Probability Distributions, Exam questions EBP038A05, 2020-2021.2A 2021

dr. N.D. van Foreest c.s. FEB, University of Groningen

June 22, 2021

README

1. Here are the questions and solutions for the exam and the resit.
2. We will not explain the solutions, as the questions are meant to *test* your understanding, not to help you learn the material.
3. Only in case you are sure you found an error, you can let me know, and include the correct answer in LaTeX.

1 Question

Let X and Y be independent and exponentially distributed with rates $\lambda_1 = 1$ and $\lambda_2 = 2$ respectively.

Ex 1.1 (0.5). Find the joint PDF $f(x,y)$ of X and Y.

Ex 1.2 (0.5). Show that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x\,\mathrm{d}y = 1$.

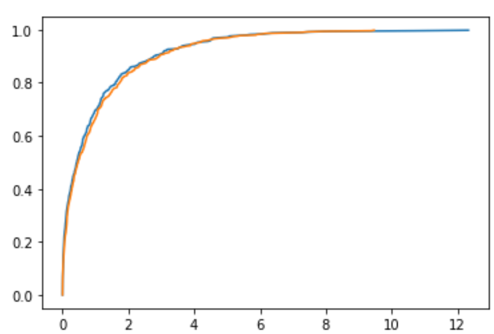
Ex 1.3 (3). Find $\mathsf{E}[|X-Y|]$, i.e., the expected distance between X and Y.

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 1.4 (1). What does the code above do and why would you expect to get the graph below as output?



2 Question

Let X and Y be i.i.d and Unif(1,3) distributed.

Ex 2.1 (0.5)**.** Find the joint PDF $f(x,y)$ of X and Y.

Ex 2.2 (0.5)**.** Show that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$

Ex 2.3 (3)**.** Find $SD(|X-Y|)$, the standard deviation of the distance between X and Y.

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 2.4 (0.5)**.** What does the code above do?

Ex 2.5 (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.

3 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 3.1 (0.5)**.** Find the joint PDF $f(x,y)$ of X and Y.

Ex 3.2 (0.5)**.** Show that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$

Ex 3.3 (3)**.** Find the expectation of R^2, i.e., the expected squared difference from the origin.

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 3.4 (0.5)**.** What does the code above do?

Ex 3.5 (0.5)**.** The code gives as output 1008.99966.
Explain why you would expect to get this output from the code.

4 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 4.1 (0.5)**.** Find the joint PDF $f(x,y)$ of X and Y.

Ex 4.2 (0.5)**.** Show that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$

Ex 4.3 (3)**.** Find the expectation of S, i.e., the squared norm of (X,Y).

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 4.4 (0.5)**.** What does the code above do? *Hint:* unit circle.

Ex 4.5 (0.5)**.** The output of the code is 0.66629. Explain this result.

5 Question

Ex 5.1 (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 $a, b > 0$.

Ex 5.2 (0.5)**.** What distributions do X_1 and X_2 have? Also give the corresponding parameters.

Ex 5.3 (2)**.** Let $B \sim \text{Beta}(p,q)$ for some $p, q > 0$. Show that $1 - B \sim \text{Beta}(q,p)$.

Ex 5.4 (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.

6 Question

Ex 6.1 (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 $S = Y_1 + Y_2$?

Ex 6.2 (1). What is the difference between the PDF of $\sum_{i=1}^{n} X_i$ 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.

Ex 6.3 (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?

7 Question

Ex 7.1 (1.5)**.** Let $X \sim \mathcal{N}(\mu, \mu^2)$ and let $Y = e^X$. Showing your work, find the PDF of Y.

Ex 7.2 (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.

Ex 7.3 (2.5)**.** Find the joint PDF of $U = Y_1 Y_2$ and $V = \frac{Y_1}{Y_2}$.

8 Question

Ex 8.1 (1.5). Let $U \sim \text{Unif}(-1,1)$. Find the PDF of $B = |U|$. What is its distribution? What is $\mathsf{E}[B]$?

Ex 8.2 (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?

Ex 8.3 (1). Let B be as you found it in part (a). Find the CDF of $X = \kappa + \lambda \ln\left(\frac{B}{1-B}\right)$.

Ex 8.4 (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.

9 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 9.1 (0.5)**.** Find $\mathsf{P}\{A_t\}$, the probability that day t hits the *record low*.

Ex 9.2 (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?

Ex 9.3 (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 $s < t$.

