(1-y) &= 1 - \int_0^{1-y} \frac{x^{p-1}(1-x)^{q-1}}{\beta(p,q)}\,\mathrm{d}x \\
&= 1 - \int_1^{y} \frac{(1-x)^{p-1}(x)^{q-1}}{\beta(p,q)}\,\mathrm{d}(1-x) \\
&= 1 - \int_y^{1} \frac{(1-x)^{p-1}(x)^{q-1}}{\beta(p,q)}\,\mathrm{d}x \\
&= 1 - \int_y^{1} \frac{(1-x)^{p-1}(x)^{q-1}}{\beta(q,p)}\,\mathrm{d}x \\
&= \int_0^{y} \frac{(1-x)^{p-1}(x)^{q-1}}{\beta(q,p)}\,\mathrm{d}x \\
&= F_D(y)
\end{aligned}$$

For $D \sim \text{Beta}(q,p)$. Since both r.v.s have support $(0,1)$ and have the same CDF on this support we conclude $1-B \sim \text{Beta}(q,p)$. *Remark.* This can also be shown by looking at the PDF, using a similar derivation.

Grading scheme:

- Noting the Beta function is symmetric 0.5pt.
- Calculating the correct inverse transformation 0.5pt.
- Correct application transformation theorem 0.5pt.
- Most of: the correct bounds, the verification the theorem is applicable, no mistakes in calculation 0.5pt.
- **OR:** The derivation as above correct 2pt.
- **OR:** A reasonable attempt at a story proof 1pt.

Ex 11.43 (1.5)**.** Let Z be a random variable on $(0,1)$. The PDF of Z is given by

$$f_Z(z) = \begin{cases} f_{X_1}(z) & \text{if } z \in (0, \frac{1}{2}] \\ f_{1-X_2}(z) & \text{if } z \in (\frac{1}{2}, 1) \\ 0 & \text{elsewhere.} \end{cases}$$

(i) Does there exist *more than one* combination of $a, b > 0$ (and $a, b \in \mathbf{R}$) such that this is a valid PDF?

(ii) Does there exist *at least one* combination of a, b as above and $a = b$ such that Z follows a Beta distribution?

Explain your answers clearly.

s.11.43. For a PDF to be valid it needs to be non-negative and integrate to 1. Clearly f_Z is non-negative, so let's check the other condition. We know from parts (b) and (c) that $1 - X_2 \sim \text{Beta}(1, b)$. Then,

$$\begin{aligned} \int_0^1 f_Z(y)\,\mathrm{d}y &= \int_0^{\frac{1}{2}} f_{X_1}(y)\,\mathrm{d}y + \int_{\frac{1}{2}}^1 f_{1-X_2}(y)\,\mathrm{d}y \\ &= \int_0^{\frac{1}{2}} a y^{a-1}\,\mathrm{d}y + \int_{\frac{1}{2}}^1 b(1-y)^{b-1}\,\mathrm{d}y \\ &= \left(\frac{1}{2}\right)^a + \left(\frac{1}{2}\right)^b. \end{aligned}$$

This should equal 1. The easy solution is $a = b = 1$. The above can also be solved to obtain $b = -\frac{\ln(1-2^{-a})}{\ln 2}$, hence there are infinitely many solutions for $a, b > 0$. For $a = b$, $a = b = 1$ is the only solution and Z then follows the Beta(1,1) distribution.

Grading scheme:

- Correct integration 0.5pt.
- Found at least one other combination or showed such a combination must exist 0.5pt.
- Noticing that for $a = b = 1$ there is a Beta distribution 0.5pt.

QUESTION

Let X be Unif(1,3) distributed and Y be exponentially distributed with rate $\lambda = 2$. X and Y are independent.

Ex 11.44 (1). Find the joint PDF $f(x,y)$ of X and Y.

s.11.44. Since X and Y are independent, their joint PDF is the product of their marginal PDFs. This gives:

$$\begin{aligned} f(x,y) &= \left(\frac{1}{3-1}\right)(2e^{-2y}) \\ &= \left(\frac{1}{2}\right)(2e^{-2y}) \\ &= e^{-2y}, \text{ for } y > 0 \text{ and } x \in [1,3] \end{aligned}$$

0.5 points for the correct expression, 0.5 points for the boundary

Ex 11.45 (2). Find $\mathsf{P}\{X \le Y\}$.

s.11.45. We have

$$\begin{aligned} \mathsf{P}\{Y \le X\} &= \int_1^3 \int_0^x e^{-2y} dy dx \\ &= \int_1^3 [-\frac{1}{2}e^{-2y}]_0^x dx \\ &= \int_1^3 (-\frac{1}{2}e^{-2x} + \frac{1}{2}) dx \\ &= -\frac{1}{2}\int_1^3 e^{-2x} dx + 1 \\ &= -\frac{1}{2}[-\frac{1}{2}e^{-2x}]_1^3 + 1 \\ &= \frac{1}{4}(e^{-6} - e^{-2}) + 1 \\ &= 1 - \frac{e^{-2} - e^{-6}}{4}. \end{aligned}$$

So $\mathsf{P}\{X \le Y\} = 1 - P(X \le Y) = \frac{e^{-2} - e^{-6}}{4}$.
One point for writing down an integral with the correct bounds. One point for the computations.

Consider the following code:

Python Code

```python
import numpy as np
np.random.seed(3)

num = 10000

# Lambda = 1/1 = 1
x = np.random.exponential(scale = 1, size = num)
# Lambda = 1/2
y = np.random.exponential(scale = 2, size = num)

result = np.zeros(num)
for i in range(0, len(result)):
    result[i] = min(x[i],y[i])

print(np.mean(result))
```

R Code

```r
set.seed(3)

num = 10000

# Lambda= = 1
x = rexp(num, 1)
# Lambda = 1/2
y = rexp(num,0.5)

result = rep(0, num)
for (i in 1:length(result)) {
  result[i] = min(x[i], y[i])
}

print(mean(result))
```

Ex 11.46 (2)**.** The output of the code above is approximately $\frac{2}{3}$. Why would you expect to get this output? Explicitly mention which convergence result you are using in your reasoning.

s.11.46. The code computes the sample mean for the minimum of $X \sim \text{Exp}(1)$ and $Y \sim \text{Exp}(\frac{1}{2})$. Since X and Y are independent, we have $\min(X,Y) \sim \text{Exp}(1+\frac{1}{2})$. So then $\mathsf{E}[\min(X,Y)] = \frac{1}{1\frac{1}{2}} = \frac{2}{3}$. By the law of large numbers, the sample mean converges to the population mean for large enough samples. Hence, we expect the output to be approximately $2/3$. *0.5 points for explaining what the code does. 1 points for computing the population mean (with correct argumentation). 0.5 points for mentioning the (strong/weak/any) law of large numbers.*

QUESTION

Tom and Jerry are the only two clerks at a local bank. Tom serves N_1 customers per hour, $N_1 \sim \text{Poisson}(\lambda_1)$; Jerry serves N_2 customers per hour, $N_2 \sim \text{Poisson}(\lambda_2)$ such that $\lambda_1 > \lambda_2 > 0$. Each customer that gets served has a probability p of applying for a credit card, independently. Let X be the number of customers that apply for credit cards per hour.

Ex 11.47 (1)**.** Show that $2N_1 + N_2$ and $2N_1 - N_2$ are **not** independent of each other.

s.11.47.

$$\begin{aligned}
&\mathsf{Cov}\,[2N_1 + N_2, 2N_1 - N_2] \\
=&\mathsf{Cov}\,[2N_1, 2N_1] - \mathsf{Cov}\,[2N_1, N_2] + \mathsf{Cov}\,[N_2, 2N_1] - \mathsf{Cov}\,[N_2, N_2] \cdots (0.5 \text{ point}) \\
=&4 \cdot Var(N_1) - Var(N_2) \\
=&4\lambda_1 - \lambda_2 \neq 0. \cdots (0.5 \text{ point})
\end{aligned}$$

Ex 11.48 (1)**.** Let $N = N_1 + N_2$. Suppose N_1 and N_2 are independent, what is the distribution of N? What is the distribution of $X|N$? What is the distribution of X?

s.11.48. Since N_1 and N_2 are independent, $N \sim Pois(\lambda_1 + \lambda_2)$.(0.5 point)
$X|N \sim Bin(N, p)$ and $X \sim Pois((\lambda_1 + \lambda_2)p)$ by the Chicken-egg theory. (0.5 point)

Ex 11.49 (2)**.** Calculate $\rho_{X,N}$, the correlation between X and N.

s.11.49. Let $Y = N - X$ be the number of customers that do not apply for a credit card. Then we know $Y \sim Pois((\lambda_1 + \lambda_2)q)$ with $q = 1 - p$, and X and Y are independent. (0.5 point)

$$\begin{aligned}
\mathsf{Cov}\,[N, X] &= \mathsf{Cov}\,[X + Y, X] \\
&= \mathsf{Cov}\,[X, X] + \mathsf{Cov}\,[Y, X] \\
&= Var(X) \cdots (0.5 \text{ point}) \\
&= (\lambda_1 + \lambda_2)p \cdots (0.5 \text{ point})
\end{aligned}$$

Then it follows that

$$\begin{aligned}
\rho_{X,N} &= \frac{\mathsf{Cov}\,[N, X]}{sd(N) \cdot sd(X)} \\
&= \frac{(\lambda_1 + \lambda_2)p}{\sqrt{\lambda_1 + \lambda_2} \cdot \sqrt{(\lambda_1 + \lambda_2)p}} \\
&= \sqrt{p}. \cdots (0.5 \text{ point})
\end{aligned}$$

Consider the following codes:

R Code

```r
library(mvtnorm)
set.seed(444)
A<-diag(x=1,nrow = 3)
B<-rep(0,3)
X<-rmvnorm(100,mean=B,sigma=A)
output=cov(X[,-3])
```

Python Code

```
1 import random
2 import numpy as np
3 random.seed(444)
4 A = np.diag([1,1,1])
5 B = np.zeros(3)
6 X = np.random.multivariate_normal(B, A, size = 100)
7 output = np.cov(X, rowvar = False)[0:2,0:2]
```

Ex 11.50 (1)**.** Explain in detail the purpose of each line of the above codes.

s.11.50. Line 1: Load the package "mvtnorm" so that we can use the function *rmvnorm*.
Line 2: Set a random seed to reproduce the same results.
Line 3: Generate a 3×3 identity matrix.
Line 4: Generate a vector of 0's.
Line 5: Generate 100 multivariate normally distributed variables $\mathbf{X} = x_1, x_2, x_3$ with mean equal to $(0, 0, 0)$ and variance equal to an identity matrix.
Line 6: Calculate the covariance matrix of the first 2 columns of matrix X.
(0.5 points for mentioning at least 3 of the above.)

QUESTION

Let X and Y be independent and $\mathcal{N}(0,1)$ distributed.

Ex 11.51 (1)**.** Show that $X - Y = \sqrt{2}Z$, where $Z \sim \mathcal{N}(0,1)$.

s.11.51. We know by independence of X and Y that $X - Y \sim \mathcal{N}(0,2)$. By the fact that $cZ \sim \mathcal{N}(0,c^2)$ for all $c \in \mathbb{R}$, using $c = \sqrt{2}$ we get the same distribution.
One point for the fact $X - Y \sim \mathcal{N}(0,2)$, one point for a correct conclusion that this equals the density of Z.

Ex 11.52 (2)**.** Consider the expectation $E|X - Y|$. Show that

$$E|X-Y| = 2\sqrt{2}\int_0^\infty z\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}\,dz.$$

You may use the result in the previous exercise and the fact that by the Fundamental Theorem of Calculus, $\int_b^a f(x)\,dx = -\int_a^b f(x)\,dx$, if $b > a$.

s.11.52. Using 2D LOTUS, substitution, and the integral equation above, we obtain

$$\begin{aligned}
E|X-Y| &= \int_\infty^\infty |\sqrt{2}z|\phi(z)dz \\
&= \int_{-\infty}^\infty |\sqrt{2}z|\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}dz \\
&= \sqrt{2}\int_{-\infty}^\infty |z|\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}dz \\
&= \sqrt{2}\int_0^\infty z\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}dz + \sqrt{2}\int_{-\infty}^0 (-z)\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}dz \\
&= \sqrt{2}\int_0^\infty z\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}dz - \sqrt{2}\int_\infty^0 u\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}du \\
&= \sqrt{2}\int_0^\infty z\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}dz + \sqrt{2}\int_0^\infty u\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}du \\
&= 2\sqrt{2}\int_0^\infty z\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}dz
\end{aligned}$$

0.5 points for the first integral, 1 point for splitting up correctly, 0.5 points for simplifying correctly.

Ex 11.53 (1)**.** Solve the integral in the previous question. *Hint: use integration by substitution* *s.11.53.* Integration by substitution yields:

$$\begin{aligned} E(|X-Y|) =& 2\sqrt{2}\int_0^\infty z\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}dz \\ =& 2\sqrt{2}\int_{0^2}^{\infty^2}\frac{1}{\sqrt{2\pi}}e^{-u}du \\ =& 2\sqrt{2}\int_{0}^{\infty^2}\frac{1}{\sqrt{2\pi}}e^{-u}du \\ =& \frac{2}{\sqrt{\pi}}\left(1-e^{-\infty^2}\right) = \frac{2}{\sqrt{\pi}} \end{aligned}$$

0.5 points for the right substitution, 0.5 points for the rest of the computations.

Consider the following code:

Python Code

```python
import numpy as np
np.random.seed(3)

num = 10000

y = np.random.normal(loc = 1, scale = np.sqrt(2), size = num)

result = np.zeros(num)
for i in range(0, len(result)):
    result[i] = np.exp(y[i])

print(np.mean(result))
```

R Code

```r
set.seed(3)

num = 10000

y = rnorm(num, mean = 1, sd = sqrt(2))

result = rep(0, num)
for (i in 1:length(result)) {
  result[i] = exp(y[i])
}

print(mean(result))
```

Ex 11.54 (1)**.** What does the code above do?

s.11.54. It loads the required packages and creates one sample with 10000 observations from a r.v. $Y \sim \mathcal{N}(1,2)$. Then for all observations y_i it calculates e^{y_i} and stores it into a vector. Finally, it estimates the mean of a log-normal r.v. $X = e^Y$.
0.5 points for mentioning that for observations of a normal r.v. the exponent is taken. 0.5 points for stating a mean of X is estimated.

QUESTION

A random point (X, Y) is chosen in the following square:

$$\{(x, y) : x^2 < \pi^2, y^2 < \pi^2\}$$

All points are equally likely to be chosen. Let N be the scaled Euclidean norm of (X, Y). So $N = c\sqrt{X^2 + Y^2}$, where $c > 0$.

Ex 11.55 (1)**.** Find the joint PDF $f(x, y)$ of X and Y.

s.11.55. The random vector (X, Y) is uniformly distributed on $(-\pi, \pi)^2$. Hence, the joint pdf is given by

$$\begin{aligned} f(x, y) =& \Big(\frac{1}{\pi - (-\pi))}\Big)\Big(\frac{1}{(\pi - (-\pi))}\Big) \\ =& \frac{1}{4\pi^2} \end{aligned}$$

for $-\pi < x < \pi$ and $-\pi < x < \pi$. *0.5 points for the correct solution, 0.5 points for the boundaries.*

Ex 11.56 (3)**.** Find the value for c such that $\mathsf{E}\left[N^2\right] = 1$.

s.11.56. Note that $N^2 = c^2(X^2 + Y^2)$. Using LOTUS:

$$\begin{aligned} \mathsf{E}\left[N^2\right] =& \mathsf{E}\left[c^2(X^2 + Y^2)\right] \\ =& \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} c^2\Big(x^2 + y^2\Big) f(x, y)\, dx\, dy \\ =& c^2 \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \Big(x^2 + y^2\Big)\Big(\frac{1}{4\pi^2}\Big)\, dx\, dy \\ =& c^2 \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \frac{x^2 + y^2}{4\pi^2}\, dx\, dy \\ =& \frac{c^2}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} x^2 + y^2\, dx\, dy \\ =& \frac{c^2}{4\pi^2}\int_{-\pi}^{\pi} \Big[\frac{x^3}{3} + y^2 x\Big]_{-\pi}^{\pi}\, dy \\ =& \frac{c^2}{4\pi^2}\int_{-\pi}^{\pi} \Big(\big(\frac{\pi^3}{3} + y^2\pi\big) - \big(\frac{(-\pi)^3}{3} - y^2\pi\big)\Big)\, dy \\ =& \frac{c^2}{4\pi^2}\int_{-\pi}^{\pi} \Big(\frac{2\pi^3}{3} + 2y^2\pi\Big)\, dy \\ =& \frac{c^2}{4\pi^2}\int_{-\pi}^{\pi} \frac{2}{3}\pi^3 + 2\pi y^2\, dy \\ =& \frac{c^2}{4\pi^2}\Big[\frac{2}{3}\pi^3 y + \frac{2}{3}\pi y^3\Big]_{-\pi}^{\pi} \\ =& \frac{c^2}{4\pi^2}\Big(\frac{2}{3}\pi^4 + \frac{2}{3}\pi^4 + \frac{2}{3}\pi^4 + \frac{2}{3}\pi^4\Big) \\ =& \frac{c^2}{4\pi^2}\frac{8}{3}\pi^4 = c^2\frac{2\pi^2}{3} \end{aligned}$$

So then since $c^2 \frac{2\pi^2}{3} = 1 \implies c^2 = \frac{3}{2\pi^2}$. We have that $c = \sqrt{\frac{3}{2\pi^2}} = \frac{\sqrt{\frac{3}{2}}}{\pi} = \frac{\sqrt{1\frac{1}{2}}}{\pi} > 0$. Where you should use $c > 0$.

1 point for $N = c^2(X^2 + Y^2)$ and writing down the integral correctly using LOTUS. 1 point for the calculations to find the expectation. 1 point for finding the correct value of c.

Consider the following code:

Python Code

```python
import math
from scipy.integrate import quad

def f(x):
    return 1/(math.pi*(1+x**2))

print(quad(f, -math.inf, math.inf))
```

R Code

```r
f = function(x){
  return(1/(pi*(1+x^2)))
}
integrate(f, -Inf, Inf)
```

Ex 11.57 (1)**.** What will the code above return? You may use the fact that the pdf of a Cauchy random variable is given by

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in (-\infty, \infty).$$

s.11.57. The code computes the integral over the entire domain of a Cauchy random variable. Hence, it returns a value of one.
0.5 points for explaining what the code does. 0.5 points for mentioning the correct output.

QUESTION

Ex 11.58 (2)**.** Let $\lambda > 0$ be given. Let $X_1, X_2, \ldots, X_n \sim \text{Expo}(\lambda)$ be independent. Let $Y_i = 2\lambda X_i$ for $i = 1, 2, \ldots, n$. Using theorems, not results (so show all calculations), what is the PDF of $S = Y_1 + Y_2$?

s.11.58. We first find the distribution of Y_1.

$$F_{Y_1}(y) = \mathsf{P}\left\{Y_1 \le y\right\} = \mathsf{P}\left\{2\lambda X_1 \le y\right\} = \mathsf{P}\left\{X_1 \le \frac{y}{2\lambda}\right\} = F_{X_1}\left(\frac{y}{2\lambda}\right).$$

We can easily find that

$$F_{X_1}(y) = \int_0^y \lambda e^{-\lambda x}\,\mathrm{d}x = 1 - e^{-\lambda y}$$

for $y > 0$, and 0 elsewhere. Then $F_{Y_1} = 1 - e^{-y/2}$, and we conclude $Y_1 \sim \text{Expo}(\frac{1}{2})$. By symmetry, $Y_2 \sim \text{Expo}(\frac{1}{2})$. Note Y_1 and Y_2 are independent. By the convolution theorem, we know that

$$\begin{aligned} f_{Y_1+Y_2}(t) &= \int_0^t f_{Y_1}(t-s) f_{Y_2}(s)\,\mathrm{d}s \\ &= \int_0^t \frac{1}{2} e^{-\frac{t-s}{2}} \frac{1}{2} e^{-\frac{s}{2}}\,\mathrm{d}s \\ &= \frac{1}{4}\int_0^t e^{-\frac{t}{2}}\,\mathrm{d}s \\ &= \frac{1}{4} t e^{-\frac{t}{2}} \end{aligned}$$

for $t > 0$, and 0 elsewhere.

Grading scheme:

- Derived the correct distribution of Y_i 0.5pt.
- Noticed that Y_i are independent to apply convolution theorem 0.5pt.
- Convolution theorem correctly applied 0.5pt.
- Most of: the correct bounds, no mistakes in calculation 0.5pt.

Ex 11.59 (1)**.** What is the difference between the PDF of $\sum\limits_{i=1}^n X_i$ and that of nX_i? Why are they different? What distributions do they follow? You can use results from the book here, so keep it brief.

s.11.59. B.H. show that $\sum\limits_{i=1}^n X_i \sim \text{Gamma}(n, \lambda)$, whereas $nX_i \sim \text{Expo}\left(\frac{\lambda}{n}\right)$. The distributions are different since the first one is a sum of independent random variables, whereas the latter is one random variable that is scaled.

Grading scheme:

- Correct distributions given 0.5pt.
- Reason why 0.5pt. (very very lenient here)

Ex 11.60 (2)**.** Let $Z \sim \chi^2(2n)$ and let S be as in part (a). Assume that Z and S are independent. Showing all calculations, what is the PDF of $W = S + Z$? What is its distribution?

s.11.60. We know the PDF of Z is given by

$$f_Z(x) = \frac{1}{2^n\Gamma(n)} x^{n-1} e^{-\frac{x}{2}}$$

for $x > 0$, and 0 elsewhere. In (a) we have shown the PDF of S. Then, by the convolution formula we get

$$\begin{aligned}
f_W(w) &= \int_0^w \frac{1}{4}(w-x)e^{-\frac{(w-x)}{2}} \frac{1}{2^n\Gamma(n)} x^{n-1} e^{-\frac{x}{2}} \,\mathrm{d}x \\
&= \frac{e^{-\frac{w}{2}}}{2^{n+2}\Gamma(n)} \left(\int_0^w w x^{n-1} \,\mathrm{d}x - \int_0^w x^n \,\mathrm{d}x \right) \\
&= \frac{e^{-\frac{w}{2}}}{2^{n+2}\Gamma(n)} \left(\frac{w^{n+1}}{n} - \frac{w^{n+1}}{n+1} \right) \\
&= \frac{w^{n+1}e^{-\frac{w}{2}}}{2^{n+2}\Gamma(n)} \left(\frac{1}{n(n+1)} \right) \\
&= \frac{w^{n+1}e^{-\frac{w}{2}}}{2^{n+2}\Gamma(n+2)}
\end{aligned}$$

for $t > 0$ and 0 elsewhere. This is the Gamma$\left(n+2, \frac{1}{2}\right)$ distribution.

Grading scheme:

- Noting Z follows a Gamma distribution with correct parameters 0.5pt. (to be lenient)
- Applying the convolution theorem 0.5pt.
- Recognition of final distribution 0.5pt.
- Most of: the correct bounds, no mistakes in calculation 0.5pt.
- No recognition that a sum of Gamma distributions with different rate parameters does not work as you want it to -0.5pt (if applicable).

QUESTION

Ex 11.61 (1.5)**.** Let $U \sim \text{Unif}(-1,1)$. Find the PDF of $B = |U|$. What is its distribution? What is $\mathsf{E}[B]$?

s.11.61. Notice that the function $|x|$ is not one-to-one on $(-1,1)$, hence we cannot use the transformation theorem here. We see that since $U \in (-1,1)$, $B \in [0,1)$. Then we see that

$$F_B(y) = \mathsf{P}\left\{B \leq y\right\} = \mathsf{P}\left\{|X| \leq y\right\} = \mathsf{P}\left\{-y \leq X \leq y\right\} = F_X(y) - F_X(-y)$$

For $y \in [0,1]$. We know $F_X(y) = \frac{y+1}{2}$, so then $F_X(y) - F_X(-y) = \frac{y+1}{2} - \frac{-y+1}{2} = y$, and we conclude that $B \sim \text{Unif}(0,1)$. Then we know $\mathsf{E}[B] = \frac{1}{2}$.

Grading scheme:

- Correct derivation with CDF 0.5pt
- No mistakes in the above 0.5pt.
- Expectation 0.5pt.

Ex 11.62 (1.5)**.** Let X be a continuous random variable such that $M_X(t) = e^t M_X(-t)$. What is $\mathsf{E}[X]$? Can you conclude that X is distributed in the same manner as B?

s.11.62. Using the given relation of M_X, we see that

$$M_X(t) = e^t M_X(-t) = e^t\,\mathsf{E}\left[e^{-tX}\right] = \mathsf{E}\left[e^{t(1-X)}\right] = M_{1-X}(t)$$

and since the MGF determines the distribution we can immediately say that $X \sim 1-X$. Then it must be that $\mathsf{E}[X] = \mathsf{E}[1-X]$ and then by linearity we have that $\mathsf{E}[X] = \frac{1}{2}$. This is not enough information to conclude what distribution X has, we can only see that it must be symmetric around $\frac{1}{2}$.

Grading scheme:

- Note $X \sim 1-X$ 0.5pt.
- Correct expectation 0.5pt.
- Cannot conclude the same distribution 0.5pt.
- **OR:** Correctly solved the differential equation/took the derivative and concluded the result 1pt.
- Cannot conclude the same distribution 0.5pt.

Ex 11.63 (1)**.** Let B be as you found it in part (a). Find the CDF of $X = \kappa + \lambda \ln\left(\frac{B}{1-B}\right)$.

s.11.63. As usual, we start with the CDF of B, this is known to be $F_B(y) = y$ for $y \in [0,1]$ (and 1 for $y > 1$, 0 for $y < 0$). Then we have that

$$\begin{aligned}
F_X(y) &= \mathsf{P}\left\{X \le y\right\} \\
&= \mathsf{P}\left\{\kappa + \lambda \ln\left(\frac{B}{1-B}\right) \le y\right\} \\
&= \mathsf{P}\left\{\ln\left(\frac{B}{1-B}\right) \le \frac{y-\kappa}{\lambda}\right\} \\
&= \mathsf{P}\left\{B \le \frac{\exp\frac{y-\kappa}{\lambda}}{1+\exp\frac{y-\kappa}{\lambda}}\right\} \\
&= F_B\left(\frac{\exp\frac{y-\kappa}{\lambda}}{1+\exp\frac{y-\kappa}{\lambda}}\right) \\
&= \frac{\exp\frac{y-\kappa}{\lambda}}{1+\exp\frac{y-\kappa}{\lambda}}
\end{aligned}$$

for $y \in \mathbf{R}$.

Grading scheme:

- CDF technique derivation 0.5pt.
- No mistakes 0.5pt.

Ex 11.64 (1)**.** Let $\kappa = 0$, $\lambda = 1$. The quantile function $Q_X(\cdot)$ is defined to be the function such that $Q_X(F_X(x)) = x$. Find $Q_X(\cdot)$ for the random variable X as in part (c). You may assume $F_X(x)$ to be strictly increasing without proof. This function Q_X is known as the 'log-odds', or 'logit' function, and is used often in regression analysis to model a binary random variable.

s.11.64. Notice that F_X maps the real line onto the interval $(0,1)$. Then, Q_X must map the interval $(0,1)$ onto $\mathbf{R}$. We find the inverse of the CDF of X as follows:

$$\begin{aligned}
&z = F_X(y) \implies \\
&z = \frac{e^y}{1+e^y} \implies \\
&z = \frac{1}{1+e^{-y}} \implies \\
&e^{-y} z = 1 - z \implies \\
&e^y = \frac{z}{1-z} \implies \\
&y = \ln\frac{z}{1-z} = Q_X(z)
\end{aligned}$$

for $z \in (0,1)$.

Grading scheme:

- Mention the idea to invert the CDF 0.5pt. (of course this includes the people who did so)
- Correct inversion 0.5pt.
- Correct bounds for the quantile function 0.5pt bonus.

QUESTION

Consider the following code:

Python Code

```python
import numpy as np
from scipy.stats import expon
np.random.seed(42)

n = 600
N = 250

X = expon(scale = 1/4).rvs([n, N])
Y = X.mean(axis = 1)

mu = 1/4
sigma = 1/4
Z = np.sqrt(N) * (Y - mu)/sigma

print((Z ** 53).mean())
```

R Code

```r
set.seed(42)

n <- 600
N <- 250

X <- matrix(rexp(N * n, rate = 4), nrow = n, ncol = N)
Y <- rowMeans(X)

mu <- 1/4
sigma <- 1/4
Z <- sqrt(N) * (Y - mu)/sigma

print(mean(Z^53))
```

Ex 11.65 (1.5)**.** Y is a vector of i.i.d. random variables.

(i) What is the length ℓ of Y?

(ii) Each element of Y is a mean of k i.i.d. Exp(λ) r.v.s. What are k and λ?

(iii) What are the (theoretical) expectation and variance of an element of Y?

s.11.65. The length of Y is $n = 600$; Each element of Y is a mean of $k = N = 250$ i.i.d. Exp(4) r.v.s. The expectation is $\frac{1}{\lambda} = \frac{1}{4}$ and the variance is $\frac{1}{k\lambda^2} = \frac{1}{4000} = 0.00025$.

Grading scheme:

- 0.5 for getting both the length $n = 600$ and $k = 250$ correct (no partial credit);
- 0.5 for $\lambda = 4$, expectation $\frac{1}{4}$ and the factor $\frac{1}{16}$ in the variance (no partial credit);
- 0.5 for the factor $\frac{1}{k}$ in the variance.

Recall that each element of Y is the mean of k i.i.d. Exp(λ) r.v.s.

Ex 11.66 (1)**.** What is the exact distribution of an element of Y?
Give its name and its parameters, and explain the answer.

s.11.66. The sum of k i.i.d. Exp(λ) r.v.s. has the Gamma(k, λ) distribution. Hence, the mean of k i.i.d. Exp(λ) r.v.s. has the Gamma($k, k\lambda$) distribution. In this exercise, $k = 250$ and $k\lambda = 1000$.
Grading scheme:

- 0.5 for Gamma with first parameter k
- 0.5 for the second parameter

Let $(Y_1, \dots, Y_\ell)$ be the elements of Y and let $(Z_1, \dots, Z_\ell)$ be the elements of Z. Recall that each Z_i depends on k because Y_i is the mean of k i.i.d. Exp(λ) r.v.s. Let T be the random variable to which Z_1 converges in the limit $k \to \infty$.

Ex 11.67 (1)**.** What is the distribution of T, and why?
Hence, what is an approximate distribution of an element of Y (e.g. Y_1) and why?

s.11.67. By the CLT, $Z_1 \stackrel{\cdot}{\sim} \text{Norm}(0, 1)$. Hence, $Y_1 = \mu + \sigma/\sqrt{n} Z_1 \stackrel{\cdot}{\sim} \text{Norm}(0.25, 0.0625/n)$.
Grading scheme:

- 0.5 for mentioning CLT and the distribution of Z_1;
- 0.5 for the approximate distribution of Y_1.

Ex 11.68 (0.5)**.** In the last line of the code, we compute $S = \frac{1}{\ell}\sum_{i=1}^{\ell} Z_i^{53}$.
If $k \to \infty$ (for fixed ℓ, e.g. $\ell = 3$), does S converge to a constant?
You do not have to give formal proofs in this subquestion, but you should give clear explanations using the appropriate theorem(s).

s.11.68. In the limit $k \to \infty$, each Z_i has the standard normal distribution by CLT, but the sum of ℓ powers of normal distributions has a non-trivial CDF.

- 0.5 for explaining that S does not converge to a constant; no points for just CLT (unless previous question was not answered).

Ex 11.69 (1)**.** In the last line of the code, we compute $S = \frac{1}{\ell}\sum_{i=1}^{\ell} Z_i^{53}$.

If $\ell \to \infty$ (for fixed k), does S converge to a constant? If so, does it converge to $\mathsf{E}\left[T^{53}\right]$?

You do not have to give formal proofs in this subquestion, but you should give clear explanations using the appropriate theorem(s).

s.11.69. By LLN, S does converge to a constant as $\ell \to \infty$, however, it converges to $\mathsf{E}\left[Z_1^{53}\right]$ for that fixed value of k. By symmetry, we have $\mathsf{E}\left[Z_1^{53}\right] = 0$. However, the gamma distribution is right-skewed, which implies $\mathsf{E}\left[T^{53}\right] > 0$. Hence, it does not converge to $\mathsf{E}\left[T^{53}\right]$.
Grading scheme:

- 0.5 for concluding that S converges to a constant using LLN.
- 0.5 for explaining why the constant is not equal to $\mathsf{E}\left[T^{53}\right]$.

QUESTION

We have a queue of people served by a single server. Let $L(t)$ be the number of people present in the system at time t. For any time $t \geq 0$ the time to the next arriving person is $X \sim \text{Exp}(\lambda)$ and, when $L(t) > 0$, the time to the next departing customer is $S \sim \text{Exp}(\mu)$. Assume that $\lambda < \mu$.

Suppose $L(0) = n \geq 0$. Then let T be the first time until the system becomes empty, i.e., $T = \inf\{t \geq 0 : L(t) = 0\}$.

Ex 11.70 (1)**.** Explain that $\lambda/(\lambda+\mu)$ is the probability an arrival occurs before a departure.

s.11.70. Standard consequence of exponential rvs.

Grading:

- The time to the next arrival is not λ. -0.5.

For the moment, assume that $\mathsf{E}[T] < \infty$.

Ex 11.71 (1)**.** Explain that for $n > 0$:

$$\mathsf{E}[T|L(0) = n] = \mathsf{E}[T|L(0) = n+1]\frac{\lambda}{\lambda+\mu} + \mathsf{E}[T|L(0) = n-1]\frac{\mu}{\lambda+\mu} + \frac{1}{\lambda+\mu}. \tag{11.0.137}$$

s.11.71. Use the memoryless property of the exp distribution. When a job arrives first, we can model this as if we start from $n+1$ until we hit 0. Likewise, when a job leaves first, we start from $n-1$. The last term is the expected time until an event happens.

Grading:

- Some people write $P(S = X)$ and give this a positive probability. That is a grave mistake: -0.5.

Ex 11.72 (1)**.** Show that $\mathsf{E}[T|L(0) = n] = n/(\mu-\lambda)$.

s.11.72. Just fill in the expression in the previous exercise and check that the RHS and LHS match.

Define $\rho = \lambda/\mu$. Assume that $L(0) \sim \text{Geo}(1-\rho)$.

Ex 11.73 (1)**.** Find a simple expression for $\mathsf{E}[T]$.

s.11.73. Use conditioning on L.

$$\mathsf{E}[T] = \mathsf{E}[\mathsf{E}[T|L(0)]] = \mathsf{E}\left[L(0)/(\mu-\lambda)\right] = \frac{\rho}{1-\rho}\frac{1}{\mu-\lambda} = \frac{\lambda}{\mu^2(1-\rho)^2}.$$

Ex 11.74 (1)**.** Up to now we simply assumed that $\mathsf{E}[T] < \infty$. Motivate intuitively that the condition $\lambda < \mu$ ensures that $\mathsf{E}[T] < \infty$.

s.11.74. If $\lambda > \mu$, jobs arrive faster than they can be served. In such cases the queueing process drifts to infinity, in expectation.

The case $\lambda = \mu$ is difficult, and I don't expect you to discuss this.

QUESTION

Bob and his father argue about tomorrow's weather. Bob thinks it will rain, but his father doesn't agree. They make the following deal. Bob will put down a glass in the back yard. At the end of the day, dad will give Bob one euro for every inch of water in the glass. To be safe, dad gives Bob a shot glass with a height of only one inch to put down in the back yard.

Let X denote the amount of rainfall tomorrow (in inches) and let Y denote the amount of rain collected in the shot glass (in inches). We assume that X has the following pdf:

$$f_X(x) = \frac{3}{4}x(2-x), \quad 0 \leq x \leq 2. \tag{11.0.138}$$

Ex 11.75 (1)**.** Compute the probability that the amount of rainfall tomorrow is at most one inch.

s.11.75. We have

$$\begin{aligned} \mathsf{P}\{X \leq 1\} &= \int_0^1 \frac{3}{4}x(2-x)\,dx && (11.0.139)\\ &= \frac{3}{4}\left[x^2 - \frac{1}{3}x^3\right]_0^1 dx && (11.0.140)\\ &= \frac{3}{4}\left(1-\frac{1}{3}\right) && (11.0.141)\\ &= 1/2. && (11.0.142) \end{aligned}$$

An argument based on the fact that f_X is a parabola centered around 1 is also fine.

Correct initial integral: 0.5 point.
Correct solution: 0.5 point.

Ex 11.76 (1.5)**.** Determine the expected amount of rain collected in the shot glass conditional on the amount of rainfall X being at most one inch.

s.11.76. Note that $Y = \min\{X, 1\}$. We find

$$\begin{aligned}
\mathsf{E}[Y|X \le 1] &= \mathsf{E}[X|X \le 1] && (11.0.143)\\
&= \int_0^1 x f_X(x|X \le 1)\,dx && (11.0.144)\\
&= \int_0^1 x \frac{f_X(x)}{\mathsf{P}\{X \le 1\}}\,dx && (11.0.145)\\
&= 2\int_0^1 x \frac{3}{4} x(2-x)\,dx && (11.0.146)\\
&= \frac{3}{2}\int_0^1 (2x^2 - x^3)\,dx && (11.0.147)\\
&= \frac{3}{2}\left[\frac{2}{3}x^3 - \frac{1}{4}x^4\right]_0^1 && (11.0.148)\\
&= \frac{3}{2}\left(\frac{2}{3} - \frac{1}{4}\right) && (11.0.149)\\
&= \frac{5}{8}. && (11.0.150)
\end{aligned}$$

Writing down $\mathsf{E}[X|X \le 1]$: 0.5 point.
Correct derivation and solution: 1 point.

Ex 11.77 (1.5)**.** What is the (unconditional) expected amount of rain collected in the shot glass?

s.11.77. By the law of total expectation,

$$\begin{aligned}
\mathsf{E}[Y] &= \mathsf{P}\{X \le 1\}\,\mathsf{E}[Y|X \le 1] + \mathsf{P}\{X > 1\}\,\mathsf{E}[Y|X > 1] && (11.0.151)\\
&= \frac{1}{2}\frac{5}{8} + \frac{1}{2} \cdot 1 && (11.0.152)\\
&= \frac{13}{16}. && (11.0.153)
\end{aligned}$$

Mentioning law of total expectation: 0.5 point.
Correctly using law of total expectation: 0.5 point.
Correct solution: 0.5 point.

To make things more interesting, dad decides to randomize the amount of euros he will give to Bob. Given an amount $Y = y$ collected in the cup, he will pay Bob an amount Z that is uniformly distributed on $[y, 2y]$.

Ex 11.78 (1)**.** What is the (unconditional) expected value of the payout Z?

s.11.78. By Adam's law,

$$\mathsf{E}[Z] = \mathsf{E}[\mathsf{E}[Z|Y]] \tag{11.0.154}$$
$$= \mathsf{E}\left[\frac{3}{2}Y\right] \tag{11.0.155}$$
$$= \frac{3}{2}\mathsf{E}[Y] \tag{11.0.156}$$
$$= \frac{3}{2}\cdot\frac{13}{16} = \frac{39}{32}. \tag{11.0.157}$$

Mentioning and using Adam's law: 0.5 point.
Correct solution: 0.5 point.

Note: if in your interpretation, $Y = y$ is actually given (I realized that the question is ambiguous), then $\frac{3}{2}y$ (lowercase!) will also be regarded as correct.

QUESTION

Consider the following code:

Python Code

```python
import numpy as np
from scipy.stats import expon
np.random.seed(42)

n = 200
N = 500

X = expon(scale = 2).rvs([N, n])
Y = X.mean(axis = 1)

mu = 2
sigma = 2
Z = np.sqrt(n) * (Y - mu)/sigma

print((Z ** 29).mean())
```

R Code

```r
set.seed(42)

n <- 200
N <- 500

X <- matrix(rexp(n * N, rate = 1/2), nrow = N, ncol = n)
Y <- rowMeans(X)

mu <- 2
sigma <- 2
Z <- sqrt(n) * (Y - mu)/sigma

print(mean(Z^29))
```

Ex 11.79 (1.5)**.** Y is a vector of i.i.d. random variables.

(i) What is the length ℓ of Y?

(ii) Each element of Y is a mean of k i.i.d. Exp(λ) r.v.s. What are k and λ?

(iii) What are the (theoretical) expectation and variance of an element of Y?

s.11.79. The length of Y is $N = 500$; Each element of Y is a mean of $k = n = 200$ i.i.d. Exp(0.5) r.v.s. The expectation is $\frac{1}{\lambda} = 2$ and the variance is $\frac{1}{k\lambda^2} = \frac{1}{200 \cdot 1/4} = 0.02$.

Grading scheme:

- 0.5 for getting both the length $N = 500$ and $k = 200$ correct (no partial credit);
- 0.5 for $\lambda = 0.5$, expectation 2 and the factor $\frac{1}{\lambda^2} = 4$ in the variance (no partial credit);
- 0.5 for the factor $\frac{1}{k}$ in the variance.

Recall that each element of Y is the mean of k i.i.d. Exp(λ) r.v.s.

Ex 11.80 (1)**.** What is the exact distribution of an element of Y?
Give its name and its parameters, and explain the answer.

s.11.80. The sum of k i.i.d. Exp(λ) r.v.s. has the Gamma(k,λ) distribution. Hence, the mean of k i.i.d. Exp(λ) r.v.s. has the Gamma$(k, k\lambda)$ distribution. In this exercise, $k = 200$ and $k\lambda = 100$.
Grading scheme:

- 0.5 for Gamma with first parameter k
- 0.5 for the second parameter

Let $(Y_1,\dots,Y_\ell)$ be the elements of Y and let $(Z_1,\dots,Z_\ell)$ be the elements of Z. Recall that each Z_i depends on k because Y_i is the mean of k i.i.d. Exp(λ) r.v.s. Let T be the random variable to which Z_1 converges in the limit $k \to \infty$.

Ex 11.81 (1)**.** What is the distribution of T, and why?
Hence, what is an approximate distribution of an element of Y (e.g. Y_1) and why?

s.11.81. By the CLT, $Z_1 \dot{\sim}$ Norm$(0,1)$. Hence, $Y_1 = \mu + \sigma/\sqrt{n} Z_1 \dot{\sim}$ Norm$(2, 4/n)$.
Grading scheme:

- 0.5 for mentioning CLT and the distribution of Z_1;
- 0.5 for the approximate distribution of Y_1.

Ex 11.82 (0.5)**.** In the last line of the code, we compute $S = \frac{1}{\ell}\sum_{i=1}^{\ell} Z_i^{29}$.
If $k \to \infty$ (for fixed ℓ, e.g. $\ell = 3$), does S converge to a constant?
You do not have to give formal proofs in this subquestion, but you should give clear explanations using the appropriate theorem(s).

s.11.82. In the limit $k \to \infty$, each Z_i has the standard normal distribution by CLT, but the sum of ℓ powers of normal distributions has a non-trivial CDF.

- 0.5 for explaining that S does not converge to a constant; no points for just CLT (unless previous question was not answered).

Ex 11.83 (1)**.** In the last line of the code, we compute $S = \frac{1}{\ell}\sum_{i=1}^{\ell} Z_i^{29}$.

If $\ell \to \infty$ (for fixed k), does S converge to a constant? If so, does it converge to $\mathsf{E}\left[T^{29}\right]$?

You do not have to give formal proofs in this subquestion, but you should give clear explanations using the appropriate theorem(s).

s.11.83. By LLN, S does converge to a constant as $\ell \to \infty$, however, it converges to $\mathsf{E}\left[Z_1^{29}\right]$ for that fixed value of k. By symmetry, we have $\mathsf{E}\left[Z_1^{29}\right] = 0$. However, the gamma distribution is right-skewed, which implies $\mathsf{E}\left[T^{29}\right] > 0$. Hence, it does not converge to $\mathsf{E}\left[T^{29}\right]$.
Grading scheme:

- 0.5 for concluding that S converges to a constant using LLN.
- 0.5 for explaining why the constant is not equal to $\mathsf{E}\left[T^{29}\right]$.

QUESTION

There are three cheese shops in town, shop A, shop B and shop C. N_1 customers enter shop A per hour, $N_1 \sim$ Poisson(λ_1); N_2 customers enter shop B per hour, $N_2 \sim$ Poisson(λ_2); N_3 customers enter shop C per hour, $N_3 \sim$ Poisson(λ_3). N_1, N_2, N_3 are independent of each other. Each customer that enters in any of the three shops, buys cheese with probability p, independently. Let X be the total number of customers that buys cheese per hour.

Ex 11.84 (1)**.** Show that $N_1 + N_2$ and $N_2 - 2N_3$ are **not** independent of each other.

s.11.84.

$$\begin{aligned}
&\mathsf{Cov}\left[N_1 + N_2, N_2 - 2N_3\right] \\
=&\mathsf{Cov}\left[N_1, N_2\right] - \mathsf{Cov}\left[N_1, 2N_3\right] + \mathsf{Cov}\left[N_2, N_2\right] - \mathsf{Cov}\left[N_2, 2N_3\right] \cdots \text{(0.5 point)} \\
=&Var(N_2) \\
=&\lambda_2 > 0. \cdots \text{(0.5 point)}
\end{aligned}$$

Ex 11.85 (1)**.** Let $N = N_1 + N_2 + N_3$. What is the distribution of N? What is the distribution of $X|N$? What is the distribution of X?

s.11.85. Since N_1, N_2 and N_3 are independent, $N \sim Pois(\lambda_1 + \lambda_2 + \lambda_3)$.(0.5 point)
$X|N \sim Bin(N, p)$, and $X \sim Pois((\lambda_1 + \lambda_2 + \lambda_3)p)$ by the Chicken-egg theory. (0.5 point)

Ex 11.86 (2)**.** Calculate $\rho_{X,N}$, the correlation between X and N.

s.11.86. Let $Y = N - X$ be the number of customers that do not buy cheese. Then we know $Y \sim Pois((\lambda_1 + \lambda_2 + \lambda_3)q)$ with $q = 1 - p$, and X and Y are independent. (0.5 point)

$$\begin{aligned}
\mathsf{Cov}\left[N, X\right] &= \mathsf{Cov}\left[X + Y, X\right] \\
&= \mathsf{Cov}\left[X, X\right] + \mathsf{Cov}\left[Y, X\right] \\
&= Var(X) \cdots \text{(0.5 point)} \\
&= (\lambda_1 + \lambda_2 + \lambda_3)p \cdots \text{(0.5 point)}
\end{aligned}$$

Then it follows that

$$\begin{aligned}
\rho_{X,N} &= \frac{\mathsf{Cov}\left[N, X\right]}{sd(N) \cdot sd(X)} \\
&= \frac{(\lambda_1 + \lambda_2 + \lambda_3)p}{\sqrt{\lambda_1 + \lambda_2 + \lambda_3} \cdot \sqrt{(\lambda_1 + \lambda_2 + \lambda_3)p}} \\
&= \sqrt{p}. \cdots \text{(0.5 point)}
\end{aligned}$$

Consider the following codes:

R Code

```
1 library(mvtnorm)
2 set.seed(777)
3 A<-matrix(c(1,2,2,4),nrow = 2)
4 B<-c(1,2)
5 X<-rmvnorm(50,mean=B,sigma=A)
6 output=cor(X[3,],X[41,])
```

Python Code

```
1 import random
2 import numpy as np
3 random.seed(777)
4 A = [[1,2],[2,4]]
5 B = [1,2]
6 X = np.random.multivariate_normal(B, A, size = 50)
7 output = np.corrcoef(X[[3,41],:].reshape(4))
```

Ex 11.87 (1)**.** Explain in detail the purpose of **each line** of the above codes.

s.11.87. Line 1: Load the package "mvtnorm" so that we can use the function *rmvnorm*.
Line 2: Set a random seed to reproduce the same results.
Line 3: Generate a 2×2 identity matrix A=$\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$.
Line 4: Generate a vector $B = (1,2)$.
Line 5: Generate 50 multivariate normally distributed variables $\mathbf{X} = x_1, x_2$ with mean equal to B and variance equal to A.
Line 6: Calculate the correlation of the 3rd and 41th rows of matrix X.
(0.5 points for mentioning at least 3 of the above.)

QUESTION

A server spends a random amount T on a job. While the server works on the job, it sometimes gets interrupted to do other tasks $\{R_i\}$. Such tasks need to be repaired before the machine can continue working again. It is clear that such interruptions can only occur when the server is working. Assume that the interruptions arrive according to a Poisson process with rate λ, i.e., the number of failures $N(t) \sim \text{Pois}(\lambda t)$. The interruptions $\{R_i\}$ are independent of N and of T and form an iid sequence with common mean $\mathsf{E}[R]$ and variance $\mathsf{V}[R]$. The duration S of a job is its own service time T plus all interruptions, hence $S = T + \sum_{i=1}^{N(T)} R_i$.