Ex 9.4 (2)**.** Let N be the number of *record low* days from day 1 up to t. Find $\mathsf{Cov}[N, I_t]$.

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 $\mathsf{E}[N]$, $\mathsf{E}[I_t]$.

10 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 10.1 (0.5)**.** Find $\mathsf{P}\{A_t\}$, the probability that month t is the *worst in a year*, $t \geq 12$.

Ex 10.2 (2)**.** Let N be the number of *worst in a year* months from the month 12 to month t. Find $\mathsf{E}[N]$ and $\mathsf{Cov}[N, I_t]$.

Ex 10.3 (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?

Ex 10.4 (1)**.** Let B_t be the event that return in month t is lower than all previous months. Find $\mathsf{P}\{B_t \cap B_{t+1}\}$.

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.

11 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 11.1 (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?

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 11.2 (2.5)**.** On average, how many of the remaining books are at least 100 pages longer than the first book you read?

Ex 11.3 (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.

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.

12 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 $i = 1,\ldots,5$.

Ex 12.1 (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?

For the next two exercises, suppose that $(X_1,\ldots,X_5)$ is now Multivariate Normal distributed with $\mathsf{Corr}\left[X_1, X_j\right] = \frac{1}{2}$ for $j = 4,5$

Ex 12.2 (2.5)**.** On average, how many of the remaining poems take at least 1 minutes more to read compared to the 1st poem?

Ex 12.3 (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.

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.

13 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

$$f_{X_i|\lambda}(x|\lambda) = \lambda e^{-\lambda x}, \qquad x > 0. \tag{1}$$

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 13.1 (2.5)**.** Suppose that John observes a first interarrival time of $X_1 = x_1$. Derive John's *posterior* distribution of λ.

Ex 13.2 (1)**.** Is John's prior distribution a *conjugate* prior?

Ex 13.3 (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!

14 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, $i = 1, \ldots, n$, 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 14.1 (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?

Ex 14.2 (1.5)**.** After an hour Denise has observed n interarrival times $Y_1 = y_1, \ldots, Y_N = y_n$. Without redoing all the math, determine Denise's posterior distribution.

Ex 14.3 (1)**.** Does Denise have a *conjugate* prior?

15 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:

$$f_X(x) = \frac{3}{4}x(2-x), \quad 0 \le x \le 2. \tag{2}$$

Let Y denote the size of the payout. Note that $Y = \min\{X, 1\}$.

Ex 15.1 (1)**.** What is the probability that the claim is at most 1000 euros?

Ex 15.2 (1.5)**.** What is the expected payout given the information that the claim is at most 1000 euros?

Ex 15.3 (1.5)**.** What is the (unconditional) expected payout?

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 15.4 (1)**.** Compute the (unconditional) expected value of T.

16 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:

$$f_X(x) = \frac{3}{4}x(2-x), \quad 0 \le x \le 2. \tag{3}$$

Ex 16.1 (1)**.** Compute the probability that the amount of rainfall tomorrow is at most one inch.

Ex 16.2 (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.

Ex 16.3 (1.5)**.** What is the (unconditional) expected amount of rain collected in the shot glass?

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 16.4 (1)**.** What is the (unconditional) expected value of the payout Z?

17 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 17.1 (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.

$$\mathsf{E}[S \mid T = t] \stackrel{a}{=} \mathsf{E}\left[T + \sum_{i=1}^{N(T)} R_i \,\middle|\, T = t\right] \tag{4}$$

$$\stackrel{b}{=} \mathsf{E}[T \mid T = t] + \mathsf{E}\left[\sum_{i=1}^{N(T)} R_i \,\middle|\, T = t\right] \tag{5}$$

$$\stackrel{c}{=} \mathsf{E}[t \mid T = t] + \mathsf{E}\left[\sum_{i=1}^{N(T)} R_i \,\middle|\, T = t\right] \tag{6}$$

$$\stackrel{d}{=} t + \mathsf{E}\left[\sum_{i=1}^{N(t)} R_i\right]. \tag{7}$$

Ex 17.2 (2)**.** Suppose $R \sim \text{Exp}(\mu)$ and $\mathsf{P}\{T = t\} = 1$, compute $\mathsf{E}[S]$.

Ex 17.3 (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
```

```
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(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.