Ex 11.88 (1.5)**.** The computation below consists of a number of steps, a, b, ... Explain for each step which property is used to ensure the step is true.

$$\mathsf{E}[S \mid T = t] \stackrel{a}{=} \mathsf{E}\left[T + \sum_{i=1}^{N(T)} R_i \,\middle|\, T = t\right] \tag{11.0.158}$$

$$\stackrel{b}{=} \mathsf{E}[T \mid T = t] + \mathsf{E}\left[\sum_{i=1}^{N(T)} R_i \,\middle|\, T = t\right] \tag{11.0.159}$$

$$\stackrel{c}{=} \mathsf{E}[t \mid T = t] + \mathsf{E}\left[\sum_{i=1}^{N(T)} R_i \,\middle|\, T = t\right] \tag{11.0.160}$$

$$\stackrel{d}{=} t + \mathsf{E}\left[\sum_{i=1}^{N(t)} R_i\right]. \tag{11.0.161}$$

s.11.88. a. substitute the definition of S

b. Linearity of expectation

c. On the set $\{T = t\}$, $T = t$. Hence we can replace T by t.

d. On the set $\{T = t\}$, $T = t$. Hence we can replace T by t. And the (conditional) expectation of a constant is that constant.
Each property missed, e.g, linearity of expectation, minus 0.5.

Ex 11.89 (2)**.** Suppose $R \sim \text{Exp}(\mu)$ and $\mathsf{P}\{T = t\} = 1$, compute $\mathsf{E}[S]$.

s.11.89. We can use the result of part 1. Since $\mathsf{P}\{T = t\} = 1$, $\mathsf{E}[R] = 1/\mu$, $\mathsf{E}[N(t)] = \lambda t$, and indepence of R_i and N, and R_i iid,

$$\begin{aligned}
\mathsf{E}[S] &= t + \mathsf{E}\left[\sum_{i=1}^{N(t)} R_i\right] \\
&= t + \mathsf{E}\left[\mathsf{E}\left[\sum_{i=1}^{N(t)} R_i \,\middle|\, N(t)\right]\right] \\
&= t + \mathsf{E}[N(t)\,\mathsf{E}[R]] \\
&= t + \mathsf{E}[N(t)]\,\mathsf{E}[R] = t + \lambda t/\mu.
\end{aligned}$$

Typically, 0.5 point for each correct line, but I subtracted points If you say that $\mathsf{E}[N(T)\mathsf{E}[R]] = \mathsf{E}[R]$, or write $n\mathsf{E}[R]$ as final answer (apparently you did not get the idea that N is an rv.)

Ex 11.90 (1.5)**.** Explain what this code computes.

Python Code

```python
import numpy as np

labda = 0.5
size = 10
num_runs = 50


def do_run():
    T = np.random.uniform(0, 20)
    N = np.random.poisson(labda * T)
    R = np.random.uniform(1, 5, size=N)
    S = T + R.sum()
    return S


samples = np.zeros(num_runs)
for i in range(num_runs):
    samples[i] = do_run()


print(samples[samples > 4].mean())
```

R Code

```r
labda <- 0.5
size <- 10
num_runs <- 50

do_run <- function() {
  bigT <- runif(n = 1, min = 0, max = 20)
  N <- rpois(n = 1, labda * bigT)
  R <- runif(n = N, min = 1, max = 5)
  S <- bigT + sum(R)
  return(S)
}

samples <- rep(0, num_runs)
for (i in 1:num_runs) {
  samples[i] <- do_run()
}

print(mean(samples[samples > 4]))
```

Hint, you should know that in P.21 (R18) the string `samples > 4` collects only the samples with value larger than 4.

s.11.90. Here is an EXPLANATION:

1. T is a simulated service time of a job without interruptions.
2. N is is the simulated number of failures that occur during the service of the job
3. R is then a vector of simulated durations of each of the interruptions
4. Hence, S is the total time the job spends at the server
5. We sample the total service time a number of times, and count how often (mean or variance) this exceeds a threshold.

The points earned depends on how clear you explanation is.

Quite a few of you don't seem to understand what it means to *explain* something. Here are some answers that I saw that certainly don't explain what the code does:

1. 'The code does what's stated in the exercise.'. What's the explanation here? The question is also not: do you understood what the code does?
2. 'T is uniform rv, N is a Poisson rv, R is a also a uniform rv. We repeat this a few times.' Like this you just read the code, but I know you can read, so this type of answer is quite useless.
3. 'We collect a subset of the samples and print that.' This answer is just a repition of the hint.

So, next time, when you are asked to explain something, explain it. In your professional carreer, if you read out loud to your customer what s/he can read for him/herself, you'll be fired pretty soon.

QUESTION

Let $Y \sim \text{Norm}(0,1)$. In this exercise, we find an upper bound for $\mathsf{P}\{|Y| > 3\}$.

Ex 11.91 (1.5)**.** Let g be a positive and increasing function, and let Z be a r.v.
Consider the following inequality:

$$\mathsf{P}\{Z \geq a\} = \mathsf{P}\left\{g(Z) \geq g(a)\right\} \leq \frac{\mathsf{E}\left[g(Z)\right]}{g(a)}.$$

(i) Explain why $\mathsf{P}\{Z \geq a\} = \mathsf{P}\left\{g(Z) \geq g(a)\right\}$ holds.

(ii) Explain why $\mathsf{P}\left\{g(Z) \geq g(a)\right\} \leq \frac{\mathsf{E}[g(Z)]}{g(a)}$ holds.

Make sure to clearly indicate where you use that g is positive and increasing.

s.11.91. (i). Since g is increasing, we have $Z \geq a$ if and only if $g(Z) \geq g(a)$, so $\{Z \geq a\}$ and $\{g(Z) \geq g(a)\}$ are the same event. Hence, $\mathsf{P}\{Z \geq a\} = \mathsf{P}\left\{g(Z) \geq g(a)\right\}$.
(ii). Since g is positive, we have $|g(Z)| = g(Z)$ and $g(a) > 0$. Hence, the inequality follows directly from Markov's inequality with r.v. $g(Z)$ and constant $g(a) > 0$.

Remarks and grading scheme:

- In part (ii), it is not sufficient to just say that it holds by Markov's inequality, also explain how you apply Markov's inequality.
- Since it was asked to indicate **clearly** where you use that g is positive and increasing, don't just say "since g is positive and increasing" twice. For part (i) the positivity is not needed, it is only needed that g is increasing. For part (ii), it is not needed that g is increasing, only the positivity is needed. Moreover, some explanation should be given how it is used.
- g is not necessarily a linear transformation, as some of you claimed.
- Not every positive and increasing function is convex or concave.
And using Jensen's inequality certainly doesn't work in (i): you are asked to prove an equality of probabilities, not an inequality of expectations.
- In part (i), the **if and only if** is really important. If you just mention that $g(Z) \geq g(a)$ if $Z \geq a$, then you are only proving that $\mathsf{P}\{Z \geq a\} \leq \mathsf{P}\left\{g(X) \geq g(a)\right\}$, because if you just say "$g(Z) \geq g(a)$ if $Z \geq a$", $g(Z) \geq g(a)$ could still be true in cases where $Z \geq a$ is not, and hence $g(Z) \geq g(a)$ can still have a larger probability. (I've not been strict on this when grading.)
- Grading: 0.5 for a sufficient explanation for (i), 0.5 for a sufficient explanation for (ii) and 0.5 for clearly indicating where it is used that g is positive and increasing and writing a clear answer overall.

Ex 11.92 (1)**.** Prove that $\mathsf{P}\{|Y| > 3\} \leq e^{-9t}\,\mathsf{E}\left[e^{tY^2}\right]$ for $t > 0$.

s.11.92. Note that $g(x) = e^{tx^2}$ is positive and increasing on $(0,\infty)$ for $t > 0$. By applying the inequality of the first question with $a = 3$ we find

$$\mathsf{P}\{|Y| > 3\} \leq e^{-9t}\,\mathsf{E}\left[e^{t|Y|^2}\right] = e^{-9t}\,\mathsf{E}\left[e^{tY^2}\right].$$

Remarks and grading scheme:

- Most of you did not apply the inequality from the previous part to solve this question, but instead used Chernoff's inequality as follows:

$$\mathsf{P}\{|Z| > 3\} = \mathsf{P}\left\{Z^2 > 9\right\} \leq e^{-9t}\,\mathsf{E}\left[e^{t|Z|^2}\right] = e^{-9t}\,\mathsf{E}\left[e^{tZ^2}\right],$$

  where the first equality holds since $|Z| > 3$ if and only if $\mathsf{P}\left\{Z^2 > 9\right\}$. This is also correct, but takes a bit more time.

- Don't write nonsense like $e^{-3t}\,\mathsf{E}\left[e^{t|Z|}\right] = e^{-9t}\,\mathsf{E}\left[e^{tZ^2}\right]$, just to make it look like you solved the exercise although you didn't.

- Grading scheme: 1 point for solving the exercise. Partial credit (0.5) for a correct and relevant application of Chernoff's inequality.

Ex 11.93 (2.5)**.** For which t do we find the best upper bound for $\mathsf{P}\{|Y| > 3\}$? Also calculate the upper bound for this value of t.

Hint 1. You may use that if $X \sim \chi^2_1$, then the MGF of X is given by $M_X(t) = (1-2t)^{-1/2}$ for $t < 1/2$. However, you should explain clearly how you use this fact.

Hint 2. Do not forget to check the second order condition of minimization.

s.11.93. Since $Y^2 \sim \chi^2_1$, we have $\mathsf{E}\left[e^{tY^2}\right] = \mathsf{E}\left[e^{tX}\right] = M_X(t) = (1-2t)^{-1/2}$.

So we minimize $e^{-9t}\,\mathsf{E}\left[e^{tY^2}\right] = e^{-9t}(1-2t)^{-1/2}$. It is easier if we take the logarithm first and minimize $-9t - \frac{1}{2}\log(1-2t)$. Its derivative to t is $-9 + \frac{1}{1-2t}$, so setting the derivative to 0 yields $t = 4/9$. The second derivative to t is $\frac{2}{(1-2t)^2} > 0$ (the value at $t = 4/9$ is 162), so the second order condition holds.

This yields $\mathsf{P}\{|Y| > 3\} \leq e^{-4}(1-8/9)^{-1/2} \approx 0.0549$.

Remarks and grading scheme:

- Unless explicitly noted, you have to give an analytical answer and derive everything using just pen and paper.

- It is essential that you include some derivations for computing the (second) derivative.

- It is not necessary to take the logarithm first, but it makes the derivation easier, and hence it is less likely that you make errors.

- Grading scheme: 0.5 for arguing that $\mathsf{E}\left[e^{tZ^2}\right] = (1-2t)^{-1/2}$ with sufficient explanation; 1 for taking the derivative and calculating that it is 0 at $t = 4/9$; (0.5 if small mistake is made but resulting t satisfies $0 < t < 1/2$, or if it is remarked that this should be the case); 0.5 for calculating the second derivative and checking the second order condition; 0.5 for filling in $t = 4/9$ to provide the upper bound (if an incorrect value of t is found, this point can be given only if the resulting bound is between 0.0001 and 1, or if it is explicitly noted that the answer does not make sense).

QUESTION

John likes to spend his time watching trains pass by on the railway near his house. John is interested in the *interarrival times* of the trains: the time between the arrivals of two subsequent trains. John knows that the interarrival times X_i, $i = 1, \ldots, n$, are i.i.d. Exponentially distributed with a rate parameter λ. Hence, given the value of λ, the pdf of interarrival time X_i is

$$f_{X_i|\lambda}(x|\lambda) = \lambda e^{-\lambda x}, \qquad x > 0. \tag{11.0.162}$$

John is interested in the value of λ. His prior belief about the distribution of λ is that it follows a Gamma(a, b) distribution with some particular choices for a and b (the exact values of a and b are not relevant for this question).

Ex 11.94 (2.5)**.** Suppose that John observes a first interarrival time of $X_1 = x_1$. Derive John's *posterior* distribution of λ.

s.11.94. By Bayes' rule we have

$$\begin{aligned} f_1(\lambda|X_1 = x_1) &= \frac{f_{X_1|\lambda}(x_1|\lambda) f_0(\lambda)}{f_{X_1}(x_1)} && (11.0.163)\\ &\propto f_{X_1|\lambda}(x_1|\lambda) f_0(\lambda) && (11.0.164)\\ &= \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda} \lambda e^{-\lambda x_1} && (11.0.165)\\ &\propto \lambda^a e^{-(b+x_1)\lambda}, && (11.0.166) \end{aligned}$$

in which we recognize the pdf of a Gamma$(a+1, b+x_1)$ distribution (up to a scaling constant). Hence, the posterior distribution λ given $X_1 = x_1$ is Gamma$(a+1, b+x_1)$.

Applying Bayes' rule: 1 point.
Finding the expression for the posterior: 1 point.
recognizing a $Gamma(a+1, b+x_1)$ dist: 0.5 point.

Ex 11.95 (1)**.** Is John's prior distribution a *conjugate* prior?

s.11.95. Yes. The posterior distribution is in the same family of distributions (Gamma) as the prior. Hence, John has a conjugate prior.

Correct answer and motivation: 1 point.

Ex 11.96 (1.5)**.** Suppose John observes the first n interarrival times, with values $X_1 = x_1, \ldots, X_n = x_n$. What is John's posterior distribution after these observations?
Hint: you don't need to redo all the math here!

s.11.96. The posterior after observing $X_1 = x_1$ becomes our new prior. Hence, our new prior is a Gamma$(a+1, b+x_1)$ distribution. From question 1 it follows that the prior after observing $X_2 = x_2$ then is a Gamma$(a+2, b+x_1+x_2)$ distribution. Hence, iterating this process, we find that the posterior distribution of λ after observing $X_1 = x_1, \ldots, X_n = x_n$ is a Gamma$(a+n, b+\sum_{i=1}^n x_i)$ distribution.

Noting that the posterior becomes the new prior: 0.5 point.
Correct derivation and answer: 1 point.

QUESTION

A mouse is trapped in a pit with three tunnels. When the mouse takes tunnel A, the time to get out of the pit is 2 minutes. Tunnel B leads back to the pit (in other words, the mouse cannot escape when it takes tunnel B) and takes 3 minutes. Tunnel C leads also back to the pit, and takes 4 minutes. Every time the mouse is in the pit, it selects a tunnel at random with equal probability. (This mouse much dumber than a real mouse.) Write X for the tunnel selected by the mouse, and let T be the time until the mouse escapes. The travel times of the tunnels are constant.

For the moment, assume that $\mathsf{E}[T] < \infty$.

Ex 11.97 (1)**.** Explain that $\mathsf{E}[T|X=B] = 3 + \mathsf{E}[T]$.

s.11.97. After selecting tunnel B, which takes 3 minutes to travel, the mouse is back in the pit again, and the process starts over again.

Ex 11.98 (1)**.** Compute $\mathsf{E}[T]$.

s.11.98.

$$\mathsf{E}[T] = \mathsf{E}[T|X=A]/3 + \mathsf{E}[T|X=B]/3 + \mathsf{E}[T|X=C]/3. \tag{11.0.167}$$
$$\mathsf{E}[T|X=A] = 2 \tag{11.0.168}$$
$$\mathsf{E}[T|X=B] = 3 + \mathsf{E}[T] \tag{11.0.169}$$
$$\mathsf{E}[T|X=C] = 4 + \mathsf{E}[T]. \tag{11.0.170}$$

Solving gives $\mathsf{E}[T] = 9$.

Grading

- Not using the result of subquestion 1: no points.

Ex 11.99 (2)**.** Compute $\mathsf{V}[T]$.

s.11.99.

$$\mathsf{V}[T|X=A] = 0 \tag{11.0.171}$$
$$\mathsf{V}[T|X=B] = \mathsf{V}[T] \tag{11.0.172}$$
$$\mathsf{V}[T|X=C] = \mathsf{V}[T] \tag{11.0.173}$$
$$\mathsf{E}[\mathsf{V}[T|X]] = \mathsf{V}[T]\,2/3. \tag{11.0.174}$$
$$\mathsf{E}[T|X] = 2\,I_{X=A} + (3+\mathsf{E}[T])\,I_{X=B} + (4+\mathsf{E}[T])\,I_{X=C} \tag{11.0.175}$$
$$= 2\,I_{X=A} + 12\,I_{X=B} + 13\,I_{X=C} \tag{11.0.176}$$
$$\mathsf{V}[\mathsf{E}[T|X]] = 4\cdot 2/9 + 144\cdot 2/9 + 169\cdot 2/9 =: \alpha \tag{11.0.177}$$
$$\mathsf{V}[T] = \mathsf{V}[T]\,2/3 + \alpha \quad \text{EVE} \tag{11.0.178}$$
$$\mathsf{V}[T] = 3\alpha. \tag{11.0.179}$$

Here we use that $I_{X=A}$ etc are independent and Bernoulli distributed with success probability p, hence $\mathsf{V}[I_{X=A}] = pq = 1/3\cdot 2/3$.

Grading

- Not using EVE: no points.
- I saw this: $\mathsf{E}\left[T^2\right] = \ldots + (3 + \mathsf{E}[])^2\, 1/3 + \ldots$. This is not correct of course.

Ex 11.100 (1)**.** Why was it actually allowed to assume that $\mathsf{E}[T] < \infty$?

s.11.100. By the strong law of large numbers, any sequence of tunnel selections that excludes tunnel A has probability zero.

Grading:

- Mention the LLN somehow. If not: 0 points.

QUESTION

A server spends a random amount T on a job. While the server works on the job, it sometimes gets interrupted to do other tasks $\{R_i\}$. Such tasks need to be repaired before the machine can continue working again. It is clear that such interruptions can only occur when the server is working. Assume that the interruptions arrive according to a Poisson process with rate λ, i.e., the number of failures $N(t) \sim \text{Pois}(\lambda t)$. The interruptions $\{R_i\}$ are independent of N and of T and form an iid sequence with common mean $\mathsf{E}[R]$ and variance $\mathsf{V}[R]$. The duration S of a job is its own service time T plus all interruptions, hence $S = T + \sum_{i=1}^{N(T)} R_i$.

Ex 11.101 (1.5)**.** In the computation of $\mathsf{V}[S]$ we encounter the following steps.

$$\mathsf{V}\left[\sum_{i=1}^{N(t)} R_i\right] = \mathsf{E}\left[\mathsf{V}\left[\sum_{i=1}^{N(t)} R_i \middle| N(t)\right]\right] + \mathsf{V}\left[\mathsf{E}\left[\sum_{i=1}^{N(t)} R_i \middle| N(t)\right]\right]. \tag{11.0.180}$$

The computation below consists of a number of steps, a, b, ... Explain for each step which property is used to ensure the step is true.

$$\mathsf{V}\left[\mathsf{E}\left[\sum_{i=1}^{N(t)} R_i \middle| N(t)\right]\right] \overset{a}{=} \mathsf{V}[N(t)\,\mathsf{E}[R]] \tag{11.0.181}$$

$$\overset{b}{=} (\mathsf{E}[R])^2\,\mathsf{V}[N(t)] \tag{11.0.182}$$

$$\overset{c}{=} (\mathsf{E}[R])^2 \lambda t. \tag{11.0.183}$$

s.11.101. a. On the set $\{N(t) = n\}$ $N(t) = n$. Hence we can replate $N(t)$ by n. Then use linearity of the expectation.

b. Since $\mathsf{E}[R]$ is a constant, we can take it out from the variance as a square.

c. Variance of Poisson rv is λt
Each property missed, e.g, linearity of expectation, minus 0.5.

Ex 11.102 (2)**.** Suppose $\mathsf{P}\{R = r\} = \mathsf{P}\{T = t\} = 1$, compute $\mathsf{E}[S]$.

s.11.102. Since $T = t$ a.s.,

$$\mathsf{E}[S] = \mathsf{E}[T] + \mathsf{E}[N(T)]\,\mathsf{E}[R] = t + r\,\mathsf{E}[N(t)] = t + r\lambda t.$$

The answer should also be simplied to show that you use all information that is available. Stopping at, e.g., $\mathsf{E}[N(t)\,\mathsf{E}[R]]$ is not completely sufficient. Here are some wrong answers. It's interesting to try to understand why.

$$\mathsf{E}[NR \,|\, N] \neq \mathsf{E}[NR] = \mathsf{E}[N]\,\mathsf{E}[R] \tag{11.0.184}$$

$$\mathsf{E}[S] \neq \mathsf{E}[T] + N(t)\,\mathsf{E}[R]. \tag{11.0.185}$$

Typically, 0.5 point for each correct line, but I subtracted points If you say that $\mathsf{E}[N(T)\,\mathsf{E}[R]] = \mathsf{E}[R]$, or write $n\,\mathsf{E}[R]$ as final answer (apparently you did not get the idea that N is an rv.)

Ex 11.103 (1.5)**.** Explain what this code computes.

Python Code

```python
import numpy as np

labda = 0.5
size = 10
num_runs = 50


def do_run():
    T = np.random.uniform(0, 20)
    N = np.random.poisson(labda * T)
    R = np.random.uniform(1, 5, size=N)
    S = T + R.sum()
    return S


samples = np.zeros(num_runs)
for i in range(num_runs):
    samples[i] = do_run()


print((samples > 8).mean())
```

R Code

```r
labda <- 0.5
size <- 10
num_runs <- 50

do_run <- function() {
  bigT <- runif(n = 1, min = 0, max = 20)
  N <- rpois(n = 1, labda * bigT)
  R <- runif(n = N, min = 1, max = 5)
  S <- bigT + sum(R)
  return(S)
}

samples <- rep(0, num_runs)
for (i in 1:num_runs) {
  samples[i] <- do_run()
}

print(mean(samples > 8))
```

Hint, you should know that in P.21 (R18) the string `samples > 8` collects only the samples with value larger than 8.

s.11.103. Here is an EXPLANATION:

1. T is a simulated service time of a job without interruptions.
2. N is is the simulated number of failures that occur during the service of the job
3. R is then a vector of simulated durations of each of the interruptions
4. Hence, S is the total time the job spends at the server
5. We sample the total service time a number of times, and count how often (mean or variance) this exceeds a threshold.

The points earned depends on how clear you explanation is.

Quite a few of you don't seem to understand what it means to *explain* something. Here are some answers that I saw that certainly don't explain what the code does:

1. 'The code does what's stated in the exercise.'. What's the explanation here? The question is also not: do you understood what the code does?
2. 'T is uniform rv, N is a Poisson rv, R is a also a uniform rv. We repeat this a few times.' Like this you just read the code, but I know you can read, so this type of answer is quite useless.
3. 'We collect a subset of the samples and print that.' This answer is just a repition of the hint.

So, next time, when you are asked to explain something, explain it. In your professional carreer, if you read out loud to your customer what s/he can read for him/herself, you'll be fired pretty soon.

QUESTION

Ex 11.104 (1.5)**.** Let $X \sim \mathcal{N}(\mu, \mu^2)$ and let $Y = e^X$. Showing your work, find the PDF of Y.

s.11.104. We start with finding the CDF of Y.

$$F_Y(y) = \mathsf{P}\left\{Y \le y\right\} = \mathsf{P}\left\{e^X \le y\right\} = \mathsf{P}\left\{X \le \ln y\right\} = F_X(\ln y).$$

Then we differentiate this integral, and we obtain our PDF. Using the FTC, we get

$$\begin{aligned}
F_X(\ln y) &= \int_{-\infty}^{\ln y} f_X(x)\,\mathrm{d}x \implies \\
\frac{\mathrm{d}}{\mathrm{d}y} F_X(\ln y) &= \frac{\mathrm{d}}{\mathrm{d}y}\int_{-\infty}^{\ln y} f_X(x)\,\mathrm{d}x \\
&= f_X(\ln y)\frac{\mathrm{d}\ln y}{\mathrm{d}y} \\
&= f_X(\ln y)\frac{1}{y} \\
&= \frac{1}{\sqrt{2\pi}\mu y}\exp\left(-\frac{1}{2\mu^2}(\ln y - \mu)^2\right)
\end{aligned}$$

for $y > 0$. Here $f_X(x)$ is the PDF of the normal random variable X.

Grading scheme:

- No deduction if second parameter is assumed to be std.dev instead of variance, even though the parametrization should be very clear in this course and other courses.
- Noticing a suitable transformation 0.5pt.
- Correctly applying the transformation theorem 0.5pt.
- Most of: the correct bounds, no mistakes in calculation 0.5pt.

Ex 11.105 (1)**.** Consider now the independent random variables $X_1, X_2 \sim \mathcal{N}(\mu, \sigma^2)$. Let $Y_1 = e^{X_1}$ and $Y_2 = e^{X_2}$. Are $Y_1 Y_2$ and $\frac{Y_1}{Y_2}$ independent? You can use results from the book here.

s.11.105. They are independent. The book proves that $X_1 + X_2$ and $X_1 - X_2$ are independent. Then it must be that $e^{X_1+X_2} = Y_1 Y_2$ and $e^{X_1-X_2} = \frac{Y_1}{Y_2}$ are also independent.

Grading scheme:

- Noticing the independence of the sum and difference of the X_i's 0.5pt.
- Transformations of independent random variables preserve independence 0.5pt. (lenient)

Ex 11.106 (2.5)**.** Find the joint PDF of $U = Y_1 Y_2$ and $V = \frac{Y_1}{Y_2}$.

s.11.106. Since $U = Y_1Y_2$ and $V = \frac{Y_1}{Y_2}$, we can write the inverse functions $Y_1 = \sqrt{UV}$ and $Y_2 = \sqrt{\frac{U}{V}}$. These functions are one-to-one and C^1, so we can write the Jacobian matrix

$$J = \begin{pmatrix} \frac{\sqrt{V}}{2\sqrt{U}} & \frac{\sqrt{U}}{2\sqrt{V}} \\ \frac{1}{2\sqrt{UV}} & -\frac{1}{2V\sqrt{UV}} \end{pmatrix},$$

which has absolute determinant $\frac{1}{2V}$. Since X_1 and X_2 are independent, it must be that Y_1 and Y_2 are independent. Then

$$f_{Y_1,Y_2}(y_1,y_2) = \frac{1}{2\pi\sigma^2 y_1 y_2}\exp\left(-\frac{1}{2\sigma^2}\left((\ln y_1-\mu)^2+(\ln y_2-\mu)^2\right)\right)$$

for $y_1, y_2 \in \mathbf{R}_+$. Then, by the transformation theorem, we have that

$$\begin{aligned} f_{U,V}(u,v) &= f_{Y_1,Y_2}\left(\sqrt{uv}, \sqrt{\frac{u}{v}}\right)\frac{1}{2v} \\ &= \frac{1}{4\pi\sigma^2 uv}\exp\left(-\frac{1}{2\sigma^2}\left((\ln\sqrt{uv}-\mu)^2+(\ln\sqrt{\frac{u}{v}}-\mu)^2\right)\right) \end{aligned}$$

For $u, v \in \mathbf{R}_+$.

Grading scheme:

- Correct inverses 0.5pt.
- Correct absolute determinant of the Jacobian matrix 0.5pt.
- Noticing independence of Y_1, Y_2 0.5pt.
- Correct application of transformation theorem 0.5pt.
- Most of: the correct bounds, no mistakes in calculation, noticing the theorem is applicable, the inverses are 1-1 and C^1 0.5pt.

QUESTION

Ex 11.107 (1)**.** Let $U \sim \text{Unif}\{-n, -n+1, \dots, n-1, n\}$ for some $n \in \mathbb{N}$. Find the PMF of $B = |U|$. What is $\mathsf{E}[B]$?

s.11.107. Since we are dealing with discrete uniform, we see that $\mathsf{P}\{B = k\} = \mathsf{P}\{U = k\} + \mathsf{P}\{U = -k\}$ for $k = 1, 2, \dots, n$, and $\mathsf{P}\{B = 0\} = \mathsf{P}\{U = 0\}$. Hence, we can immediately say that

$$f_B(b) = \begin{cases} \frac{2}{2n+1} & 0 < b \le n \\ \frac{1}{2n+1} & b = 0 \end{cases}.$$

Then,

$$\mathsf{E}[B] = 0\frac{1}{2n+1} + \sum_{i=1}^{n} \frac{2i}{2n+1} = \frac{n(n+1)}{2}\frac{2}{2n+1} = \frac{n^2+n}{2n+1}$$

Grading scheme:

- Correct PMF 0.5pt.
- Correct expectation 0.5pt.

Ex 11.108 (2)**.** Now, consider a random variable X distributed according to a Beta(p, q) distribution. Since this distribution is only defined on $(0, 1)$, we will transform it to be more general. Consider the random variable $Z = bX + a(1 - X)$, for some $a, b \in \mathbb{R}$ such that $a < b$, and find its PDF.

s.11.108. First, note that $Z = a + (b - a)X$. Then, we know that

$$F_Z(y) = \mathsf{P}\left\{X(b-a) + a \le y\right\} = \mathsf{P}\left\{X \le \frac{y-a}{b-a}\right\} = \int_0^{\frac{y-a}{b-a}} f_X(s)\,\mathrm{d}s.$$

Then, by the FTC, we get

$$f_Z(y) = f_X(\frac{y-a}{b-a})\frac{1}{b-a} = \frac{(y-a)^{p-1}(b-y)^{q-1}}{\beta(p,q)(b-a)^{p+q-1}}$$

for $a < y < b$.

Grading scheme:

- Correct rewriting of Z 0.5pt.
- Correct CDF 0.5pt.
- Correct PDF 0.5pt.

- No mistakes, correct bounds etc. 0.5pt.

Ex 11.109 (2)**.** Consider again Z as in the previous exercise. Assume that $a = -b$, that $b > 0$, and that $p = q = 2$. What is the PDF of $|Z|$?

s.11.109. We start as usual by considering the CDF of $Q = |Z|$. This shows us that

$$F_Q(y) = \mathsf{P}\left\{Q \leq y\right\} = \mathsf{P}\left\{|Z| \leq y\right\} = \mathsf{P}\left\{-y \leq Z \leq y\right\} = F_Z(y) - F_Z(-y).$$

We see that

$$F_Z(y) = \int_{-b}^{y} \frac{3}{4b^3}(b^2 - s^2)\,\mathrm{d}s = \frac{3}{4b}y - \frac{1}{4b^3}y^3 + \frac{1}{2}$$

after filling all values given and integrating. This holds for $0 < y < b$, the CDF is 1 for $y \geq b$, and 0 for $y \leq 0$. Then it must be that

$$f_Q(y) = \frac{\mathrm{d}}{\mathrm{d}y}(F_Z(y) - F_Z(-y)) = \frac{3}{2b} - \frac{3}{2b^3}y^2$$

for $0 < y < b$.

Grading scheme:

- Difference of CDF 0.5pt.
- Difference of CDF of Z correct, and derivative 1pt.
- Most of: no mistakes, correct bounds etc. 0.5pt.

QUESTION

Consider the following code:

Python Code

```python
import numpy as np
from scipy.stats import expon
np.random.seed(42)

n = 750
N = 300

X = expon(scale = 3).rvs([n, N])
Y = X.mean(axis = 1)

mu = 3
sigma = 3
Z = np.sqrt(N) * (Y - mu)/sigma

print((Z ** 71).mean())
```

R Code

```r
set.seed(42)

n <- 750
N <- 300

X <- matrix(rexp(n * N, rate = 1/3), nrow = n, ncol = N)
Y <- rowMeans(X)

mu <- 3
sigma <- 3
Z <- sqrt(N) * (Y - mu)/sigma

print(mean(Z^71))
```

Ex 11.110 (1.5)**.** Y is a vector of i.i.d. random variables.

(i) What is the length ℓ of Y?

(ii) Each element of Y is a mean of k i.i.d. Exp(λ) r.v.s. What are k and λ?

(iii) What are the (theoretical) expectation and variance of an element of Y?

s.11.110. The length of Y is $n = 750$; Each element of Y is a mean of $k = N = 300$ i.i.d. Exp(1/3) r.v.s. The expectation is $\frac{1}{\lambda} = 3$ and the variance is $\frac{1}{k\lambda^2} = \frac{1}{300 \cdot 1/9} = 0.03$.

Grading scheme:

- 0.5 for getting both the length $n = 750$ and $k = 300$ correct (no partial credit);
- 0.5 for $\lambda = 1/3$, expectation 3 and the factor $\frac{1}{\lambda^2} = \frac{1}{1/9}$ in the variance (no partial credit);
- 0.5 for the factor $\frac{1}{k}$ in the variance.

Recall that each element of Y is the mean of k i.i.d. Exp(λ) r.v.s.

Ex 11.111 (1)**.** What is the exact distribution of an element of Y?
Give its name and its parameters, and explain the answer.

s.11.111. The sum of k i.i.d. Exp(λ) r.v.s. has the Gamma(k, λ) distribution. Hence, the mean of k i.i.d. Exp(λ) r.v.s. has the Gamma($k, k\lambda$) distribution. In this exercise, $k = 300$ and $k\lambda = 100$.
Grading scheme:

- 0.5 for Gamma with first parameter k
- 0.5 for the second parameter

Let $(Y_1, \dots, Y_\ell)$ be the elements of Y and let $(Z_1, \dots, Z_\ell)$ be the elements of Z. Recall that each Z_i depends on k because Y_i is the mean of k i.i.d. Exp(λ) r.v.s. Let T be the random variable to which Z_1 converges in the limit $k \to \infty$.

Ex 11.112 (1)**.** What is the distribution of T, and why?
Hence, what is an approximate distribution of an element of Y (e.g. Y_1) and why?

s.11.112. By the CLT, $Z_1 \dot{\sim} \text{Norm}(0, 1)$. Hence, $Y_1 = \mu + \sigma/\sqrt{n} Z_1 \dot{\sim} \text{Norm}(3, 9/n)$.
Grading scheme:

- 0.5 for mentioning CLT and the distribution of Z_1;
- 0.5 for the approximate distribution of Y_1.

Ex 11.113 (0.5)**.** In the last line of the code, we compute $S = \frac{1}{\ell} \sum_{i=1}^{\ell} Z_i^{71}$.
If $k \to \infty$ (for fixed ℓ, e.g. $\ell = 4$), does S converge to a constant?
You do not have to give formal proofs in this subquestion, but you should give clear explanations using the appropriate theorem(s).

s.11.113. In the limit $k \to \infty$, each Z_i has the standard normal distribution by CLT, but the sum of ℓ powers of normal distributions has a non-trivial CDF.

- 0.5 for explaining that S does not converge to a constant; no points for just CLT (unless previous question was not answered).

Ex 11.114 (1)**.** In the last line of the code, we compute $S = \frac{1}{\ell}\sum_{i=1}^{\ell} Z_i^{71}$.

If $\ell \to \infty$ (for fixed k), does S converge to a constant? If so, does it converge to $\mathsf{E}\left[T^{71}\right]$?

You do not have to give formal proofs in this subquestion, but you should give clear explanations using the appropriate theorem(s).

s.11.114. By LLN, S does converge to a constant as $\ell \to \infty$, however, it converges to $\mathsf{E}\left[Z_1^{71}\right]$ for that fixed value of k. By symmetry, we have $\mathsf{E}\left[Z_1^{71}\right] = 0$. However, the gamma distribution is right-skewed, which implies $\mathsf{E}\left[T^{71}\right] > 0$. Hence, it does not converge to $\mathsf{E}\left[T^{71}\right]$.
Grading scheme:

- 0.5 for concluding that S converges to a constant using LLN.
- 0.5 for explaining why the constant is not equal to $\mathsf{E}\left[T^{71}\right]$.

QUESTION

John is an archer who likes to shoot at small targets. To find out his skill level, John plays the following game. He shoots arrows at a target and counts the number of times he successfully hits the target. He keeps counting until he has missed r times, at which moment the current round of the game stops. His score for the round is the total number of successful shots in the round. John plays n rounds in total and we assume that all shots are independent and have the same (unknown) success probability p. John is interested in finding out his skill level. That is, he is interested in learning the value of p.

Given the value of p, John's score Y_i for the ith round of the game follows a negative binomial distribution with parameters r and p. That is, for every $i = 1, \dots, n$, we have that $Y_i | p \sim \text{NB}(r, p)$, with a corresponding pmf defined by

$$\mathsf{P}\{Y_i = y_i | p\} = \binom{y_i + r - 1}{y_i} (1-p)^r p^{y_i}, \tag{11.0.186}$$

for $y_i = 0, 1, 2, \dots$. John's prior belief about the distribution of p is that it follows a Beta(a, b) distribution with given values for a and b (the exact values of a and b are not relevant for this question).

Ex 11.115 (2.5)**.** In the first round, John gets a score of $Y_1 = y_1$. What is John's *posterior* distribution of p, given this first observation?

s.11.115. Using Bayes' rule we have

$$\begin{aligned}
f_1(p|Y_1 = y_1) &= \frac{\mathsf{P}\{Y_1 = y_1|p\} f_0(p)}{\mathsf{P}\{Y_1 = y_1\}} &(11.0.187)\\
&\propto \mathsf{P}\{Y_1 = y_1|p\} f_0(p) &(11.0.188)\\
&= \binom{y_1 + r - 1}{y_1} (1-p)^r p^{y_1} \frac{1}{B(a,b)} p^{a-1} (1-p)^{b-1} &(11.0.189)\\
&\propto (1-p)^{r+b-1} p^{a+y_1-1}, &(11.0.190)
\end{aligned}$$

in which we recognize the pdf of a Beta$(a + y_1, b + r)$ distribution (up to a constant factor). Hence, the posterior distribution of p given $Y_1 = y_1$ is a Beta$(a + y_1, b + r)$ distribution.

Ex 11.116 (1)**.** Is John's prior distribution a *conjugate* prior?

s.11.116. Yes. The posterior distribution is in the same family of distributions (Beta) as the prior. Hence, Amy has a conjugate prior.

Ex 11.117 (1.5)**.** Suppose John plays n rounds and observes the scores $Y_1 = y_1, \dots, Y_n = y_n$. What is his posterior distribution after these observations?
Hint: you don't need to do a lot of math here! Instead, use the result of the first question above and a logical argument.

s.11.117. The posterior after observing $Y_1 = y_1$ becomes our new prior. Hence, our new prior is a Beta$(a + y_1, b + r)$ distribution. From question 1 it follows that the prior after observing $Y_2 = y_2$ then is a Beta$(a + y_1 + y_2, b + 2r)$ distribution. Hence, iterating this process, we find that the posterior distribution of p after observing $Y_1 = y_1, \ldots, Y_n = y_n$ is a Beta$(a + \sum_{i=1}^n y_i, b + rn)$ distribution.

QUESTION

Let $Y \sim \text{Norm}(0,1)$. In this exercise, we find an upper bound for $\mathsf{P}\{|Y| > 4\}$.

Ex 11.118 (1.5)**.** If $X \sim \text{Gamma}(a, \lambda)$ then the rth moment of X is given by $\frac{\Gamma(a+r)}{\lambda^r \Gamma(a)}$. Use this to prove that $\mathsf{E}\left[Y^{2n+2}\right] = (2n+1)\,\mathsf{E}\left[Y^{2n}\right]$ for all positive integers n.

Hint. Use the chi-square distribution.

s.11.118. Let $V = Y^2$, then $V \sim \chi^2_1$, so $V \sim \text{Gamma}(\frac{1}{2}, \frac{1}{2})$. We have

$$\mathsf{E}\left[Y^{2n+2}\right] = \mathsf{E}\left[V^{n+1}\right] = \frac{\Gamma(n+3/2)}{(1/2)^{n+1}\Gamma(1/2)} = \frac{2^{n+1}\Gamma(n+3/2)}{\Gamma(1/2)}$$
$$\mathsf{E}\left[Y^{2n}\right] = \mathsf{E}\left[V^{n}\right] = \frac{\Gamma(n+1/2)}{(1/2)^{n}\Gamma(1/2)} = \frac{2^{n}\Gamma(n+1/2)}{\Gamma(1/2)}.$$

Since $\Gamma(n+3/2) = (n+1/2)\Gamma(n+1/2)$ we conclude that

$$\mathsf{E}\left[Y^{2n+2}\right] = \frac{2^{n+1}\Gamma(n+3/2)}{\Gamma(1/2)} = 2(n+1/2)\frac{2^{n}\Gamma(n+1/2)}{\Gamma(1/2)} = (2n+1)\,\mathsf{E}\left[Y^{2n}\right].$$

Remarks and grading scheme:

- It is NOT true that $\mathsf{E}\left[Y^{2n+2}\right] = \mathsf{E}\left[Y^2\right]\mathsf{E}\left[Y^{2n}\right]$. This would only be true if Y^2 and Y^{2n} would be uncorrelated. But clearly, they are positively correlated: if Y^2 is large, then so is Y^{2n}.
- While induction is a good strategy to try when you have to prove a statement for all $n \in \mathbb{N}$, it is certainly not the only option.
- Grading: 0.5 for introducing $V = Y^2$ and arguing $V \sim \text{Gamma}(\frac{1}{2}, \frac{1}{2})$ with some explanation,
  0.5 for using the given expression to calculate $\mathsf{E}\left[Y^{2n+2}\right]$ and $\mathsf{E}\left[Y^{2n}\right]$,
  0.5 for using a property of the Gamma function to finish the answer.

Ex 11.119 (1)**.** Use the previous exercise to calculate $\mathsf{E}\left[Y^4\right]$ and $\mathsf{E}\left[Y^8\right]$.

s.11.119. By applying the previous exercise with $n = 1$, we obtain that $\mathsf{E}\left[Y^4\right] = 3\,\mathsf{E}\left[Y^2\right] = 3$. By applying the previous exercise with $n = 3$ and $n = 2$, we obtain that $\mathsf{E}\left[Y^8\right] = 7\,\mathsf{E}\left[Y^6\right] = 7 \cdot 5\,\mathsf{E}\left[Y^4\right] = 7 \cdot 5 \cdot 3 = 105$.

Remarks and grading scheme:

- If the exercise explicitly asks to use the previous exercise, don't do it in a different way.

- You should really know that $\mathsf{E}\left[Y^2\right] = 1$ for $Y \sim \text{Norm}(0,1)$.
- Grading: 0.5 for a correct solution for $\mathsf{E}\left[Y^4\right]$ and 0.5 for a correct solution for $\mathsf{E}\left[Y^8\right]$.

Ex 11.120 (1)**.** We now provide a bound for $\mathsf{P}\{|Y| > 4\}$.

(i) Prove that $\mathsf{P}\{|Y| > 4\} = \mathsf{P}\left\{Y^4 > 256\right\}$.

(ii) Use this to prove that $\mathsf{P}\{|Y| > 4\} \leq \frac{3}{256}$.

s.11.120. We have $\mathsf{P}\{|Y| > 4\} = \mathsf{P}\left\{Y^4 > 256\right\}$ since $|Y| > 4$ if and only if $Y^4 > 256$. The inequality now directly follows from Markov's inequality.

Remarks and grading scheme:

- This exercise was made quite well.
- Grading scheme: 0.5 for (i) and 0.5 for (ii).

Ex 11.121 (1.5)**.** Prove that $\mathsf{P}\{|Y| > 4\} \leq \frac{\mathsf{E}[Y^{2n}]}{16^n}$ for all $n \in \mathbb{N}$.
For what value(s) of n do we obtain the strongest bound for $\mathsf{P}\{|Y| > 4\}$?

s.11.121. By Markov's inequality,

$$\mathsf{P}\{|Y| > 4\} = \mathsf{P}\left\{Y^{2n} > 16^n\right\} \leq \frac{\mathsf{E}\left[Y^{2n}\right]}{16^n}.$$

From the formula for $\mathsf{E}\left[Y^{2n}\right]$ we see that $\mathsf{E}\left[Y^{2(n+1)}\right] = (2n+1)\,\mathsf{E}\left[Y^{2n}\right]$. We now consider what happens when incrementing n. If $n < 8$ then $2n+1 < 16$, so then incrementing n improves the bound, but for $n \geq 8$ the bound becomes weaker. So we get the best possible bound for $n = 8$.
Remarks and grading scheme:

- Many people tried to use induction in this exercise, but that doesn't work. While it is good to think of induction when you have to prove a statement for all $n \in \mathbb{N}$, it is certainly not the only option.
- In fact, following a similar strategy as in the previous exercise would have helped here.
- Grading scheme: 0.5 for $\mathsf{P}\{|Y| > 4\} \leq \frac{\mathsf{E}[Y^{2n}]}{16^n}$, 1 for showing that the best bound is obtained for $n = 8$.

QUESTION

We have a queue of people served by a potentially infinite number of servers. Let $L(t)$ be the number of people present in the system at time t. For any time $t \geq 0$ the time to the next arriving person is $X \sim \text{Exp}(\lambda)$, and given $L(t) = n$ customers in the system at time t, the time to the next departing customer is $S \sim \text{Exp}(\mu n)$. The rvs S and X are independent, and $\lambda, \mu > 0$. Write $B(h)$ for the number of arrivals during an interval of length h, and $D(h)$ for the number of departures. (Hint: recall the relation between the Poisson distribution and the exponential distribution.)

In the sequel, take h positive, but very, very small, i.e, $h \ll 1$. With this, we use the shorthand $o(h)$ to capture all terms of a polynomial in h with a power higher than 1, for instance,

$$2h + 3h^2 + 44h^{21} = 2h + o(h). \tag{11.0.191}$$

Like this we can hide all nonlinear terms of a polynomial in the $o(h)$ function. This is easy when we want to take limits, for example,

$$\lim_{h\to 0} \frac{2h + 3h^2 + 44h^{21}}{h} = 2 + \lim_{h\to 0} \frac{o(h)}{h} = 2 + 0. \tag{11.0.192}$$

In other words, when computing this limit for $h \to 0$, we don't care about the details in $o(h)$ because $o(h)/h \to 0$ anyway.

Ex 11.122 (1)**.** Explain that

$$\mathsf{P}\{B(h) = 1, D(h) = 0 | L(0) = n\} = \lambda h e^{-\lambda h} e^{-\mu n h}. \tag{11.0.193}$$

s.11.122. Since job interarrival and departure times are exponentially distributed, we can use that $B(h) \sim \text{Pois}(\lambda h)$ and $D(h) = 0 \implies S > h$, hence $\mathsf{P}\{S > h | L(0) = n\} = e^{-\mu n h}$.

Mentioning that both are Poisson is also fine, but see the next question.

Ex 11.123 (1)**.** Use the first degree Taylor's expansion, $f(h) \approx f(0) + hf'(0) + o(h)$, to motivate that

$$\mathsf{P}\{B(h) = 0, D(h) = 1 | L(0) = n\} = n\mu h + o(h). \tag{11.0.194}$$

s.11.123.

$$\begin{aligned} \mathsf{P}\{B(h) = 0, D(h) = 1 | L(0) = n\} &= e^{-\lambda h} \mu n h e^{-\mu n h} + o(h() && (11.0.195) \\ &= (1 - \lambda h)\mu n h (1 - \mu n h) + o(h) = \mu n h + o(h). && (11.0.196) \end{aligned}$$

Note that $X > h \implies B(h) = 0$. We also know that for $h << 1$, the rv $D(h)$ is nearly Poisson distributed with mean $\mu n h$. The first $o(h)$ is necessary because during the time h also two departures can occur and then the departure rates are not the same. Before the departure, people leave at rate μn, but after the first departure they leave at rate $\mu(n-1)$. However, since two or more departures have very small, in fact have $o(h)$ probability, we can capture all such details in the $o(h)$ terms.

I don't require the explanation about this subtle point.

Ex 11.124 (2)**.** Explain that

$$\mathsf{E}[L(t+h)|L(t)=n] = n + (\lambda - \mu n)h + o(h). \tag{11.0.197}$$

s.11.124. Use conditional expectation and the above results to see that

$$\begin{aligned}
\mathsf{E}[L(t+h)|L(t)=n] &= n\,\mathsf{P}\{B(h)=0, D(h)=0\} + (n+1)\,\mathsf{P}\{B(h)=1, D(h)=0\} && (11.0.198)\\
&\quad + (n-1)\,\mathsf{P}\{B(h)=0, D(h)=1\} + o(h) && (11.0.199)\\
&= ne^{-\lambda h}e^{-\mu n h} + (n+1)\lambda h + (n-1)\mu n h + o(h) && (11.0.200)\\
&= n(1-\lambda h)(1-\mu n h) + (n+1)\lambda h + (n-1)\mu n h + o(h) && (11.0.201)\\
&= n - n(\lambda + \mu n)h + (n+1)\lambda h + (n-1)\mu n h + o(h) && (11.0.202)\\
&= n + (\lambda - \mu n)h + o(h). && (11.0.203)
\end{aligned}$$

Write $M(t) = \mathsf{E}[L(t)]$.

Ex 11.125 (1)**.** Derive that

$$M(t+h) = M(t) + (\lambda - \mu M(t))h + o(h). \tag{11.0.204}$$

s.11.125. Replace n by $L(t)$ in $\mathsf{E}[L(t+h)|L(t)]$ to see that

$$\mathsf{E}[L(t+h)|L(t)] = L(t) + (\lambda - \mu L(t))h + o(h). \tag{11.0.205}$$

Take expectations left and right and use Adam's law.

QUESTION

Ex 11.126 (0.5)**.** Consider a random variable $X \sim \mathcal{N}(\mu, \sigma^2)$. Write down the PDF of the random variable $Y = e^X$. You do not have to elaborate on your answer, but make sure to get everything correct.

s.11.126. The PDF of Y is given by

$$= \frac{1}{\sqrt{2\pi}\sigma y} \exp\left(-\frac{1}{2\sigma^2}(\ln y - \mu)^2\right)$$

for $y > 0$.