18 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 18.1 (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.

$$\mathsf{E}\left[\sum_{i=1}^{N(t)} R_i \,\middle|\, N(t) = n\right] \stackrel{a}{=} \mathsf{E}\left[\sum_{i=1}^{n} R_i\right] \tag{8}$$

$$\stackrel{b}{=} n\,\mathsf{E}[R] \tag{9}$$

$$\mathsf{E}\left[\sum_{i=1}^{N(t)} R_i \,\middle|\, N(t)\right] \stackrel{c}{=} N(t)\,\mathsf{E}[R] \tag{10}$$

Ex 18.2 (2)**.** Suppose R is equal to the constant r and $T \sim \text{Exp}(\mu)$, compute $\mathsf{E}[S]$.

Ex 18.3 (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].var())
```

R Code

```r
labda <- 0.5
size <- 10
num_runs <- 50

do_run <- function() {
  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.

19 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 19.1 (1.5)**.** In the computation of $\mathsf{V}[S]$ we encounter the following steps.

$$\mathsf{V}\left[\sum_{i=1}^{N(t)} R_i\right] = \mathsf{E}\left[\mathsf{V}\left[\sum_{i=1}^{N(t)} R_i \,\middle|\, N(t)\right]\right] + \mathsf{V}\left[\mathsf{E}\left[\sum_{i=1}^{N(t)} R_i \,\middle|\, N(t)\right]\right]. \tag{11}$$

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.

$$\mathsf{V}\left[\sum_{i=1}^{N(t)} R_i \,\middle|\, N(t) = n\right] \overset{a}{=} \mathsf{V}\left[\sum_{i=1}^{n} R_i\right] \tag{12}$$

$$\overset{b}{=} n\mathsf{V}[R] \tag{13}$$

$$\mathsf{E}\left[\mathsf{V}\left[\sum_{i=1}^{N(t)} R_i \,\middle|\, N(t)\right]\right] \overset{c}{=} \mathsf{E}[N(t)\mathsf{V}[R]] \tag{14}$$

$$\overset{d}{=} \lambda t \mathsf{V}[R]. \tag{15}$$

Ex 19.2 (2)**.** Suppose R is equal to the constant r and $T \sim \text{Unif}([0, a])$, compute $\mathsf{E}[S]$.