Grading scheme:

- Correct 0.5pt.

Ex 11.127 (1.5)**.** Consider now the random variable $W_k = \frac{k}{5Y^2}$. What is the distribution of W_k? You can use results from the book here.

s.11.127. Notice that, since Y is a normal rv, the log of Y is log-normal. Then, taking the ln on both sides, we get that

$$\ln W_k = \ln k - \ln 5 - 2\ln Y_k.$$

From the book, we know that if $\ln Y \sim \mathcal{N}(\mu, \sigma^2)$, then it must be that

$$-2\ln Y \sim \mathcal{N}(-2\mu, 4\sigma^2),$$

and that

$$\ln W_k = \ln k - \ln 5 - 2\ln Y \sim \mathcal{N}(\ln k - \ln 5 - 2\mu, 4\sigma^2).$$

Thus, $W_k \sim \mathcal{LN}(\ln k - \ln 5 - 2\mu, 4\sigma^2)$

Grading scheme:

- The idea to take logs 0.5pt.
- The rest correct 1pt.

Ex 11.128 (1)**.** Calculate $\mathsf{P}\left\{\frac{W_k}{W_l} = \frac{k}{l}\right\}$ for some $l > k > 0$. Are $W_k W_l$ and $\frac{W_k}{W_l}$ independent?

s.11.128. Clearly, $W_k = \frac{k}{l} W_l$, and thus we can see that

$$W_k W_l = \frac{k}{l} W_k^2,$$
$$\frac{W_k}{W_l} = \frac{k}{l}.$$

These are independent, since one is a constant.

Grading scheme:

- Correct probability 0.5pt.
- Correct conclusion 0.5pt.

Ex 11.129 (2)**.** Let $X_1, X_2 \sim X$ be IID random variables, where X has PDF

$$f_X(x) = \frac{1}{\sqrt{2\pi x}} \exp\left(-\frac{x}{2}\right)$$

for $x > 0$. Find the joint PDF of the random variables $U = X_1 + X_2$ and $V = X_1 - X_2$.

s.11.129. Since $U = X_1 + X_2$ and $V = X_1 - X_2$, we can write the inverse functions $X_1 = \frac{1}{2}(U+V)$ and $X_2 = \frac{1}{2}(U-V)$. These functions are one-to-one and C^1, so we can write the Jacobian matrix

$$J = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix},$$

which has absolute determinant $\frac{1}{2}$. Since X_1 and X_2 are independent,

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi\sqrt{x_1 x_2}} \exp\left(-\frac{x_1+x_2}{2}\right)$$

for $x_1, x_2 \in \mathbf{R}_+$. Then, by the transformation theorem, we have that

$$\begin{aligned} f_{U,V}(u,v) &= f_{X_1,X_2}\left(\frac{1}{2}(u+v), \frac{1}{2}(u-v)\right)\frac{1}{2} \\ &= \frac{1}{2\pi\sqrt{u^2-v^2}} \exp\left(-\frac{u}{2}\right) \end{aligned}$$

For $-\infty < v < u < \infty$.

Grading scheme:

- Correct inverses 0.5pt.
- Correct absolute determinant of the Jacobian matrix 0.5pt.
- Correct application of transformation theorem 0.5pt.
- Most of: the correct bounds, no mistakes in calculation, noticing the theorem is applicable, the inverses are 1-1 and C^1, independence of X_1 and X_2 etc. 0.5pt.

QUESTION

A random point (X, Y) is chosen in the following square:

$$\{(x, y) : -\sqrt{\pi} < x < \sqrt{\pi}, -\sqrt{\pi} < y < \sqrt{\pi}\}.$$

All points are equally likely to be chosen. Let R be its distance from the origin.

Ex 11.130 (0.5)**.** Find the joint PDF $f(x, y)$ of X and Y.

s.11.130. So we can see that both X and Y are uniformly distributed on $(-\sqrt{\pi}, \sqrt{\pi})$. Then their joint PDF is simply:

$$\begin{aligned} f(x, y) &= \left(\frac{1}{\sqrt{\pi} - (-\sqrt{\pi}))}\right)\left(\frac{1}{(\sqrt{\pi} - (-\sqrt{\pi}))}\right) \\ &= \frac{1}{4\pi} \end{aligned}$$

One mistake, zero points.

Ex 11.131 (0.5)**.** Show that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$

s.11.131.

$$\begin{aligned} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dx\, dy &= \int_{-\sqrt{\pi}}^{\sqrt{\pi}}\int_{-\sqrt{\pi}}^{\sqrt{\pi}} \frac{1}{4\pi}\, dx\, dy \\ &= 1 \end{aligned}$$

One mistake, zero points

Ex 11.132 (3)**.** Find the expectation of R^2, i.e., the expected squared difference from the origin.

s.11.132. Note that $R = \sqrt{X^2 + X^2}$, so then $R^2 = X^2 + Y^2$. Using LOTUS:

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(x^2+y^2\right) f(x,y)\,dx\,dy &= \int_{-\sqrt{\pi}}^{\sqrt{\pi}}\int_{-\sqrt{\pi}}^{\sqrt{\pi}}\left(x^2+y^2\right)\left(\frac{1}{4\pi}\right)dx\,dy \\
&= \int_{-\sqrt{\pi}}^{\sqrt{\pi}}\int_{-\sqrt{\pi}}^{\sqrt{\pi}}\frac{x^2+y^2}{4\pi}\,dx\,dy \\
&= \frac{1}{4\pi}\int_{-\sqrt{\pi}}^{\sqrt{\pi}}\int_{-\sqrt{\pi}}^{\sqrt{\pi}} x^2+y^2\,dx\,dy \\
&= \frac{1}{4\pi}\int_{-\sqrt{\pi}}^{\sqrt{\pi}}\left[\frac{x^3}{3}+y^2x\right]_{-\sqrt{\pi}}^{\sqrt{\pi}}dy \\
&= \frac{1}{4\pi}\int_{-\sqrt{\pi}}^{\sqrt{\pi}}\left(\frac{\pi^{\frac{3}{2}}}{3}+\sqrt{\pi}y^2-\frac{(-\pi)^{\frac{3}{2}}}{3}+\sqrt{\pi}y^2\right)dy \\
&= \frac{1}{4\pi}\int_{-\sqrt{\pi}}^{\sqrt{\pi}}\left(\frac{\pi^{\frac{3}{2}}}{3}+\sqrt{\pi}y^2+\frac{\pi^{\frac{3}{2}}}{3}+\sqrt{\pi}y^2\right)dy \\
&= \frac{1}{4\pi}\int_{-\sqrt{\pi}}^{\sqrt{\pi}}\frac{2\pi^{\frac{3}{2}}}{3}+2\sqrt{\pi}y^2\,dy \\
&= \frac{1}{4\pi}\left[\frac{2\pi^{\frac{3}{2}}}{3}y+\frac{2\sqrt{\pi}y^3}{3}\right]_{-\sqrt{\pi}}^{\sqrt{\pi}} \\
&= \frac{1}{4\pi}\left(\frac{2\pi^2}{3}+\frac{2\pi^2}{3}+\frac{2(-\pi)^2}{3}+\frac{2(-\pi)^2}{3}\right) \\
&= \frac{1}{4\pi}\frac{8\pi^2}{3}=\frac{2\pi}{3}
\end{aligned}$$

One point for finding that $R^2 = X^2 + Y^2$ and writing down the integral correctly using LOTUS. 2 points for the remaining calculations.

Consider the following code:

Python Code

```python
import numpy as np
np.random.seed(3)

num = 100000

x = np.random.normal(loc = 50, scale = 200, size = num)
y = np.random.normal(loc = 20, scale = 100, size = num)

result = np.zeros(num)
for i in range(0,num):
    result[i] = x[i]*y[i]

print(np.mean(result))
```

Ex 11.133 (0.5)**.** What does the code above do?

s.11.133. It loads the required packages and creates two samples with 100000 observations from respectively a $\mathcal{N}(50,200)$- and $\mathcal{N}(20,100)$-distribution. Then for all paired observations it computes the product and takes the mean to estimate $E(XY)$.
0.5 points if it is mentioned a product is taken and an average is computed/estimated.

Ex 11.134 (0.5)**.** The code gives as output 1008.99966.
Explain why you would expect to get this output from the code.

s.11.134. As the samples are generated independently we would expect $E(XY) = E(x)E(Y) = 50 * 20 \approx 1000$. This is indeed shown by the code.
0.5 points if independence is mentioned. Which would then result in $E(XY) = E(X)E(Y)$

QUESTION

Catherine and Denny are playing a game. Each player throws a single (fair) die. The die has three pairs of identical sides: a pair of ones, a pair of twos and a pair of threes. Let X and Y denote the outcome of Catherine and Denny's throw, respectively. The person that throws the highest number wins. If both throw the same number, they have a draw. The final score C of Catherine is determined as follows:

- If Catherine loses, then she gets zero points.
- If Catherine and Denny draw, then she gets 0.5 point.
- If Catherine wins, then her score is the difference $X-Y$ in the numbers they threw.

The final score D for Denny is determined analogously. Assume that the dice are fair and that all throws are independent.

Ex 11.135 (1)**.** Find the joint distribution of X and Y conditional on Catherine winning.

s.11.135. The initial distribution is $\mathsf{P}\{X=x, Y=y\} = 1/9$, $x, y = 1,2,3$. Catherine wins iff $X > Y$, i.e., iff $(X, Y) \in \{(2,1),(3,1),(3,2)\}$. So we have

$$\begin{aligned}
\mathsf{P}\{X=x, Y=y|X>Y\} &= \frac{\mathsf{P}\{X=x, Y=y, X>Y\}}{\mathsf{P}\{X>Y\}} && (11.0.206)\\
&= \frac{1/9}{1/3} = 1/3, && (11.0.207)
\end{aligned}$$

for $(x,y) \in \{(2,1),(3,1),(3,2)\}$.

Ex 11.136 (1.5)**.** What is Catherine's expected score conditional on Catherine winning?

s.11.136. We have

$$C = \begin{cases} 0 & X<Y \\ 1/2 & X=Y \\ X-Y & X>Y. \end{cases} \qquad (11.0.208)$$

Hence,

$$\begin{aligned}
\mathsf{E}[C|X>Y] &= \mathsf{E}[X-Y|X>Y] && (11.0.209)\\
&= \frac{1}{3}(2-1) + \frac{1}{3}(3-1) + \frac{1}{3}(3-2) && (11.0.210)\\
&= 4/3. && (11.0.211)
\end{aligned}$$

Ex 11.137 (1.5)**.** Determine Catherine's (unconditional) expected score.

s.11.137. We have, by the law of total expectation,

$$\begin{aligned}
\mathsf{E}[C] &= \mathsf{P}\{X < Y\}\,\mathsf{E}[C|X < Y] + \mathsf{P}\{X = Y\}\,\mathsf{E}[C|X = Y] + \mathsf{P}\{X > Y\}\,\mathsf{E}[C|X > Y] && (11.0.212)\\
&= 0 + \frac{1}{3}\cdot\frac{1}{2} + \frac{1}{3}\frac{4}{3} && (11.0.213)\\
&= 11/18. && (11.0.214)
\end{aligned}$$

To spice things up, Catherine and Denny decide to play for money. After playing the dice game and scoring C points, Catherine receives an amount of S euros, where S is determined randomly. Here, conditional on the outcome of C, S follows a uniform distribution on $[C, 2C]$.

Ex 11.138 (1)**.** What is the (unconditional) expected reward for Catherine? That is, compute $\mathsf{E}[S]$.

s.11.138. By Adam's law,

$$\begin{aligned}
\mathsf{E}[S] &= \mathsf{E}[\mathsf{E}[S|C]] && (11.0.215)\\
&= \mathsf{E}\left[\frac{3}{2}C\right] && (11.0.216)\\
&= \frac{3}{2}\mathsf{E}[C] && (11.0.217)\\
&= \frac{3}{2}\frac{11}{18} && (11.0.218)\\
&= \frac{33}{36}. && (11.0.219)
\end{aligned}$$

QUESTION

A random point (X, Y) is chosen in the following square:

$$\{(x, y) : x^2 < e, y^2 < e\}$$

All points are equally likely to be chosen. Let S be the Euclidean norm of (X, Y).

Ex 11.139 (1)**.** Find the joint PDF $f(x, y)$ of X and Y.

s.11.139. Since (X, Y) is uniformly distributed on $(-\sqrt{e}, \sqrt{e})^2$, their joint PDF is simply:

$$\begin{aligned} f(x, y) =& \Big(\frac{1}{\sqrt{e}-(-\sqrt{e}))}\Big)\Big(\frac{1}{(\sqrt{e}-(-\sqrt{e}))}\Big) \\ =& \frac{1}{4e}, \end{aligned}$$

for $x \in (-\sqrt{e}, \sqrt{e})$ and $y \in (-\sqrt{e}, \sqrt{e})$. *0.5 points for the solution, 0.5 points for the correct bounds.*

Ex 11.140 (3)**.** Find the expectation of S^2, i.e., the squared norm of (X, Y).

s.11.140. Note that $S = X^2 + Y^2$. Using LOTUS and symmetry in x and y:

$$\begin{aligned} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Big(x^2+y^2\Big)f(x,y)\,dx\,dy =& \int_{-\sqrt{e}}^{\sqrt{e}}\int_{-\sqrt{e}}^{\sqrt{e}}\Big(x^2+y^2\Big)\Big(\frac{1}{4e}\Big)\,dx\,dy \\ =& \frac{1}{4e}\int_{-\sqrt{e}}^{\sqrt{e}}\int_{-\sqrt{e}}^{\sqrt{e}}(x^2+y^2)\,dx\,dy \\ =& \frac{1}{4e}\int_{-\sqrt{e}}^{\sqrt{e}}\int_{-\sqrt{e}}^{\sqrt{e}}x^2\,dx\,dy + \frac{1}{4e}\int_{-\sqrt{e}}^{\sqrt{e}}\int_{-\sqrt{e}}^{\sqrt{e}}y^2\,dx\,dy+ \\ =& \frac{2}{4e}\int_{-\sqrt{e}}^{\sqrt{e}}\int_{-\sqrt{e}}^{\sqrt{e}}x^2\,dx\,dy \\ =& \frac{1}{2e}\int_{-\sqrt{e}}^{\sqrt{e}}[\frac{1}{3}x^3]_{\sqrt{e}}^{\sqrt{e}}\,dy \\ =& \frac{1}{6e}\int_{-\sqrt{e}}^{\sqrt{e}}(e^{3/2}+e^{3/2})\,dy \\ =& \frac{2}{6e}2\sqrt{e}e^{3/2} \\ =& \frac{4e^2}{6e} = \frac{2}{3}e. \end{aligned}$$

One point for finding $S^2 = X^2 + Y^2$ and writing down the integral correctly using LOTUS. Two points for the remaining calculations.

Consider the following code:

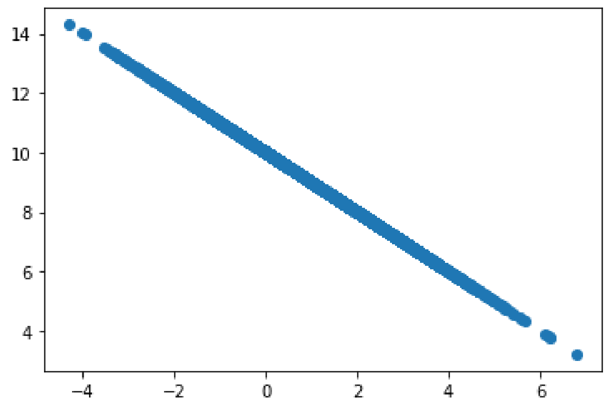
Python Code

```python
import numpy as np
import matplotlib.pyplot as plt
np.random.seed(3)

num = 10000

x = np.random.normal(loc = 1, scale = np.sqrt(2), size = num)
y = 10 - x

print(np.corrcoef(x,y))

plt.scatter(x,y)
```

R Code

```r
set.seed(3)

num = 10000

x = rnorm(num, mean = 1, sd = sqrt(2))
y = 10 - x

print(cor(x,y))

plot(x,y)
```

Ex 11.141 (1)**.** This code gives the value -1 and the following graph.



Explain, what the relationship is between the numerical and graphical output and why the output is -1.

s.11.141. The output of -1 means that the correlation between X and Y, is -1. This makes sense since $Y = 10 - X$, so obviously the correlation is -1 as Y completely determines the value of X.

If Y goes up by 1, X always goes down by exactly 1. This can also be seen in the graph where X and Y always sum op to 10 and there is are linear negative relationship between them. No points are deviated from the line, X and Y will always move together.

QUESTION

N number of employees will participate in the pension system. There are currently two types of pension schemes, Plan A and Plan B. Each employee independently chooses Plan A with probability p, and Plan B with probability $1-p$.

Ex 11.142 (2)**.** Suppose $N \sim \text{Pois}(\lambda)$. Let X_A be the number of people that choose Plan A and $X_B = N - X_A$ be the number of people that choose Plan B. Find $Var(X_A - X_B)$ and $\rho_{X_B,N}$.

s.11.142. First notice that $X_A|N \sim \text{Bin}(N, p)$ and $X_B|N \sim \text{Bin}(N, 1-p)$.By the chicken-egg theory, $X_A \sim \text{Pois}(\lambda p)$ and $X_B \sim \text{Pois}(\lambda(1-p))$ are independent. (0.5 points)
Then it follows

$$Var(X_A - X_B) = Var(X_A) + Var(X_B) = \lambda p + \lambda(1-p) = \lambda \cdots \text{ (0.5 points)}$$

And

$$\begin{aligned}
&\text{Cov}[X_B, N] \\
=&\text{Cov}[X_B, X_A + X_B] \\
=&\text{Cov}[X_B, X_A] + \text{Cov}[X_B, X_B] \\
=&Var(X_B) \\
=&\lambda(1-p) \cdots \text{ (0.5 points)}
\end{aligned}$$

Then

$$\rho_{X_B,N} = \frac{\text{Cov}[X_B, N]}{sd(X_B)sd(N)} = \frac{\lambda(1-p)}{\sqrt{\lambda(1-p)}\sqrt{\lambda}} = \sqrt{1-p} \cdots \text{ (0.5 points)}$$

Ex 11.143 (2)**.** Suppose $N = 500$. Two new pension schemes are now introduced, called Plan C and Plan D. Each of the 500 employees now independently chooses one of the four pension schemes with equal probabilities $\frac{1}{4}$. Let X_i be the number of employees that choose Plan i, i=A,B,C,D, $\sum_i X_i = N = 500$. Find $\text{Cov}[X_B, X_C]$ and ρ_{X_B,X_C}.

s.11.143. First notice that $X_j \sim \text{Bin}(500, \frac{1}{4})$, $j = A, B, C, D$.(0.5 points)
Then we know $\boldsymbol{X} = (X_A, X_B, X_C, X_D) \sim Mult_4(500, \frac{1}{4})$.(0.5 points)
Using the property of a Multinomial distribution,

$$\text{Cov}[X_B, X_C] = -\frac{500}{4^2}, \cdots \text{ (0.5 points)}$$

$$\rho_{X_B,X_C} = \frac{\text{Cov}[X_B, X_C]}{sd(X_B)sd(X_C)} = -\frac{500/4^2}{500(1/4)(3/4)} = -\frac{1}{3}. \cdots \text{ (0.5 points)}$$

Consider the following codes:

R Code

```
1 library(mvtnorm)
2 set.seed(999)
3 A<-c(1,2)
4 B<-c(2,3)
5 C<-A+B
6 D<-cbind(A,B)
7 X<-rmvnorm(200,mean=C,sigma=D)
8 output<-colMeans(X)
```

Python Code

```
1 import random
2 import numpy as np
3 random.seed(999)
4 A = np.array([1,2])
5 B = np.array([2,3])
6 C = A+B
7 D = np.transpose([A,B])
8 X = np.random.multivariate_normal(C, D, size = 200)
9 output = X.mean(axis=0)
```

Ex 11.144 (1)**.** Explain in detail the purpose of **Line 1, 2, 5, 6, 7, 8** of the above codes.

s.11.144. Line 1: Load the package "mvtnorm" so that we can use the function *rmvnorm*.
Line 2: Set a random seed to reproduce the same results.
Line 5: Generate a vector $C = (3,5)$.
Line 6: Generate a 2×2 matrix of D=$\begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$.
Line 7: Generate 200 multivariate normally distributed variables $\mathbf{X} = x_1, x_2$ with mean equal to $(3,5)$ and variance equal to matrix D.
Line 8: Calculate the average value for each column of matrix X.
(0.5 points for mentioning at least 3 of the above.)

QUESTION

An investor wants to keep track of the daily return of his portfolio. Let X_t be the portfolio daily return on day t, where $X_1, X_2, \ldots$ are i.i.d. r.v.s from a continuous distribution. We say that day t hits a *record low* if the return on day t is lower than on all previous $t-1$ days. Let A_t be the event that day t hits a *record low*, and let I_t be the indicator r.v. that is 1 if day t hits a *record low* and 0 otherwise.

Ex 11.145 (0.5)**.** Find $\mathsf{P}\{A_t\}$, the probability that day t hits the *record low*.

s.11.145. Since all of the first t days are equally likely to have the lowest return, by symmetry, $\mathsf{P}\{A_t\} = \frac{1}{t}$.

Ex 11.146 (1)**.** Find $\mathsf{P}\{A_t \cap A_{t+1}\}$, the probability that the *record low* is hit on both day t and day $t+1$. Are A_t and A_{t+1} independent?

s.11.146. To solve this exercise we use permutations. First notice that in $t+1$ days, there are in total $(t+1)!$ possible combination of the daily returns. Since only the lowest 2 daily returns should be on day $t+1$ and day t, the order of the remaining $t-1$ daily returns does not matter. Thus, $\mathsf{P}\{A_t \cap A_{t+1}\} = \frac{(t-1)!}{(t+1)!} = \frac{1}{t(t+1)}$.

Ex 11.147 (1.5)**.** Show that A_s and A_t are independent if $s < t$. This means that whether day s hits a *record low* does not influence whether day t hits a *record low*, with $s < t$.

s.11.147. To solve this exercise we use permutations. First notice that in t days, there are in total $t!$ possible combination of the daily returns. Since the lowest daily return should be on day t, we need to sort the rest $t-1$ daily returns. Further notice that between day $s+1$ and $t-1$ the daily returns only have to be higher than X_{t+1}, no other restrictions. So we need $t-(s+1)$ our of the remaining $t-1$ daily returns to fill the days between day $s+1$ and $t-1$, in total $\binom{t-1}{t-s-1}(t-s-1)!$ possible combinations. Finally, the lowest out of the remaining s daily returns need to be on day s, and the order of the remaining does not matter. Thus we have:

$$\begin{aligned}
\mathsf{P}\{A_s \cap A_t\} &= \frac{\binom{t-1}{t-s-1}(t-s-1)!(s-1)!}{t!} \\
&= \frac{(s-1)!(t-1)!}{s!t!} \\
&= \frac{1}{st} \\
&= \mathsf{P}\{A_s\}\,\mathsf{P}\{A_t\}.
\end{aligned}$$

Ex 11.148 (2)**.** Let N be the number of *record low* days from day 1 up to t. Find $\mathsf{Cov}[N, I_t]$.

s.11.148. First notice that $\mathsf{Cov}[N, I_t] = \mathsf{E}[NI_t] - \mathsf{E}[N]\,\mathsf{E}[I_t]$. Since $\mathsf{E}[I_t] = \mathsf{P}\{A_t\} = \frac{1}{t}$, we need to find out $\mathsf{E}[NI_t]$ and $\mathsf{E}[N]$. Since $N = \sum_{k=1}^t I_k$,

$$\begin{aligned}
\mathsf{E}[N] &= \mathsf{E}\left[\sum_{k=1}^t I_k\right] \\
&= \sum_{k=1}^t \mathsf{E}[I_k] \\
&= \sum_{k=1}^t \frac{1}{k}
\end{aligned}$$

$$\begin{aligned}
\mathsf{E}[NI_t] &= \mathsf{E}[\mathsf{E}[NI_t \,|\, I_t]] \\
&= \mathsf{E}\left[\sum_{k=1}^t I_k I_t \,\middle|\, I_t = 1\right] \mathsf{P}\{I_t = 1\} \\
&= \mathsf{E}\left[\sum_{k=1}^{t-1} I_k + 1\right] \mathsf{P}\{I_t = 1\} \\
&= \frac{1}{t}\sum_{k=1}^{t-1}\frac{1}{k} + \frac{1}{t}
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathsf{Cov}[N, I_t] &= \mathsf{E}[NI_t] - \mathsf{E}[N]\,\mathsf{E}[I_t] \\
&= \frac{1}{t}\sum_{k=1}^{t-1}\frac{1}{k} + \frac{1}{t} - \frac{1}{t}\sum_{k=1}^{t}\frac{1}{k} \\
&= \frac{1}{t} - \frac{1}{t^2} = \frac{t-1}{t^2}.
\end{aligned}$$

Alternatively, note that $\mathsf{Cov}\left[I_i, I_j\right] = 0$ if $i \neq j$ since I_i and I_j are independent if $i \neq j$. Hence,

$$\begin{aligned}
\mathsf{Cov}[N, I_t] &= \mathsf{Cov}\left[\sum_{k=1}^t I_k, I_t\right] \\
&= \sum_{k=1}^t \mathsf{Cov}[I_k, I_t] \\
&= \mathsf{Cov}[I_t, I_t] = \mathsf{V}[I_t] = \mathsf{E}\left[I_t^2\right] - \mathsf{E}[I_t]^2 \\
&= \frac{1}{t} - \frac{1}{t^2} = \frac{t-1}{t^2}.
\end{aligned}$$

Remarks and grading scheme:

1. Ex 3.2: only full points if the permutation is well explained. 0.5 point for no or bad explanation.
2. Ex 3.4: 0.5 point for writing out the formula for covariance. 0.5 point for correctly calculated $\mathsf{E}[N]$, $\mathsf{E}[I_t]$.

QUESTION

X number of people will get vaccinated for Covid-19. There are currently two types of vaccines, Vaccine A and Vaccine B. Each person independently choose vaccine A with probability p, and vaccine B with probability $1-p$.

Ex 11.149 (2)**.** Suppose $X \sim \text{Poi}(\lambda)$. Let X_A be the number of people that choose Vaccine A and $X_B = X - X_A$ be the number of people that choose Vaccine B. Find $Var(X_A - X_B)$ and $\rho_{X_A,X}$.

s.11.149. First notice that $X_A|X \sim \text{Bin}(X,p)$ and $X_B|X \sim \text{Bin}(X,1-p)$. By the chicken-egg theory, $X_A \sim \text{Pois}(\lambda p)$ and $X_B \sim \text{Pois}(\lambda(1-p))$ are independent. (0.5 points)
Then it follows

$$Var(X_A - X_B) = Var(X_A) + Var(X_B) = \lambda p + \lambda(1-p) = \lambda \cdots \text{ (0.5 points)}$$

And

$$\begin{aligned} &\operatorname{Cov}[X_A, X] \\ =&\operatorname{Cov}[X_A, X_A + X_B] \\ =&\operatorname{Cov}[X_A, X_A] + \operatorname{Cov}[X_A, X_B] \\ =&Var(X_A) \\ =&\lambda p \cdots \text{ (0.5 points)} \end{aligned}$$

Then

$$\rho_{X_A,X} = \frac{\operatorname{Cov}[X_A, X]}{sd(X_A)sd(X)} = \frac{\lambda p}{\sqrt{\lambda p}\sqrt{\lambda}} = \sqrt{p} \cdots \text{ (0.5 points)}$$

Ex 11.150 (2)**.** Suppose $X = 1000$. A new vaccine is now available, called Vaccine C. Each of the 1000 people now independently chooses one of the three vaccines with equal probabilities $\frac{1}{3}$. Let X_A be the number of people that choose Vaccine i, i=A,B,C and $\sum_i X_i = X = 1000$.
Calculate $\operatorname{Cov}[X_A, X_C]$ and ρ_{X_A,X_C}.

s.11.150. First notice that $X_j \sim \text{Bin}(1000, \frac{1}{3})$, $j = A, B, C$. (0.5 points)
Then we know $\boldsymbol{X} = (X_A, X_B, X_C) \sim Mult_3(1000, \frac{1}{3})$. (0.5 points)
Using the property of a Multinomial distribution,

$$\operatorname{Cov}[X_A, X_C] = -\frac{1000}{3^2}, \cdots \text{ (0.5 points)}$$

$$\rho_{X_A,X_C} = \frac{\operatorname{Cov}[X_A, X_C]}{sd(X_A)sd(X_C)} = -\frac{1000/3^2}{1000(1/3)(2/3)} = -\frac{1}{2}. \cdots \text{ (0.5 points)}$$

Consider the following codes:

R Code

```r
1 library(mvtnorm)
2 set.seed(888)
3 A<-c(1,2,1)
4 B<-c(2,3,1)
5 C<-c(1,1,8)
6 D<-cbind(A,B,C)
7 X<-rmvnorm(200,mean=A,sigma=D)
8 output=cor(X[,1]+X[,2],X[,3])
```

Python Code

```python
1 import random
2 import numpy as np
3 random.seed(888)
4 A = [1,2,1]
5 B = [2,3,1]
6 C = [1,1,8]
7 D = np.transpose([A,B,C])
8 X = np.random.multivariate_normal(A, D, size = 200)
9 output = np.corrcoef(X[:,0]+X[:,1]+X[:,2])
```

Ex 11.151 (1)**.** Explain in detail the purpose of **Line 1,2, 6,7,8** of the above codes.

s.11.151. Line 1: Load the package "mvtnorm" so that we can use the function *rmvnorm*.
Line 2: Set a random seed to reproduce the same results.
Line 6: Combine vector A, B, C to generate a matrix D=$\begin{pmatrix} 1 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 8 \end{pmatrix}$
Line 7: Generate 200 multivariate normally distributed variables $\mathbf{X} = x_1, x_2, x_3$ with mean equal to $(1,2,1)$ and variance equal to matrix D.
Line 8: Calculate the correlation of the sum of column 1-2 of matrix X and the 3rd column of matrix X.
(0.5 points for correct answer of at least 3 of the above.)

QUESTION

A random point (X,Y) is chosen in the following square:

$$\{(x,y) : x^2 < 7, y^2 < 7\}$$

All points are equally likely to be chosen. Let S be the squared norm of (X,Y).

Ex 11.152 (0.5)**.** Find the joint PDF $f(x,y)$ of X and Y.

s.11.152. So we can see that both X and Y are uniformly distributed on $(-\sqrt{7},\sqrt{7})$. Then their joint PDF is simply:

$$\begin{aligned} f(x,y) =& \left(\frac{1}{\sqrt{7}-(-\sqrt{7}))}\right)\left(\frac{1}{(\sqrt{7}-(-\sqrt{7}))}\right) \\ =& \frac{1}{28} \end{aligned}$$

One mistake, zero points.

Ex 11.153 (0.5)**.** Show that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$

s.11.153.

$$\begin{aligned} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy =& \int_{-\sqrt{7}}^{\sqrt{7}}\int_{-\sqrt{7}}^{\sqrt{7}} \frac{1}{28}\,dx\,dy \\ =& 1 \end{aligned}$$

One mistake, zero points.

Ex 11.154 (3)**.** Find the expectation of S, i.e., the squared norm of (X,Y).

s.11.154. Note that $S = X^2 + Y^2$. Using LOTUS:

$$
\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(x^2+y^2\right)f(x,y)\,dx\,dy &= \int_{-\sqrt{7}}^{\sqrt{7}}\int_{-\sqrt{7}}^{\sqrt{7}}\left(x^2+y^2\right)\left(\frac{1}{28}\right)dx\,dy \\
&= \int_{-\sqrt{7}}^{\sqrt{7}}\int_{-\sqrt{7}}^{\sqrt{7}}\frac{x^2+y^2}{28}\,dx\,dy \\
&= \frac{1}{28}\int_{-\sqrt{7}}^{\sqrt{7}}\int_{-\sqrt{7}}^{\sqrt{7}}x^2+y^2\,dx\,dy \\
&= \frac{1}{28}\int_{-\sqrt{7}}^{\sqrt{7}}\left[\frac{x^3}{3}+y^2x\right]_{-\sqrt{7}}^{\sqrt{7}}dy \\
&= \frac{1}{28}\int_{-\sqrt{7}}^{\sqrt{7}}\left(\frac{7^{\frac{3}{2}}}{3}+\sqrt{7}y^2-\frac{(-7)^{\frac{3}{2}}}{3}+\sqrt{7}y^2\right)dy \\
&= \frac{1}{28}\int_{-\sqrt{7}}^{\sqrt{7}}\left(\frac{7^{\frac{3}{2}}}{3}+\sqrt{7}y^2+\frac{7^{\frac{3}{2}}}{3}+\sqrt{7}y^2\right)dy \\
&= \frac{1}{28}\int_{-\sqrt{7}}^{\sqrt{7}}\frac{2}{3}7^{\frac{3}{2}}+2\sqrt{7}y^2\,dy \\
&= \frac{1}{28}\left[\frac{2}{3}7^{\frac{3}{2}}y+\frac{2}{3}\sqrt{7}y^3\right]_{-\sqrt{7}}^{\sqrt{7}} \\
&= \frac{1}{28}\left(\frac{2}{3}7^2+\frac{2}{3}7^2+\frac{2}{3}(-7)^2+\frac{2}{3}(-7)^2\right) \\
&= \frac{1}{28}\frac{8}{3}7^2=\frac{14}{3}=4\frac{2}{3}
\end{aligned}
$$

One point for finding $S^2 = X^2 + Y^2$ and writing down the integral correctly using LOTUS. Two points for the remaining calculations.

Consider the following code:

Python Code

```python
import numpy as np
import math
np.random.seed(3)

num = 100000
distances = np.zeros(num)
for i in range(0,num):
    angle = 2*math.pi*np.random.uniform(0,1,1)
    position = np.sqrt(np.random.uniform(0,1,1))
    x = np.cos(angle)*position
    y = np.sin(angle)*position
    distances[i] = np.sqrt(x**2 + y**2)

print(np.mean(distances))
```

Ex 11.155 (0.5)**.** What does the code above do? *Hint:* unit circle.

s.11.155. It loads the required packages and defines a vector of zeros of length `N`. Then in the for-loop it creates a random point in the unit circle. It computes the distance from this point to the origin and stores it in `distances`. Finally, it takes the mean to estimate the mean distance from a randomly chosen point in the unit circle to the origin.
0.5 points for mentioning an average distance is between a random point and the origin is estimated. This point has an x-coordinate and a y-coordinate.

Ex 11.156 (0.5)**.** The output of the code is 0.66629. Explain this result.

s.11.156. The mean distance from a randomly selected point in the unit circle to the origin is $\approx \frac{2}{3}$.
Trivial

QUESTION

A portfolio manager wants to investigate the monthly return of a particular portfolio, starting from an arbitrary month in history. Let X_t be the portfolio monthly return on month t, with $X_1, X_2, \ldots$ i.i.d. We say that month t is *worst in a year* if the return in month t is lower than all previous 11 months. Let A_t be the event that month t is the *worst in a year*, and let I_t be the indicator r.v. that is 1 if month t hits a *worst in a year* an 0 otherwise.

Ex 11.157 (0.5)**.** Find $\mathsf{P}\{A_t\}$, the probability that month t is the *worst in a year*, $t \geq 12$.

s.11.157. Since the current month and all previous 11 months are equally likely to have the lowest return, by symmetry, $\mathsf{P}\{A_t\} = \frac{1}{12}$.

Ex 11.158 (2)**.** Let N be the number of *worst in a year* months from the month 12 to month t. Find $\mathsf{E}[N]$ and $\mathsf{Cov}[N, I_t]$.

s.11.158. Since $N = \sum_{k=12}^{t} I_k$ and $\mathsf{E}[I_t] = \mathsf{P}\{A_t\} = \frac{1}{12}$,

$$\mathsf{E}[N] = \mathsf{E}\left[\sum_{k=12}^{t} I_k\right] = \sum_{k=12}^{t} \mathsf{E}[I_k] = \frac{t-11}{12}.$$

Then notice that $\mathsf{Cov}[N, I_t] = \mathsf{E}[NI_t] - \mathsf{E}[N]\,\mathsf{E}[I_t]$. Since $\mathsf{E}[I_t] = \mathsf{P}\{A_t\} = \frac{1}{12}$ and $\mathsf{E}[N] = \frac{t-11}{12}$ we need to find $\mathsf{E}[NI_t]$.

$$\begin{aligned}
\mathsf{E}[NI_t] &= \mathsf{E}[\mathsf{E}[NI_t \mid I_t]] \\
&= \mathsf{E}\left[\sum_{k=12}^{t} I_k I_t \,\middle|\, I_t = 1\right] \mathsf{P}\{I_t = 1\} \\
&= \mathsf{E}\left[\sum_{k=12}^{t-1} I_k + 1\right] \mathsf{P}\{I_t = 1\} \\
&= \left(\frac{t-12}{12} + 1\right)\frac{1}{12} = \frac{t}{144}
\end{aligned}$$

Thus:

$$\begin{aligned}
\mathsf{Cov}[N, I_t] &= \mathsf{E}[NI_t] - \mathsf{E}[N]\,\mathsf{E}[I_t] \\
&= \frac{t}{144} - \frac{t-11}{12}\frac{1}{12} = \frac{11}{144}.
\end{aligned}$$

Ex 11.159 (1.5)**.** Find $\mathsf{P}\{A_t \cap A_{t+1}\}$, the probability that two consecutive months are both *worst in a year*. Are A_t and A_{t+1} independent?

s.11.159. To solve this exercise we use permutations. First notice that since we involve 2 months, there are in total 12+1=13 possible combination of the monthly returns. Since only the lowest 2 monthly returns should be on month $t+1$ and month t, the order of the remaining 13-2=11 monthly returns does not matter. Thus, $\mathsf{P}\{A_t \cap A_{t+1}\} = \frac{(13-2)!}{13!} = \frac{1}{12*13} = \frac{1}{156}$.
Since $\mathsf{P}\{A_t\} = \mathsf{P}\{A_{t+1}\} = \frac{1}{12}$, $\mathsf{P}\{A_t \cap A_{t+1}\} \neq \mathsf{P}\{A_t\}\,\mathsf{P}\{A_{t+1}\}$, and A_t and A_{t+1} are not independent.

Ex 11.160 (1)**.** Let B_t be the event that return in month t is lower than all previous months. Find $\mathsf{P}\{B_t \cap B_{t+1}\}$.

s.11.160. To solve this exercise we use permutations. First notice that in $t+1$ months, there are in total $(t+1)!$ possible combination of the monthly returns. Since only the lowest 2 monthly returns should be on month $t+1$ and month t, the order of the remaining $t-1$ monthly returns does not matter. Thus, $\mathsf{P}\{B_t \cap B_{t+1}\} = \frac{(t-1)!}{(t+1)!} = \frac{1}{t(t+1)}$.

Remarks and grading scheme:

1. First notice that A_t and A_{t+1} are not independent. This is illustrated in Exercise 3.3. Many students wrongly assumed their independence.
2. Ex 3.2: 0.5 point for correctly calculated $\mathsf{E}[N]$, $\mathsf{E}[I_t]$. 0.5 point for writing out the formula for covariance.
3. Ex 3.3: No point if you assume them to be independent before solving the question. Even though you might have also got the same answer in the end by accident.

QUESTION

Suppose you received a collection of books as your birthday gift. You already read 2 of them and there are still 4 books left. Let X_1, X_2 be the number of pages (in hundreds of pages) of the first 2 books you read, and let $X_3, \dots, X_6$ be the number of pages (in hundreds of pages) of the remaining books. Assume that $X_i \sim \text{Norm}(4,1)$ for $i = 1, \dots, 6$.

Ex 11.161 (1.5)**.** First assume that the number of pages of the books are all independent. What is the expected number of remaining books that have more pages than each of the 2 books you have already read?

s.11.161. Let I_i be the indicator r.v. for the ith book having more page than each of book 1 and book 2. Then:

$$\begin{aligned} \mathsf{P}\{I_i = 1\} &= \mathsf{P}\{X_i > X_1, X_i > X_2\} \\ &= \mathsf{P}\{X_i = \max\{X_1, X_2, X_i\}\} \\ &= \frac{1}{3}, \end{aligned}$$

by symmetry. Then $\mathsf{E}\left[\Sigma_{i=3}^{6} I_i\right] = \frac{1}{3} \cdot 4 = \frac{4}{3}$.

For the next two exercises, suppose that $(X_1, \dots, X_6)$ is now Multivariate Normal distributed with $\mathsf{Corr}\left[X_1, X_j\right] = \frac{1}{2}$ for $3 \le j \le 6$.

Ex 11.162 (2.5)**.** On average, how many of the remaining books are at least 100 pages longer than the first book you read?

s.11.162. In order to answer this question, we want to know $P(X_i - X_1 > 1)$, for $i = 3, \dots, 6$. We first consider $i = 3$. Since X_3 and X_1 are jointly normal distributed, $X_3 - X_1$ is also normally distributed, with $\mathsf{E}[X_3 - X_1] = \mathsf{E}[X_3] - \mathsf{E}[X_1] = 0$ and

$$\begin{aligned} \mathsf{V}[X_3 - X_1] &= \mathsf{V}[X_3] + \mathsf{V}[-X_1] + 2\,\mathsf{Cov}[X_3, -X_1] \\ &= 1 + 1 - 2 \cdot \frac{1}{2} \cdot 1 \cdot 1 \\ &= 1 \end{aligned}$$

Then we know $X_3 - X_1$ follows a standard normal distribution and $P(X_3 - X_1 > 1) = 0.16$ Similarly, $P(X_4 - X_1 > 1) = P(X_5 - X_1 > 1) = P(X_6 - X_1 > 1) = 0.16$. So the average number of the remaining books that has 100 pages more than the first book is $0.16 \cdot 4 = 0.64$.

Ex 11.163 (1)**.** Show that there exists a constant c such that $X_1 - cX_3$ and X_3 are independent, and determine the value of c.

s.11.163. $\mathsf{Cov}[X_1 - cX_3, X_3] = \mathsf{Cov}[X_1 - cX_3, X_3] = \mathsf{Cov}[X_1, X_3] - c\,\mathsf{V}[X_3] = \frac{1}{2} - c$, so for $c = \frac{1}{2}$, we have that $X_1 - cX_3$ and X_3 are uncorrelated. Since $(X_1, \dots, X_6)$ has the multivariate normal distribution, it follows that $X_1 - cX_3$ and X_3 are independent.

Remarks and grading scheme:

1. Ex 3.1: Many students assume that $X_i > X_1$ and $X_i > X_2$ is independent. This is not the case. In fact, If $X_i > X_1$, then it's more likely that X_i is large. In consequence, it is also more likely that $X_i > X_2$.

2. Ex 3.1: 0.5 point for multiply your probability with 4 (even if it is calculated wrongly). Full point(1.5) for correct answer.

3. Ex 3.2: 0.5 point for mentioning that $X_3 - X_1$ is normally distributed ,0.5 point for correctly calculated $\mathsf{E}[X_3 - X_1]$ and 0.5 point for correctly calculated $Var(X_3 - X_1)$.

4. Ex 3.3: 0.5 point for writing out the formula for covariance.

QUESTION

A server spends a random amount T on a job. While the server works on the job, it sometimes gets interrupted to do other tasks $\{R_i\}$. Such tasks need to be repaired before the machine can continue working again. It is clear that such interruptions can only occur when the server is working. Assume that the interruptions arrive according to a Poisson process with rate λ, i.e., the number of failures $N(t) \sim \text{Pois}(\lambda t)$. The interruptions $\{R_i\}$ are independent of N and of T and form an iid sequence with common mean $\mathsf{E}[R]$ and variance $\mathsf{V}[R]$. The duration S of a job is its own service time T plus all interruptions, hence $S = T + \sum_{i=1}^{N(T)} R_i$.

Ex 11.164 (1.5)**.** In the computation of $\mathsf{V}[S]$ we encounter the following steps.

$$\mathsf{V}\left[\sum_{i=1}^{N(t)} R_i\right] = \mathsf{E}\left[\mathsf{V}\left[\sum_{i=1}^{N(t)} R_i \,\middle|\, N(t)\right]\right] + \mathsf{V}\left[\mathsf{E}\left[\sum_{i=1}^{N(t)} R_i \,\middle|\, N(t)\right]\right]. \tag{11.0.220}$$

The computation below consists of a number of steps, a, b, ... Explain for each step which property is used to ensure the step is true.

$$\mathsf{V}\left[\sum_{i=1}^{N(t)} R_i \,\middle|\, N(t) = n\right] \stackrel{a}{=} \mathsf{V}\left[\sum_{i=1}^{n} R_i\right] \tag{11.0.221}$$

$$\stackrel{b}{=} n\mathsf{V}[R] \tag{11.0.222}$$

$$\mathsf{E}\left[\mathsf{V}\left[\sum_{i=1}^{N(t)} R_i \,\middle|\, N(t)\right]\right] \stackrel{c}{=} \mathsf{E}[N(t)\mathsf{V}[R]] \tag{11.0.223}$$

$$\stackrel{d}{=} \lambda t\mathsf{V}[R]. \tag{11.0.224}$$

s.11.164. a. On the set $\{N(t) = n\}$ $N(t) = n$. Hence we can replate $N(t)$ by n.

b. Variance of sum of iid rvs is sum of variance of rv.

c. Use a and b, substitute $N(t)$ for n and use the definition of conditional expectation.

d. Since $\mathsf{V}[R]$ is a constant, we can take out of the expectation. Expectation of Poisson rv is λt
Each property missed, e.g, linearity of expectation, minus 0.5.

Ex 11.165 (2)**.** Suppose R is equal to the constant r and $T \sim \text{Unif}([0, a])$, compute $\mathsf{E}[S]$.

s.11.165. Since $T = t$ a.s.,

$$\mathsf{E}[S] = \mathsf{E}[T] + \mathsf{E}[N(T)]\,\mathsf{E}[R] = a/2 + r\lambda\mathsf{E}[T] = a/2 + r\lambda a/2.$$

We know that $\mathsf{E}[T] = a/2$, Hence writing $\mathsf{E}[N(t)] = \lambda t$, is not ok. Some other strange things that I saw:

$$R \neq rT \tag{11.0.225}$$

$$\mathsf{E}[R] \neq ra/2 \tag{11.0.226}$$

$$\tag{11.0.227}$$

Typically, 0.5 point for each correct line, but I subtracted points If you say that $\mathsf{E}[N(T)\mathsf{E}[R]] = \mathsf{E}[R]$, or write $n\mathsf{E}[R]$ as final answer (apparently you did not get the idea that N is an rv.)

Ex 11.166 (1.5)**.** Explain what this code computes.

Python Code

```python
import numpy as np

labda = 0.5
size = 10
num_runs = 50


def do_run():
    T = np.random.uniform(0, 20)
    N = np.random.poisson(labda * T)
    R = np.random.uniform(1, 5, size=N)
    S = T + R.sum()
    return S


samples = np.zeros(num_runs)
for i in range(num_runs):
    samples[i] = do_run()


print((samples > 8).sum())
```

R Code

```r
labda <- 0.5
size <- 10
num_runs <- 50

do_run <- function() {
  bigT <- runif(n = 1, min = 0, max = 20)
  N <- rpois(n = 1, labda * bigT)
  R <- runif(n = N, min = 1, max = 5)
  S <- bigT + sum(R)
  return(S)
}

samples <- rep(0, num_runs)
for (i in 1:num_runs) {
  samples[i] <- do_run()
}

print(sum(samples > 8))
```

Hint, you should know that in P.21 (R18) the string `samples > 8` collects only the samples with value larger than 8.

s.11.166. Here is an EXPLANATION:

1. T is a simulated service time of a job without interruptions.
2. N is is the simulated number of failures that occur during the service of the job
3. R is then a vector of simulated durations of each of the interruptions
4. Hence, S is the total time the job spends at the server
5. We sample the total service time a number of times, and count how often (mean or variance) this exceeds a threshold.

The points earned depends on how clear you explanation is.

Quite a few of you don't seem to understand what it means to *explain* something. Here are some answers that I saw that certainly don't explain what the code does:

1. 'The code does what's stated in the exercise.'. What's the explanation here? The question is also not: do you understood what the code does?
2. 'T is uniform rv, N is a Poisson rv, R is a also a uniform rv. We repeat this a few times.' Like this you just read the code, but I know you can read, so this type of answer is quite useless.
3. 'We collect a subset of the samples and print that.' This answer is just a repition of the hint.

So, next time, when you are asked to explain something, explain it. In your professional carreer, if you read out loud to your customer what s/he can read for him/herself, you'll be fired pretty soon.

QUESTION

Consider the following code:

Python Code

```python
import numpy as np
from scipy.stats import expon
np.random.seed(42)

n = 100
N = 1000

X = expon(scale = 1/2).rvs([N, n])
Y = X.mean(axis = 1)

mu = 1/2
sigma = 1/2
Z = np.sqrt(n) * (Y - mu)/sigma

print((Z ** 37).mean())
```

R Code

```r
set.seed(42)

n <- 100
N <- 1000

X <- matrix(rexp(N * n, rate = 2), nrow = N, ncol = n)
Y <- rowMeans(X)

mu <- 1/2
sigma <- 1/2
Z <- sqrt(n) * (Y - mu)/sigma

print(mean(Z^37))
```

Ex 11.167 (1.5)**.** Y is a vector of i.i.d. random variables.