Ex 19.3 (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

```
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(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.

20 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 20.1 (1.5)**.** In the computation of $\mathsf{V}[S]$ we encounter the following steps.

$$\mathsf{V}\left[\sum_{i=1}^{N(t)} R_i\right] = \mathsf{E}\left[\mathsf{V}\left[\sum_{i=1}^{N(t)} R_i \middle| N(t)\right]\right] + \mathsf{V}\left[\mathsf{E}\left[\sum_{i=1}^{N(t)} R_i \middle| N(t)\right]\right]. \tag{16}$$

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.

$$\mathsf{V}\left[\mathsf{E}\left[\sum_{i=1}^{N(t)} R_i \middle| N(t)\right]\right] \stackrel{a}{=} \mathsf{V}[N(t)\mathsf{E}[R]] \tag{17}$$

$$\stackrel{b}{=} (\mathsf{E}[R])^2 \mathsf{V}[N(t)] \tag{18}$$

$$\stackrel{c}{=} (\mathsf{E}[R])^2 \lambda t. \tag{19}$$

Ex 20.2 (2)**.** Suppose $\mathsf{P}\{R = r\} = \mathsf{P}\{T = t\} = 1$, compute $\mathsf{E}[S]$.

Ex 20.3 (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
```

```
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(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.

21 Question

Let $Z \sim \text{Norm}(0,1)$. In this exercise, we find an upper bound for $\mathsf{P}\{|Z| > 2\}$.

Ex 21.1 (1.5)**.** Let f be a positive and increasing function, and let X be a r.v.
Consider the following inequality:

$$\mathsf{P}\{X \geq a\} = \mathsf{P}\{f(X) \geq f(a)\} \leq \frac{\mathsf{E}[f(X)]}{f(a)}.$$

(i) Explain why $\mathsf{P}\{X \geq a\} = \mathsf{P}\{f(X) \geq f(a)\}$ holds.

(ii) Explain why $\mathsf{P}\{f(X) \geq f(a)\} \leq \frac{\mathsf{E}[f(X)]}{f(a)}$ holds.

Make sure to clearly indicate where you use that f is positive and increasing.

Ex 21.2 (1)**.** Prove that $\mathsf{P}\{|Z| > 2\} \leq e^{-4t}\,\mathsf{E}\left[e^{tZ^2}\right]$ for $t > 0$.

Ex 21.3 (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 $Y \sim \chi^2_1$, 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.

22 Question

Let $Z \sim \text{Norm}(0,1)$. In this exercise, we find an upper bound for $\mathsf{P}\{|Z| > 3\}$.

Ex 22.1 (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.

Ex 22.2 (0.5)**.** Use the previous exercise to calculate $\mathsf{E}\left[Z^4\right]$.

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 22.3 (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 $\mathsf{P}\{|Z| > 3\} \leq \frac{1}{27}$.

Ex 22.4 (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.

23 Question

Let $Y \sim \text{Norm}(0,1)$. In this exercise, we find an upper bound for $\mathsf{P}\{|Y| > 3\}$.

Ex 23.1 (1.5)**.** Let g be a positive and increasing function, and let Z be a r.v.
Consider the following inequality:

$$\mathsf{P}\{Z \geq a\} = \mathsf{P}\{g(Z) \geq g(a)\} \leq \frac{\mathsf{E}[g(Z)]}{g(a)}.$$

(i) Explain why $\mathsf{P}\{Z \geq a\} = \mathsf{P}\{g(Z) \geq g(a)\}$ holds.

(ii) Explain why $\mathsf{P}\{g(Z) \geq g(a)\} \leq \frac{\mathsf{E}[g(Z)]}{g(a)}$ holds.

Make sure to clearly indicate where you use that g is positive and increasing.

Ex 23.2 (1)**.** Prove that $\mathsf{P}\{|Y| > 3\} \leq e^{-9t}\,\mathsf{E}\left[e^{tY^2}\right]$ for $t > 0$.

Ex 23.3 (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.

24 Question

Let $Y \sim \text{Norm}(0,1)$. In this exercise, we find an upper bound for $\mathsf{P}\{|Y| > 4\}$.

Ex 24.1 (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.

Ex 24.2 (1)**.** Use the previous exercise to calculate $\mathsf{E}\left[Y^4\right]$ and $\mathsf{E}\left[Y^8\right]$.

Ex 24.3 (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 $\mathsf{P}\{|Y| > 4\} \leq \frac{3}{256}$.

Ex 24.4 (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 $\mathsf{P}\{|Y| > 4\}$?

25 Question

Let X be Unif(1,3) distributed and Y be exponentially distributed with rate $\lambda = 2$. X and Y are independent.

Ex 25.1 (1)**.** Find the joint PDF $f(x,y)$ of X and Y.

Ex 25.2 (2)**.** Find $\mathsf{P}\{X \le Y\}$.

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 25.3 (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.

26 Question

Let X and Y be independent and $\mathcal{N}(0,1)$ distributed.

Ex 26.1 (1). Show that $X - Y = \sqrt{2}Z$, where $Z \sim \mathcal{N}(0,1)$.

Ex 26.2 (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 $b > a$.

Ex 26.3 (1). Solve the integral in the previous question. *Hint: use integration by substitution*

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 26.4 (1). What does the code above do?

27 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 27.1 (1)**.** Find the joint PDF $f(x,y)$ of X and Y.

Ex 27.2 (3)**.** Find the expectation of S^2, i.e., the squared norm of (X,Y).

Consider the following code:

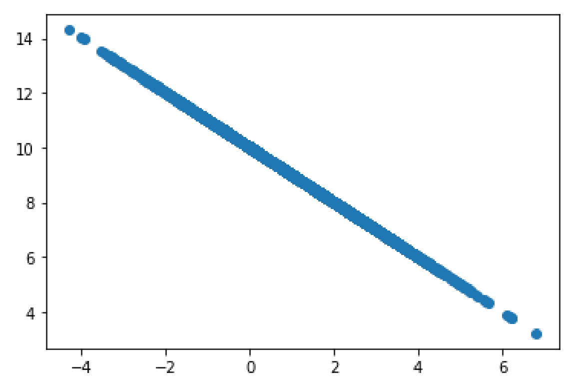
Python Code

```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 27.3 (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.

28 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 28.1 (1)**.** Find the joint PDF $f(x,y)$ of X and Y.

Ex 28.2 (3)**.** Find the value for c such that $\mathsf{E}\left[N^2\right] = 1$.