(i) What is the length ℓ of Y?

(ii) Each element of Y is a mean of k i.i.d. Exp(λ) r.v.s. What are k and λ?

(iii) What are the (theoretical) expectation and variance of an element of Y?

s.11.167. The length of Y is $N = 1000$; Each element of Y is a mean of $k = n = 100$ i.i.d. Exp(2) r.v.s. The expectation is $\frac{1}{\lambda} = \frac{1}{2}$ and the variance is $\frac{1}{k\lambda^2} = \frac{1}{400} = 0.0025$.

Grading scheme:

- 0.5 for getting both the length $N = 1000$ and $k = 100$ correct (no partial credit);
- 0.5 for $\lambda = 2$, expectation $\frac{1}{2}$ and the factor $\frac{1}{4}$ in the variance (no partial credit);
- 0.5 for the factor $\frac{1}{k}$ in the variance.

Recall that each element of Y is the mean of k i.i.d. Exp(λ) r.v.s.

Ex 11.168 (1)**.** What is the exact distribution of an element of Y?
Give its name and its parameters, and explain the answer.

s.11.168. The sum of k i.i.d. Exp(λ) r.v.s. has the Gamma(k, λ) distribution. Hence, the mean of k i.i.d. Exp(λ) r.v.s. has the Gamma$(k, k\lambda)$ distribution. In this exercise, $k = 100$ and $k\lambda = 200$.
Grading scheme:

- 0.5 for Gamma with first parameter k
- 0.5 for the second parameter

Let $(Y_1, \dots, Y_\ell)$ be the elements of Y and let $(Z_1, \dots, Z_\ell)$ be the elements of Z. Recall that each Z_i depends on k because Y_i is the mean of k i.i.d. Exp(λ) r.v.s. Let T be the random variable to which Z_1 converges in the limit $k \to \infty$.

Ex 11.169 (1)**.** What is the distribution of T, and why?
Hence, what is an approximate distribution of an element of Y (e.g. Y_1) and why?

s.11.169. By the CLT, $Z_1 \overset{\cdot}{\sim} \mathrm{Norm}(0,1)$. Hence, $Y_1 = \mu + \sigma/\sqrt{n} Z_1 \overset{\cdot}{\sim} \mathrm{Norm}(0.5, 0.25/n)$.
Grading scheme:

- 0.5 for mentioning CLT and the distribution of Z_1;
- 0.5 for the approximate distribution of Y_1.

Ex 11.170 (0.5)**.** In the last line of the code, we compute $S = \frac{1}{\ell}\sum_{i=1}^{\ell} Z_i^{37}$.
If $k \to \infty$ (for fixed ℓ, e.g. $\ell = 3$), does S converge to a constant?
You do not have to give formal proofs in this subquestion, but you should give clear explanations using the appropriate theorem(s).

s.11.170. In the limit $k \to \infty$, each Z_i has the standard normal distribution by CLT, but the sum of ℓ powers of normal distributions has a non-trivial CDF.

- 0.5 for explaining that S does not converge to a constant; no points for just CLT (unless previous question was not answered).

Ex 11.171 (1)**.** In the last line of the code, we compute $S = \frac{1}{\ell}\sum_{i=1}^{\ell} Z_i^{37}$.

If $\ell \to \infty$ (for fixed k), does S converge to a constant? If so, does it converge to $\mathsf{E}\left[T^{37}\right]$?

You do not have to give formal proofs in this subquestion, but you should give clear explanations using the appropriate theorem(s).

s.11.171. By LLN, S does converge to a constant as $\ell \to \infty$, however, it converges to $\mathsf{E}\left[Z_1^{37}\right]$ for that fixed value of k. By symmetry, we have $\mathsf{E}\left[Z_1^{37}\right] = 0$. However, the gamma distribution is right-skewed, which implies $\mathsf{E}\left[T^{37}\right] > 0$. Hence, it does not converge to $\mathsf{E}\left[T^{37}\right]$.
Grading scheme:

- 0.5 for concluding that S converges to a constant using LLN.
- 0.5 for explaining why the constant is not equal to $\mathsf{E}\left[T^{37}\right]$.

QUESTION

Let $Z \sim \text{Norm}(0,1)$. In this exercise, we find an upper bound for $\mathsf{P}\{|Z|>2\}$.

Ex 11.172 (1.5)**.** Let f be a positive and increasing function, and let X be a r.v.
Consider the following inequality:

$$\mathsf{P}\{X \geq a\} = \mathsf{P}\left\{f(X) \geq f(a)\right\} \leq \frac{\mathsf{E}\left[f(X)\right]}{f(a)}.$$

(i) Explain why $\mathsf{P}\{X \geq a\} = \mathsf{P}\left\{f(X) \geq f(a)\right\}$ holds.

(ii) Explain why $\mathsf{P}\left\{f(X) \geq f(a)\right\} \leq \frac{\mathsf{E}[f(X)]}{f(a)}$ holds.

Make sure to clearly indicate where you use that f is positive and increasing.

s.11.172. (i). Since f is increasing, we have $X \geq a$ if and only if $f(X) \geq f(a)$, so $\{X \geq a\}$ and $\{f(X) \geq f(a)\}$ are the same event. Hence, $\mathsf{P}\{X \geq a\} = \mathsf{P}\left\{f(X) \geq f(a)\right\}$.
(ii). Since f is positive, we have $|f(X)| = f(X)$ and $f(a) > 0$. Hence, the inequality follows directly from Markov's inequality with r.v. $f(X)$ and constant $f(a) > 0$.

Remarks and grading scheme:

- In part (ii), it is not sufficient to just say that it holds by Markov's inequality, also explain how you apply Markov's inequality.
- Since it was asked to indicate **clearly** where you use that f is positive and increasing, don't just say "since f is positive and increasing" twice. For part (i) the positivity is not needed, it is only needed that f is increasing. For part (ii), it is not needed that f is increasing, only the positivity is needed. Moreover, some explanation should be given how it is used.
- f is not necessarily a linear transformation, as some of you claimed.
- Not every positive and increasing function is convex or concave.
  And using Jensen's inequality certainly doesn't work in (i): you are asked to prove an equality of probabilities, not an inequality of expectations.
- In part (i), the **if and only if** is really important. If you just mention that $f(X) \geq f(a)$ if $X \geq a$, then you are only proving that $\mathsf{P}\{X \geq a\} \leq \mathsf{P}\left\{f(X) \geq f(a)\right\}$, because if you just say "$f(X) \geq f(a)$ if $X \geq a$", $f(X) \geq f(a)$ could still be true in cases where $X \geq a$ is not, and hence $f(X) \geq f(a)$ can still have a larger probability. (I've not been strict on this when grading.)
- Grading: 0.5 for a sufficient explanation for (i), 0.5 for a sufficient explanation for (ii) and 0.5 for clearly indicating where it is used that f is positive and increasing and writing a clear answer overall.

Ex 11.173 (1)**.** Prove that $\mathsf{P}\{|Z| > 2\} \leq e^{-4t}\,\mathsf{E}\left[e^{tZ^2}\right]$ for $t > 0$.

s.11.173. Note that $f(x) = e^{tx^2}$ is positive and increasing on $(0, \infty)$ for $t > 0$. By applying the inequality of the first question with $a = 2$ we find

$$\mathsf{P}\{|Z| > 2\} \leq e^{-4t}\,\mathsf{E}\left[e^{t|Z|^2}\right] = e^{-4t}\,\mathsf{E}\left[e^{tZ^2}\right].$$

Remarks and grading scheme:

- Most of you did not apply the inequality from the previous part to solve this question, but instead used Chernoff's inequality as follows:

  $$\mathsf{P}\{|Z| > 2\} = \mathsf{P}\left\{Z^2 > 4\right\} \leq e^{-4t}\,\mathsf{E}\left[e^{t|Z|^2}\right] = e^{-4t}\,\mathsf{E}\left[e^{tZ^2}\right],$$

  where the first equality holds since $|Z| > 2$ if and only if $\mathsf{P}\left\{Z^2 > 4\right\}$. This is also correct, but takes a bit more time.

- Don't write nonsense like $e^{-2t}\,\mathsf{E}\left[e^{t|Z|}\right] = e^{-4t}\,\mathsf{E}\left[e^{tZ^2}\right]$, just to make it look like you solved the exercise although you didn't.

- Grading scheme: 1 point for solving the exercise. Partial credit (0.5) for a correct and relevant application of Chernoff's inequality.

Ex 11.174 (2.5)**.** For which t do we find the best upper bound for $\mathsf{P}\{|Z| > 2\}$? Also calculate the upper bound for this value of t.

Hint 1. You may use that if $Y \sim \chi_1^2$, then the MGF of Y is given by $M_Y(t) = (1-2t)^{-1/2}$ for $t < 1/2$. However, you should explain clearly how you use this fact.

Hint 2. Do not forget to check the second order condition of minimization.

s.11.174. Since $Z^2 \sim \chi_1^2$, we have $\mathsf{E}\left[e^{tZ^2}\right] = \mathsf{E}\left[e^{tY}\right] = M_Y(t) = (1-2t)^{-1/2}$.

So we minimize $e^{-4t}\,\mathsf{E}\left[e^{tZ^2}\right] = e^{-4t}(1-2t)^{-1/2}$. It is easier if we take the logarithm first and minimize $-4t - \frac{1}{2}\log(1-2t)$. Its derivative to t is $-4 + \frac{1}{1-2t}$, so setting the derivative to 0 yields $t = 3/8$. The second derivative to t is $\frac{2}{(1-2t)^2} > 0$ (the value at $t = 3/8$ is 32), so the second order condition holds.

This yields $\mathsf{P}\{|Z| > 2\} \leq e^{-3/2}(1-3/4)^{-1/2} \approx 0.446$.

Remarks and grading scheme:

- Unless explicitly noted, you have to give an analytical answer and derive everything using just pen and paper.

- It is essential that you include some derivations for computing the (second) derivative.

- It is not necessary to take the logarithm first, but it makes the derivation easier, and hence it is less likely that you make errors.

- Grading scheme: 0.5 for arguing that $\mathsf{E}\left[e^{tZ^2}\right] = (1-2t)^{-1/2}$ with sufficient explanation; 1 for taking the derivative and calculating that it is 0 at $t = 3/8$; (0.5 if small mistake is made but resulting t satisfies $0 < t < 1/2$, or if it is remarked that this should be the case); 0.5 for calculating the second derivative and checking the second order condition; 0.5 for filling in $t = 3/8$ to provide the upper bound (if an incorrect value of t is found, this point can be given only if the resulting bound is between 0.01 and 1, or if it is explicitly noted that the answer does not make sense).

QUESTION

Amy and Bob are playing a dice game. Every (fair) die has three pairs of identical sides: a pair of ones, a pair of twos and a pair of threes. Each player throws a single die. Let X and Y denote the outcome of Amy and Bob's throw, respectively. The person that throws the highest number wins. If both throw the same number, they have a draw. The final score A of Amy is determined as follows:

- If Amy loses, then she gets zero points.
- If Amy and Bob draw, then she gets 0.5 point.
- If Amy wins, then her score is the difference $X - Y$ in the numbers they threw.

The final score B for Bob is determined analogously. Assume that the dice are fair and that all throws are independent.

Ex 11.175 (1)**.** Determine the joint distribution of X and Y conditional on Amy winning.

s.11.175. The initial distribution is $\mathsf{P}\{X = x, Y = y\} = 1/9$, $x, y = 1, 2, 3$. Amy wins iff $X > Y$, i.e., iff $(X, Y) \in \{(2, 1), (3, 1), (3, 2)\}$. So we have

$$\mathsf{P}\{X = x, Y = y | X > Y\} = \frac{\mathsf{P}\{X = x, Y = y, X > Y\}}{\mathsf{P}\{X > Y\}} \tag{11.0.228}$$

$$= \frac{1/9}{1/3} = 1/3, \tag{11.0.229}$$

for $(x, y) \in \{(2, 1), (3, 1), (3, 2)\}$.

Ex 11.176 (1.5)**.** Find Amy's expected score conditional on Amy winning.

s.11.176. We have

$$A = \begin{cases} 0 & X < Y \\ 1/2 & X = Y \\ X - Y & X > Y. \end{cases} \tag{11.0.230}$$

Hence,

$$\mathsf{E}[A|X > Y] = \mathsf{E}[X - Y|X > Y] \tag{11.0.231}$$

$$= \frac{1}{3}(2-1) + \frac{1}{3}(3-1) + \frac{1}{3}(3-2) \tag{11.0.232}$$

$$= 4/3. \tag{11.0.233}$$

Ex 11.177 (1.5)**.** Find Amy's (unconditional) expected score.

s.11.177. We have, by the law of total expectation,

$$
\begin{aligned}
\mathsf{E}[A] &= \mathsf{P}\{X<Y\}\,\mathsf{E}[A|X<Y] + \mathsf{P}\{X=Y\}\,\mathsf{E}[A|X=Y] + \mathsf{P}\{X>Y\}\,\mathsf{E}[A|X>Y] && (11.0.234)\\
&= 0 + \frac{1}{3}\cdot\frac{1}{2} + \frac{1}{3}\frac{4}{3} && (11.0.235)\\
&= 11/18. && (11.0.236)
\end{aligned}
$$

To make the game more interesting, Amy and Bob decide to play for money. After playing the dice game and scoring A points, Amy receives an amount of T euros, where T is determined randomly. Here, conditional on the outcome of A, T follows a uniform distribution on $[A, 2A]$.

Ex 11.178 (1)**.** What is the (unconditional) expected reward for Amy? That is, compute $\mathsf{E}[T]$.

s.11.178. By Adam's law,

$$
\begin{aligned}
\mathsf{E}[T] &= \mathsf{E}[\mathsf{E}[T|A]] && (11.0.237)\\
&= \mathsf{E}\left[\frac{3}{2}A\right] && (11.0.238)\\
&= \frac{3}{2}\mathsf{E}[A] && (11.0.239)\\
&= \frac{3}{2}\frac{11}{18} && (11.0.240)\\
&= \frac{33}{36}. && (11.0.241)
\end{aligned}
$$

QUESTION

Let $Z \sim \text{Norm}(0,1)$. In this exercise, we find an upper bound for $\mathsf{P}\{|Z| > 3\}$.

Ex 11.179 (1.5)**.** If $X \sim \text{Gamma}(a, \lambda)$ then the rth moment of X is given by $\frac{\Gamma(a+r)}{\lambda^r \Gamma(a)}$. Use this to prove that $\mathsf{E}\left[Z^{2n+2}\right] = (2n+1)\,\mathsf{E}\left[Z^{2n}\right]$ for all positive integers n.

Hint. Use the chi-square distribution.

s.11.179. Let $Y = Z^2$, then $Y \sim \chi_1^2$, so $Y \sim \text{Gamma}(\frac{1}{2}, \frac{1}{2})$. We have

$$\mathsf{E}\left[Z^{2n+2}\right] = \mathsf{E}\left[Y^{n+1}\right] = \frac{\Gamma(n+3/2)}{(1/2)^{n+1}\Gamma(1/2)} = \frac{2^{n+1}\Gamma(n+3/2)}{\Gamma(1/2)}$$
$$\mathsf{E}\left[Z^{2n}\right] = \mathsf{E}\left[Y^{n}\right] = \frac{\Gamma(n+1/2)}{(1/2)^{n}\Gamma(1/2)} = \frac{2^{n}\Gamma(n+1/2)}{\Gamma(1/2)}.$$

Since $\Gamma(n+3/2) = (n+1/2)\Gamma(n+1/2)$ we conclude

$$\mathsf{E}\left[Z^{2n+2}\right] = \frac{2^{n+1}\Gamma(n+3/2)}{\Gamma(1/2)} = 2(n+1/2)\frac{2^{n}\Gamma(n+1/2)}{\Gamma(1/2)} = (2n+1)\,\mathsf{E}\left[Z^{2n}\right].$$

Remarks and grading scheme:

- It is NOT true that $\mathsf{E}\left[Z^{2n+2}\right] = \mathsf{E}\left[Z^{2}\right]\mathsf{E}\left[Z^{2n}\right]$. This would only be true if Z^2 and Z^{2n} would be uncorrelated. But clearly, they are positively correlated: if Z^2 is large, then so is Z^{2n}.
- While induction is a good strategy to try when you have to prove a statement for all $n \in \mathbb{N}$, it is certainly not the only option.
- Grading: 0.5 for introducing $Y = Z^2$ and arguing $Y \sim \text{Gamma}(\frac{1}{2}, \frac{1}{2})$ with some explanation, 0.5 for using the given expression to calculate $\mathsf{E}\left[Z^{2n+2}\right]$ and $\mathsf{E}\left[Z^{2n}\right]$ and 0.5 for using a property of the Gamma function to finish the answer.

Ex 11.180 (0.5)**.** Use the previous exercise to calculate $\mathsf{E}\left[Z^4\right]$.

s.11.180. By using the previous exercise with $n = 1$, we get $\mathsf{E}\left[Z^4\right] = 3\,\mathsf{E}\left[Z^2\right] = 3$.

Remarks and grading scheme:

- If the exercise explicitly asks to use the previous exercise, don't do it in a different way.
- You should really know that $\mathsf{E}\left[Z^2\right] = 1$ for $Z \sim \text{Norm}(0,1)$.
- Grading: 0.5 for a correct solution.

Ex 11.181 (1)**.** We now provide a bound for $\mathsf{P}\{|Z| > 3\}$.

(i) Prove that $\mathsf{P}\{|Z| > 3\} = \mathsf{P}\left\{Z^4 > 81\right\}$.

(ii) Use this to prove that $\mathsf{P}\{|Z| > 3\} \leq \frac{1}{27}$.

s.11.181. We have $\mathsf{P}\{|Z| > 3\} = \mathsf{P}\left\{Z^4 > 81\right\}$ since $|Z| > 3$ if and only if $Z^4 > 81$. The inequality now directly follows from Markov's inequality.

Remarks and grading scheme:

- This exercise was made quite well.
- Grading scheme: 0.5 for (i) and 0.5 for (ii).

Ex 11.182 (2)**.** Prove that $\mathsf{P}\{|Z| > 3\} \leq \frac{\mathsf{E}[Z^{2n}]}{9^n}$ for all $n \in \mathbb{N}$. For what value(s) of n do we obtain the strongest bound for $\mathsf{P}\{|Z| > 3\}$? Also provide this upper bound.

s.11.182. By Markov's inequality,

$$\mathsf{P}\{|Z| > 3\} = \mathsf{P}\left\{Z^{2n} > 9^n\right\} \leq \frac{\mathsf{E}\left[Z^{2n}\right]}{9^n}.$$

From the formula for $\mathsf{E}\left[Z^{2n}\right]$ we see that $\mathsf{E}\left[Z^{2(n+1)}\right] = (2n+1)\,\mathsf{E}\left[Z^{2n}\right]$. We now consider what happens when incrementing n. If $n < 4$ then $2n+1 < 9$, so then incrementing n improves the bound, for $n = 4$ incrementing n doesn't change the bound and for $n > 4$ the bound becomes weaker. So we get the best possible bound for $n = 4$ and $n = 5$. We have $\mathsf{E}\left[Z^8\right] = 7\,\mathsf{E}\left[Z^6\right] = 7 \cdot 5\,\mathsf{E}\left[Z^4\right] = 105$. Hence,

$$\mathsf{P}\{|Z| > 3\} \leq \frac{105}{9^4} \approx 0.016.$$

Remarks and grading scheme:

- Many people tried to use induction in this exercise, but that doesn't work. While it is good to think of induction when you have to prove a statement for all $n \in \mathbb{N}$, it is certainly not the only option.
- In fact, following a similar strategy as in the previous exercise would have helped here.
- Grading scheme: 0.5 for $\mathsf{P}\{|Z| > 3\} \leq \frac{\mathsf{E}[Z^{2n}]}{9^n}$, 1 for showing that the best bound is obtained for $n = 4$ and $n = 5$, and 0.5 for calculating the resulting bound.

QUESTION

A server spends a random amount T on a job. While the server works on the job, it sometimes gets interrupted to do other tasks $\{R_i\}$. Such tasks need to be repaired before the machine can continue working again. It is clear that such interruptions can only occur when the server is working. Assume that the interruptions arrive according to a Poisson process with rate λ, i.e., the number of failures $N(t) \sim \text{Pois}(\lambda t)$. The interruptions $\{R_i\}$ are independent of N and of T and form an iid sequence with common mean $\mathsf{E}[R]$ and variance $\mathsf{V}[R]$. The duration S of a job is its own service time T plus all interruptions, hence $S = T + \sum_{i=1}^{N(T)} R_i$.

Ex 11.183 (1.5)**.** The computation below consists of a number of steps, a, b, ... Explain for each step which property is used to ensure the step is true.

$$\mathsf{E}\left[\sum_{i=1}^{N(t)} R_i \,\middle|\, N(t) = n\right] \stackrel{a}{=} \mathsf{E}\left[\sum_{i=1}^{n} R_i\right] \tag{11.0.242}$$

$$\stackrel{b}{=} n\,\mathsf{E}[R] \tag{11.0.243}$$

$$\mathsf{E}\left[\sum_{i=1}^{N(t)} R_i \,\middle|\, N(t)\right] \stackrel{c}{=} N(t)\,\mathsf{E}[R] \tag{11.0.244}$$

s.11.183. We can use the result of part 1.

a. On the set $\{N(t) = n\}$ $N(t) = n$. Hence we can replate $N(t)$ by n.

b. Linearity of expectation

c. definition of conditional expectation.

Each property missed, e.g, linearity of expectation, minus 0.5.

Ex 11.184 (2)**.** Suppose R is equal to the constant r and $T \sim \text{Exp}(\mu)$, compute $\mathsf{E}[S]$.

s.11.184.

$$\mathsf{E}[S] = \mathsf{E}[T] + \mathsf{E}[N(T)]\,\mathsf{E}[R] = 1/\mu + r\lambda\,\mathsf{E}[T] = 1/\mu + r\lambda/\mu.$$

Typically, 0.5 point for each correct line, but I subtracted points If you say that $\mathsf{E}[N(T)\,\mathsf{E}[R]] = \mathsf{E}[R]$, or write $n\,\mathsf{E}[R]$ as final answer (apparently you did not get the idea that N is an rv.)

Ex 11.185 (1.5)**.** Explain what this code computes.

Python Code

```
1  import numpy as np
2
3  labda = 0.5
4  size = 10
5  num_runs = 50
6
7
8  def do_run():
9      T = np.random.uniform(0, 20)
10     N = np.random.poisson(labda * T)
11     R = np.random.uniform(1, 5, size=N)
12     S = T + R.sum()
13     return S
14
15
16 samples = np.zeros(num_runs)
17 for i in range(num_runs):
18     samples[i] = do_run()
19
20
21 print(samples[samples > 4].var())
```

R Code

```
1  labda <- 0.5
2  size <- 10
3  num_runs <- 50
4
5  do_run <- function() {
6    bigT <- runif(n = 1, min = 0, max = 20)
7    N <- rpois(n = 1, labda * bigT)
8    R <- runif(n = N, min = 1, max = 5)
9    S <- bigT + sum(R)
10   return(S)
11 }
12
13 samples <- rep(0, num_runs)
14 for (i in 1:num_runs) {
15   samples[i] <- do_run()
16 }
17
18 print(var(samples[samples > 4]))
```

Hint, you should know that in P.21 (R18) the string `samples > 4` collects only the samples with value larger than 4.

s.11.185. Here is an EXPLANATION:

1. T is a simulated service time of a job without interruptions.
2. N is is the simulated number of failures that occur during the service of the job
3. R is then a vector of simulated durations of each of the interruptions
4. Hence, S is the total time the job spends at the server
5. We sample the total service time a number of times, and count how often (mean or variance) this exceeds a threshold.

The points earned depends on how clear you explanation is.

Quite a few of you don't seem to understand what it means to *explain* something. Here are some answers that I saw that certainly don't explain what the code does:

1. 'The code does what's stated in the exercise.'. What's the explanation here? The question is also not: do you understood what the code does?
2. 'T is uniform rv, N is a Poisson rv, R is a also a uniform rv. We repeat this a few times.' Like this you just read the code, but I know you can read, so this type of answer is quite useless.
3. 'We collect a subset of the samples and print that.' This answer is just a repition of the hint.

So, next time, when you are asked to explain something, explain it. In your professional carreer, if you read out loud to your customer what s/he can read for him/herself, you'll be fired pretty soon.

QUESTION

We have a population of $X(t)$ individuals at time t. At time t, the time to the next birth is $Z \sim \text{Exp}(\lambda X(t)+\theta)$, and the time to the next death is $Y \sim \text{Exp}(\mu X(t))$; $\lambda, \mu, \theta \geq 0$, and rvs Y and Z are independent. Write $B(h)$ for the number of births during an interval of length h, and $D(h)$ for the number of deaths. (Hint, recall the relation between the Poisson distribution and the exponential distribution.)

In the sequel, take h positive, but very, very small, i.e, $h \ll 1$. With this, we use the shorthand $o(h)$ to capture all terms of a polynomial in h with a power higher than 1, for instance,

$$2h+3h^2+44h^{21} = 2h+o(h). \tag{11.0.245}$$

Like this we can hide all nonlinear terms of a polynomial in the $o(h)$ function. This is easy when we want to take limits, for example,

$$\lim_{h\to 0} \frac{2h+3h^2+44h^{21}}{h} = 2+\lim_{h\to 0}\frac{o(h)}{h} = 2+0. \tag{11.0.246}$$

In other words, when computing this limit for $h \to 0$, we don't care about the details in $o(h)$ because $o(h)/h \to 0$ anyway.

Ex 11.186 (1)**.** Provide intuitive motivation about the correctness of the following equality:

$$\mathsf{P}\{B(h)=1, D(h)=0|X(0)=n\} = (\lambda n+\theta)he^{-(\lambda n+\theta)h}e^{-\mu n h}+o(h). \tag{11.0.247}$$

The $o(h)$ here is a subtlety to get the mathematics correct, but you don't have to explain why this term is necessary.

s.11.186. Since births and deaths are exponentially distributed, we can use that $B(h) \sim \text{Pois}((\lambda X(t)+\theta)h)$ and $D(h) \sim \text{Pois}(\mu X(0)h)$ when $X(0)=n$.

The subtlety is due to the fact that during the time h also multiple arrivals and departures can occur, but since these rates depend on the number people in the system, these rates need not be constant during the time interval h. However, since such events have very small, in fact have $o(h)$ probability, we can capture all such details in the $o(h)$ terms.

Grading: mention the use of exponential and Poisson distribution: +1/2.

Ex 11.187 (1)**.** Use the first degree Taylor's expansion, $f(h) \approx f(0)+hf'(0)+o(h)$, to show that

$$\mathsf{P}\{B(h)=0, D(h)=1|X(0)=n\} = n\mu h+o(h). \tag{11.0.248}$$

s.11.187.

$$\mathsf{P}\{B(h)=0, D(h)=1|X(0)=n\} = e^{-(\lambda n+\theta)h}\mu n h e^{-\mu n h} = (1-(\lambda n+\theta)h)\mu n h(1-\mu n h)+o(h) = \mu n h+o(h). \tag{11.0.249}$$

Grading:

- Skipping the algebra: -1/2.

Ex 11.188 (2). Explain that

$$\mathsf{E}[X(t+h)|X(t)=n] = n + (\lambda n + \theta - \mu n)h + o(h). \tag{11.0.250}$$

s.11.188.

$$\begin{aligned}
\mathsf{E}[X(t+h)|X(t)=n] &= n\,\mathsf{P}\{B(h)=0, D(h)=0\} + (n+1)\,\mathsf{P}\{B(h)=1, D(h)=0\} && (11.0.251)\\
&\quad + (n-1)\,\mathsf{P}\{B(h)=0, D(h)=1\} + o(h) && (11.0.252)\\
&= ne^{-(\lambda n+\theta)h}e^{-\mu n h} + (n+1)(\lambda n+\theta)h + (n-1)\mu n h + o(h) && (11.0.253)\\
&= n(1-(\lambda n+\theta)h)(1-\mu n h) + (n+1)(\lambda n+\theta)h + (n-1)\mu n h + o(h) && (11.0.254)\\
&= n + (\lambda n + \theta - \mu n)h + o(h). && (11.0.255)
\end{aligned}$$

Grading:

- Show also how to simplify the results of the first question. If not, -1/2.

Write $M(t) = \mathsf{E}[X(t)]$.

Ex 11.189 (1). Derive that

$$M(t+h) = M(t) + (\lambda - \mu)M(t)h + \theta h + o(h). \tag{11.0.256}$$

s.11.189. Replace n by $X(t)$ in $\mathsf{E}[X(t+h)|X(t)]$ to see that

$$\mathsf{E}[X(t+h)|X(t)] = X(t) + (\lambda - \mu)X(t)h + \theta h + o(h). \tag{11.0.257}$$

Take expectations left and right and use Adam's law.

Grading:

- No points for not mentioning Adam's law, or showing in some way that you used it.
- Using Adam's law in the wrong way, i.e, not replacing the n by $X(t)$ at most 1/2.

QUESTION

Let X and Y be i.i.d and Unif(1,3) distributed.

Ex 11.190 (0.5)**.** Find the joint PDF $f(x, y)$ of X and Y.

s.11.190. Since X and Y are independent, their joint PDF is the product of their marginal PDFs. This gives:

$$f(x, y) = \left(\frac{1}{(3-1)}\right)\left(\frac{1}{(3-1)}\right) = \frac{1}{4}$$

For $1 < x < 3$ and $1 < y < 3$ and 0 otherwise.
One mistake, zero points.

Ex 11.191 (0.5)**.** Show that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$

s.11.191. Calculating this integral gives:

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dx\, dy =& \int_{1}^{3}\int_{1}^{3} \frac{1}{4}\, dx\, dy \\ =& 1\end{aligned}$$

One mistake, zero points.

Ex 11.192 (3)**.** Find $SD(|X-Y|)$, the standard deviation of the distance between X and Y.

s.11.192. Step 1. Find the expectation $E(|X-Y|)$. Using LOTUS.

$$\begin{aligned}E(|X-Y|) =& \int_{1}^{3}\int_{1}^{3} |x-y| \left(\frac{1}{4}\right) dx\, dy \\ =& \int_{1}^{3}\int_{y}^{3} (x-y) \left(\frac{1}{4}\right) dx\, dy + \int_{1}^{3}\int_{1}^{y} (y-x) \left(\frac{1}{4}\right) dx\, dy \\ =& \frac{1}{3} + \frac{1}{3} \\ =& \frac{2}{3}\end{aligned}$$

Step 2. Find the squared expectation $|X-Y|^2$. Using LOTUS.

$$\begin{aligned}E(|X-Y|^2) =& \int_{1}^{3}\int_{1}^{3} |x-y|^2 \left(\frac{1}{4}\right) dx\, dy \\ =& \int_{1}^{3}\int_{1}^{3} (x-y)^2 \left(\frac{1}{4}\right) dx\, dy \\ =& \int_{1}^{3}\int_{1}^{3} (x^2-2xy+y^2) \left(\frac{1}{4}\right) dx\, dy \\ =& \frac{2}{3}\end{aligned}$$

Step 3. Find the variance of $|X-Y|$.

$$\begin{aligned}\mathsf{V}[|X-Y|] &= E(|X-Y|^2) - E(|X-Y|)^2 \\ &= \frac{2}{3} - \left(\frac{2}{3}\right)^2 \\ &= \frac{2}{9}\end{aligned}$$

Step 4. Find the standard deviation of $|X-Y|$.

$$\begin{aligned}SD(|X-Y|) &= \sqrt{\mathsf{V}[|X-Y|]} \\ &= \sqrt{\frac{2}{9}} \\ &= 0.4714\end{aligned}$$

One point for writing down the integral for $E|X-Y|$ and splitting it up correctly. One point for $E|X-Y|^2$. One point for finding $SD(|X-Y|)$ in the correct way.

Consider the following code:

Python Code

```python
import numpy as np
np.random.seed(3)

num = 100000

x = np.random.normal(loc = 50, scale = 200, size = num)

result1 = np.zeros(num)
for i in range(0,num):
    result1[i] = abs(x[i]-50)<2*200

print(np.sum(result1)/num)
```

Ex 11.193 (0.5)**.** What does the code above do?

s.11.193. It loads the required packages and creates one sample with 100000 observations from a $\mathcal{N}(50, 200)$-distribution. Then for all observations it subtracts its mean and tests if the new value is within 2 standard deviations of the mean.
0.5 points for mentioning the mean is subtracted and it is checked if the value found is smaller than 2 times the s.d.

Ex 11.194 (0.5)**.** The code gives as output 0.95429. Explain why you would expect to get this output from the code. *Hint:* use Theorem 5.4.5 in the book.

s.11.194. By Theorem 5.4.5 we get that $P(|X-\mu| < 2\sigma) \approx 0.95$, this is also shown in the code. *0.5 points for making a comparison between the theorem and the answer in the code. Conclusion should be that they give similar results.*

Let $\{X_k\}$ be a set of iid demands at a shop, distributed as the common rv X, with mean $\mu = \mathsf{E}[X]$ and std $\sigma = \sqrt{\mathsf{V}[X]}$. A random number N of people visit the shop on some day. Let $D_N = \sum_{i=1}^N X_k$.

Ex 11.195 (1)**.** When the rvs X and Y are independent, then $\mathsf{E}[Y|X] = \mathsf{E}[Y]$. It is clear that $\mathsf{E}[Y]$ is a number. Now BH say that $\mathsf{E}[Y|X]$ is a rv. But a rv is not a number. Explain how it can be that $\mathsf{E}[Y|X] = \mathsf{E}[Y]$.

s.11.195.

Ex 11.196. Use the definition of $\mathsf{E}[Y\,|\,A]$ to show that $\mathsf{E}[X\,|\,X=x] = x$.

s.11.196.

We break a stick of length 1 at a uniformly distributed point X. Then we choose the smallest of the two parts. Let the length of this be Y.

Ex 11.197 (1)**.** Find an expression for $\mathsf{E}[Y|X=x]$.

s.11.197.

Ex 11.198 (0.5)**.** Find an expression for $\mathsf{E}[Y|X]$.

s.11.198.

Ex 11.199 (1.5)**.** Compute $\mathsf{V}[Y]$?

s.11.199.

Ex 11.200 (1)**.** What does the code below compute?

s.11.200.

Python Code

```python
import random

random.seed(3)

num = 5
tot = 0
for i in range(num):
    X = random.uniform(0, 1)
    Y = random.uniform(0, 1)
    tot += abs(Y - X)

print(tot / num)
```

R Code

```r
x = 3
```

QUESTION

Ex 11.201 (0.5)**.** Let $\lambda > 0$ be some parameter. Let $X_1, X_2, \dots, X_n \sim \text{Expo}(\lambda)$ be independent. Find the distribution of $\min\{X_1, X_2, \dots, X_n\}$. You can use results from the book here.

s.11.201. By the book, we know that $\min\{X_1, X_2, \dots, X_n\} \sim \text{Expo}(n\lambda)$.

Grading scheme:

- Correct 0.5pt.

Ex 11.202 (2)**.** From here on, consider the random variables $X, Y \sim \text{Expo}(\lambda)$, again for $\lambda > 0$. Assume X, Y are independent. We will in steps show the distribution of $|X - Y|$.
To start, consider the PDF of a random variable W, which is as follows:

$$f_W(w) = \frac{\lambda}{2} e^{-\lambda|w|},$$

for $w \in \mathbf{R}$. Find the moment-generating function of W. As a hint, be careful of what assumptions are necessary to make sure the required integral(s) converge.

s.11.202. We start with the definition:

$$\begin{aligned}
M_W(t) &= \mathsf{E}\left[e^{Wt}\right] \\
&= \int_{-\infty}^{\infty} e^{wt} \frac{\lambda}{2} e^{-\lambda|w|} \,\mathrm{d}w \\
&= \frac{\lambda}{2} \int_{-\infty}^{0} e^{(t+\lambda)w} \,\mathrm{d}w + \frac{\lambda}{2} \int_{0}^{\infty} e^{(t-\lambda)w} \,\mathrm{d}w \\
&= \frac{\lambda}{2} \frac{1}{\lambda + t} + \frac{\lambda}{2} \frac{1}{\lambda - t} \\
&= \frac{\lambda^2}{\lambda^2 - t^2},
\end{aligned}$$

where we used the assumptions $t > -\lambda$ and $t < \lambda$ to make the first and second integral converge respectively. Hence, this MGF is defined only for $|t| < \lambda$.

Grading scheme:

- Definition 0.5pt.
- Split the integral 0.5pt.
- Correct integral calculation 0.5pt.
- Correct bounds 0.5pt.

Ex 11.203 (1)**.** Show that $X - Y \sim W$. You may use any known results from *previous* courses.

s.11.203. We know that $M_{X-Y}(t) = M_X(t)M_Y(-t)$, and that $M_X(t) = M_Y(t) = \frac{\lambda}{\lambda-t}$ for $t < \lambda$. Then, we show that

$$\frac{\lambda}{\lambda - t}\frac{\lambda}{\lambda + t} = \frac{\lambda^2}{\lambda^2 - t^2},$$

for $t < \lambda$ *and* $-t < \lambda$, or $|t| < \lambda$. Since the MGF uniquely determines the distribution, we know that $X - Y \sim W$.

Grading scheme:

- Correct MGF and bounds 1pt.

Ex 11.204 (1.5)**.** Finally, calculate the MGF of the random variable $|X - Y|$. Do you recognize it?

s.11.204. Let $Q = |X - Y| \sim |W|$. We start with the definition:

$$\begin{aligned} M_Q(t) &= \mathsf{E}\left[e^{Qt}\right] \\ &= \mathsf{E}\left[e^{|W|t}\right] \\ &= \int_{-\infty}^{\infty} e^{|w|t}\frac{\lambda}{2}e^{-\lambda|w|}\,\mathrm{d}w \\ &= \frac{\lambda}{2}\int_{-\infty}^{0} e^{(\lambda-t)w}\,\mathrm{d}w + \frac{\lambda}{2}\int_{0}^{\infty} e^{(t-\lambda)w}\,\mathrm{d}w \\ &= \frac{\lambda}{2}\left(\frac{1}{\lambda - t} - \frac{1}{t - \lambda}\right) \\ &= \frac{\lambda}{\lambda - t}, \end{aligned}$$

where we need $t < \lambda$ to make both integrals converge. This is again an exponential MGF!

Grading scheme:

- Correct integration 1pt.
- Correct bounds 0.5pt.

QUESTION

Amy is playing a game. She throws a basketball at a hoop and counts the number of times she successfully throws the ball through the hoop. She keeps counting until she has missed r times, at which moment the current round of the game stops. Her score for the round is the total number of successful throws in the round. Amy plays n rounds in total. We assume that all throws are independent and have the same (unknown) success probability p. Amy is interested in finding out her skill level. That is, she is interested in the value of p.

Given the value of p, Amy's score X_i for the ith round of the game follows a negative binomial distribution with parameters r and p. That is, for every $i = 1, \dots, n$, we have that $X_i | p \sim \mathrm{NB}(r, p)$, with a corresponding pmf defined by

$$\mathsf{P}\{X_i = x_i | p\} = \binom{x_i + r - 1}{x_i} (1-p)^r p^{x_i}, \tag{11.0.258}$$

for $x_i = 0, 1, 2, \dots$. Amy's prior belief about the distribution of p is that it follows a Beta(a, b) distribution with given values for a and b (the exact values of a and b are not relevant for this question).

Ex 11.205 (2.5)**.** In the first round, Amy gets a score of $X_1 = x_1$. Find Amy's *posterior* distribution of p, given this observation.

s.11.205. Using Bayes' rule we have

$$\begin{aligned}
f_1(p | X_1 = x_1) &= \frac{\mathsf{P}\{X_1 = x_1 | p\} f_0(p)}{\mathsf{P}\{X_1 = x_1\}} && (11.0.259)\\
&\propto \mathsf{P}\{X_1 = x_1 | p\} f_0(p) && (11.0.260)\\
&= \binom{x_1 + r - 1}{x_1} (1-p)^r p^{x_1} \frac{1}{B(a,b)} p^{a-1} (1-p)^{b-1} && (11.0.261)\\
&\propto (1-p)^{r+b-1} p^{a+x_1-1}, && (11.0.262)
\end{aligned}$$

in which we recognize the pdf of a Beta$(a + x_1, b + r)$ distribution (up to a constant factor). Hence, the posterior distribution of p given $X_1 = x_1$ is a Beta$(a + x_1, b + r)$ distribution.

Ex 11.206 (1)**.** Is Amy's prior distribution a *conjugate* prior?

s.11.206. Yes. The posterior distribution is in the same family of distributions (Beta) as the prior. Hence, Amy has a conjugate prior.

Ex 11.207 (1.5)**.** Suppose Amy plays n rounds and observes the scores $X_1 = x_1, \dots, X_n = x_n$. What is Amy's posterior distribution after these observations?
Hint: you don't need to do a lot of math here! Instead, use the result of the first question above and a logical argument.

s.11.207. The posterior after observing $X_1 = x_1$ becomes our new prior. Hence, our new prior is a Beta$(a + x_1, b + r)$ distribution. From question 1 it follows that the prior after observing $X_2 = x_2$ then is a Beta$(a + x_1 + x_2, b + 2r)$ distribution. Hence, iterating this process, we find that the posterior distribution of p after observing $X_1 = x_1, \dots, X_n = x_n$ is a Beta$(a + \sum_{i=1}^n x_i, b + rn)$ distribution.

QUESTION

Ex 11.208 (1)**.** Let X follow the student's t distribution with v degrees of freedom. Consider the random variable $Y = \frac{1}{X}$. Find the CDF of Y, $F_Y(y)$, in terms of $\mathsf{P}\left\{X \le \frac{1}{y}\right\}$.

s.11.208. We start by trying to find a formula for $F_Y(y)$. After drawing the function $y = \frac{1}{x}$ in the x, y-plane, it becomes obvious that

$$F_Y(y) = \mathsf{P}\left\{Y \le y\right\} = \begin{cases} \mathsf{P}\{X \le 0\} + \mathsf{P}\left\{X \ge \frac{1}{y}\right\} & \text{if } y > 0 \\ \mathsf{P}\left\{\frac{1}{y} \le X \le 0\right\} & \text{if } y < 0 \end{cases}.$$

Draw it if this is not clear!
To calculate less, we notice that $\mathsf{P}\{X \le 0\} = \frac{1}{2}$, by symmetry. Then,

$$F_Y(y) = \begin{cases} \frac{1}{2} + \mathsf{P}\left\{X \ge \frac{1}{y}\right\} & \text{if } y > 0 \\ \frac{1}{2} & \text{if } y = 0 \\ \frac{1}{2} - \mathsf{P}\left\{X \le \frac{1}{y}\right\} & \text{if } y < 0 \end{cases}.$$

See also exercise 8.9.11.

Grading scheme:

- Correct cases 0.5pt.
- No mistakes etc. 0.5pt.

Ex 11.209 (1)**.** Show that $f_Y(y) = \frac{1}{y^2} f_X(\frac{1}{y})$ for all $y \neq 0$.

s.11.209.

$$\begin{aligned} f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) &= \begin{cases} \frac{\mathrm{d}}{\mathrm{d}y}(\frac{1}{2} + \mathsf{P}\left\{X \ge \frac{1}{y}\right\}) & \text{if } y > 0 \\ \frac{\mathrm{d}}{\mathrm{d}y}(\frac{1}{2} - \mathsf{P}\left\{X \le \frac{1}{y}\right\}) & \text{if } y < 0 \end{cases} \\ &= \begin{cases} -\frac{\mathrm{d}}{\mathrm{d}y} \mathsf{P}\left\{X \le \frac{1}{y}\right\} & \text{if } y > 0 \\ -\frac{\mathrm{d}}{\mathrm{d}y} \mathsf{P}\left\{X \le \frac{1}{y}\right\} & \text{if } y < 0 \end{cases}. \end{aligned}$$

We can write, by definition:

$$\mathsf{P}\left\{X \le \frac{1}{y}\right\} = \int_{-\infty}^{\frac{1}{y}} f_X(s)\,\mathrm{d}s,$$

such that by the FTC,

$$-\frac{\mathrm{d}}{\mathrm{d}y} \mathsf{P}\left\{X \le \frac{1}{y}\right\} = \frac{1}{y^2} f_X(\frac{1}{y})$$

for $y \neq 0$.

Grading scheme:

- Correct calculations 0.5pt.
- No mistakes etc. 0.5pt.
- Alternatively, use the transformation theorem to show this, if you didn't use the result from part (a). Be careful to correctly apply it.

Ex 11.210 (1)**.** Let Y be as in the previous question. What distribution does Y follow when $\nu = 1$?

s.11.210. When $\nu = 1$, X follows a Cauchy distribution. Then, Y must also be Cauchy.

Grading scheme:

- Correct 1pt.

Ex 11.211 (0.5)**.** What happens to the t distribution when $\nu \to \infty$?

s.11.211. It converges to the standard normal distribution.

Grading scheme:

- Correct 0.5pt.

Ex 11.212 (1.5)**.** Let $\nu > 1$. For what value(s) of y is $f_Y(y)$ maximal? You may neglect the possibility that $y = 0$.

s.11.212. For $\nu > 1$, we have that

$$f_Y(y) \propto \left(y + \frac{1}{\nu y}\right)^{\frac{\nu}{2}+\frac{1}{2}}.$$

The FOC tells us that the mode is at

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}y}\left(y + \frac{1}{\nu y}\right)^{\frac{\nu}{2}+\frac{1}{2}} = \left(y + \frac{1}{\nu y}\right)^{\frac{\nu}{2}-\frac{1}{2}}\left(1 - \frac{1}{\nu y^2}\right) = 0 \implies \\
&1 - \frac{1}{\nu y^2} = 0 \implies \\
&y = \pm\frac{\sqrt{\nu}}{\nu}.
\end{aligned}$$

Clearly, these must be the maximum values f_Y takes on; if they were minima the PDF would not integrate to unity, and they cannot be saddle points (the only other option as the PDF is symmetric) since then there would be a different maximum, which the FOC would show. Alternatively, you could look at the second derivative.

Grading scheme:

- Correct FOC 1pt.
- Something about it being a maximum (lenient) 0.5pt.

QUESTION

Denise is the proud owner of a small supermarket. In order to gain some insight into the behavior of her customers, she analyzes their arrival times. In particular, she is interested in the customers' *interarrival times*. Denise knows that the interarrival times Y_i, $i = 1,\dots,n$, are i.i.d. Exponentially distributed with a rate parameter λ (i.e., with a mean value of $1/\lambda$). However, Denise does not know the value of λ. Her prior belief about λ is captured by a Gamma(a,b) distribution, with some particular values of $a,b>0$.

Ex 11.213 (2.5)**.** Denise starts observing the customers' interarrival times. For the first customer she observes $Y_1 = y_1$. What is Denise's *posterior* distribution of λ after this observation?

s.11.213. By Bayes' rule we have

$$
\begin{aligned}
f_1(\lambda|Y_1 = y_1) &= \frac{f_{Y_1|\lambda}(y_1|\lambda) f_0(\lambda)}{f_{Y_1}(y_1)} && (11.0.263)\\
&\propto f_{Y_1|\lambda}(y_1|\lambda) f_0(\lambda) && (11.0.264)\\
&= \frac{b^a}{\Gamma(a)}\lambda^{a-1}e^{-b\lambda}\lambda e^{-\lambda x_1} && (11.0.265)\\
&\propto \lambda^a e^{-(b+x_1)\lambda}, && (11.0.266)
\end{aligned}
$$

in which we recognize the pdf of a Gamma$(a+1, b+y_1)$ distribution (up to a scaling constant). Hence, the posterior distribution λ given $Y_1 = y_1$ is Gamma$(a+1, b+y_1)$.

Applying Bayes' rule: 1 point.
Finding the expression for the posterior: 1 point.
recognizing a $Gamma(a+1, b+x_1)$ dist: 0.5 point.

Ex 11.214 (1.5)**.** After an hour Denise has observed n interarrival times $Y_1 = y_1,\dots,Y_N = y_n$. Without redoing all the math, determine Denise's posterior distribution.

s.11.214. The posterior after observing $Y_1 = y_1$ becomes our new prior. Hence, our new prior is a Gamma$(a+1, b+y_1)$ distribution. From question 1 it follows that the prior after observing $Y_2 = y_2$ then is a Gamma$(a+2, b+y_1+y_2)$ distribution. Iterating this process (i.e., by mathematical induction), we find that the posterior distribution of λ after observing $Y_1 = y_1,\dots,Y_n = y_n$ is a Gamma$(a+n, b+\sum_{i=1}^n y_i)$ distribution.

Noting that the posterior becomes the new prior: 0.5 point.
Correct derivation and answer: 1 point.

Ex 11.215 (1)**.** Does Denise have a *conjugate* prior?

s.11.215. Yes. The posterior distribution is in the same family of distributions (Gamma) as the prior. Hence, John has a conjugate prior.

Correct answer and motivation: 1 point.