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 28.3 (1)**.** What will the code above return? You may use the fact that the pdf of a Cauchy random variable is given by

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in (-\infty, \infty).$$

29 Question

Ex 29.1 (1). Let X follow the student's t distribution with ν degrees of freedom. Consider the random variable $Y = \frac{1}{X}$. Find the CDF of Y, $F_Y(y)$, in terms of $\mathsf{P}\left\{X \le \frac{1}{y}\right\}$.

Ex 29.2 (1). Show that $f_Y(y) = \frac{1}{y^2} f_X(\frac{1}{y})$ for all $y \neq 0$.

Ex 29.3 (1). Let Y be as in the previous question. What distribution does Y follow when $\nu = 1$?

Ex 29.4 (0.5). What happens to the t distribution when $\nu \to \infty$?

Ex 29.5 (1.5). Let $\nu > 1$. For what value(s) of y is $f_Y(y)$ maximal? You may neglect the possibility that $y = 0$.

30 Question

Ex 30.1 (0.5). Let $\lambda > 0$ be some parameter. Let $X_1, X_2, \ldots, X_n \sim \text{Expo}(\lambda)$ be independent. Find the distribution of $\min\{X_1, X_2, \ldots, X_n\}$. You can use results from the book here.

Ex 30.2 (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.

Ex 30.3 (1). Show that $X - Y \sim W$. You may use any known results from *previous* courses.

Ex 30.4 (1.5). Finally, calculate the MGF of the random variable $|X - Y|$. Do you recognize it?

31 Question

Ex 31.1 (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.

Ex 31.2 (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.

Ex 31.3 (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?

Ex 31.4 (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 $V = X_1 - X_2$.

32 Question

Ex 32.1 (1)**.** Let $U \sim \mathrm{Unif}\{-n, -n+1, \ldots, n-1, n\}$ for some $n \in \mathbb{N}$. Find the PMF of $B = |U|$. What is $\mathsf{E}[B]$?

Ex 32.2 (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.

Ex 32.3 (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 $|Z|$?

33 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 33.1 (1)**.** Show that $2N_1 + N_2$ and $2N_1 - N_2$ are **not** independent of each other.

Ex 33.2 (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?

Ex 33.3 (2)**.** Calculate $\rho_{X,N}$, the correlation between X and N.

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

```python
import random
import numpy as np
random.seed(444)
A = np.diag([1,1,1])
B = np.zeros(3)
X = np.random.multivariate_normal(B, A, size = 100)
output = np.cov(X, rowvar = False)[0:2,0:2]
```

Ex 33.4 (1)**.** Explain in detail the purpose of each line of the above codes.

34 Question

There are three cheese shops in town, shop A, shop B and shop C. N_1 customers enter shop A per hour, $N_1 \sim \text{Poisson}(\lambda_1)$; N_2 customers enter shop B per hour, $N_2 \sim \text{Poisson}(\lambda_2)$; N_3 customers enter shop C per hour, $N_3 \sim \text{Poisson}(\lambda_3)$. 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 34.1 (1)**.** Show that $N_1 + N_2$ and $N_2 - 2N_3$ are **not** independent of each other.

Ex 34.2 (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?

Ex 34.3 (2)**.** Calculate $\rho_{X,N}$, the correlation between X and N.

Consider the following codes:

R Code

```r
library(mvtnorm)
set.seed(777)
A<-matrix(c(1,2,2,4),nrow = 2)
B<-c(1,2)
X<-rmvnorm(50,mean=B,sigma=A)
output=cor(X[3,],X[41,])
```

Python Code

```python
import random
import numpy as np
random.seed(777)
A = [[1,2],[2,4]]
B = [1,2]
X = np.random.multivariate_normal(B, A, size = 50)
output = np.corrcoef(X[[3,41],:].reshape(4))
```

Ex 34.4 (1)**.** Explain in detail the purpose of **each line** of the above codes.

35 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 35.1 (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 $\rho_{X_A,X}$.

Ex 35.2 (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 $\text{Cov}[X_A, X_C]$ and ρ_{X_A,X_C}.

Consider the following codes:

R Code

```
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

```
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 35.3 (1)**.** Explain in detail the purpose of **Line 1,2, 6,7,8** of the above codes.

36 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 36.1 (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 $\rho_{X_B,N}$.

Ex 36.2 (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}.

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 36.3 (1)**.** Explain in detail the purpose of **Line 1, 2, 5, 6, 7, 8** of the above codes.

37 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 $i = 1, \ldots, n$, we have that $X_i | p \sim \mathrm{NB}(r, p)$, with a corresponding pmf defined by

$$\mathsf{P}\{X_i = x_i | p\} = \binom{x_i + r - 1}{x_i} (1-p)^r p^{x_i}, \tag{20}$$

for $x_i = 0, 1, 2, \ldots$. 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 37.1 (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.

Ex 37.2 (1)**.** Is Amy's prior distribution a *conjugate* prior?

Ex 37.3 (1.5)**.** Suppose Amy plays n rounds and observes the scores $X_1 = x_1, \ldots, X_n = x_n$. 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.

38 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 $Y_i | p \sim \mathrm{NB}(r, p)$, with a corresponding pmf defined by

$$\mathsf{P}\{Y_i = y_i | p\} = \binom{y_i + r - 1}{y_i} (1-p)^r p^{y_i}, \tag{21}$$

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 38.1 (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?

Ex 38.2 (1)**.** Is John's prior distribution a *conjugate* prior?

Ex 38.3 (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.

39 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 39.1 (1)**.** Determine the joint distribution of X and Y conditional on Amy winning.

Ex 39.2 (1.5)**.** Find Amy's expected score conditional on Amy winning.

Ex 39.3 (1.5)**.** Find Amy's (unconditional) expected score.

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 39.4 (1)**.** What is the (unconditional) expected reward for Amy? That is, compute $\mathsf{E}[T]$.

40 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 40.1 (1)**.** Find the joint distribution of X and Y conditional on Catherine winning.

Ex 40.2 (1.5)**.** What is Catherine's expected score conditional on Catherine winning?

Ex 40.3 (1.5)**.** Determine Catherine's (unconditional) expected score.

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 40.4 (1)**.** What is the (unconditional) expected reward for Catherine? That is, compute $\mathsf{E}[S]$.

41 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,

$$2h + 3h^2 + 44h^{21} = 2h + o(h). \tag{22}$$

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,

$$\lim_{h\to 0} \frac{2h + 3h^2 + 44h^{21}}{h} = 2 + \lim_{h\to 0} \frac{o(h)}{h} = 2 + 0. \tag{23}$$

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 41.1 (1)**.** Explain that

$$\mathsf{P}\{B(h) = 1, D(h) = 0 | L(0) = n\} = \lambda h e^{-\lambda h} e^{-\mu n h}. \tag{24}$$

Ex 41.2 (1)**.** Use the first degree Taylor's expansion, $f(h) \approx f(0) + hf'(0) + o(h)$, to motivate that

$$\mathsf{P}\{B(h) = 0, D(h) = 1 | L(0) = n\} = n\mu h + o(h). \tag{25}$$

Ex 41.3 (2)**.** Explain that

$$\mathsf{E}[L(t+h) | L(t) = n] = n + (\lambda - \mu n)h + o(h). \tag{26}$$

Write $M(t) = \mathsf{E}[L(t)]$.

Ex 41.4 (1)**.** Derive that

$$M(t+h) = M(t) + (\lambda - \mu M(t))h + o(h). \tag{27}$$

42 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,

$$2h + 3h^2 + 44h^{21} = 2h + o(h). \tag{28}$$

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,

$$\lim_{h\to 0} \frac{2h + 3h^2 + 44h^{21}}{h} = 2 + \lim_{h\to 0} \frac{o(h)}{h} = 2 + 0. \tag{29}$$

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 42.1 (1)**.** Provide intuitive motivation about the correctness of the following equality:

$$\mathsf{P}\{B(h) = 1, D(h) = 0 | X(0) = n\} = (\lambda n + \theta) h e^{-(\lambda n+\theta)h} e^{-\mu n h} + o(h). \tag{30}$$

The $o(h)$ here is a subtlety to get the mathematics correct, but you don't have to explain why this term is necessary.

Ex 42.2 (1)**.** Use the first degree Taylor's expansion, $f(h) \approx f(0) + hf'(0) + o(h)$, to show that

$$\mathsf{P}\{B(h) = 0, D(h) = 1 | X(0) = n\} = n\mu h + o(h). \tag{31}$$

Ex 42.3 (2)**.** Explain that

$$\mathsf{E}[X(t+h) | X(t) = n] = n + (\lambda n + \theta - \mu n)h + o(h). \tag{32}$$

Write $M(t) = \mathsf{E}[X(t)]$.

Ex 42.4 (1)**.** Derive that

$$M(t+h) = M(t) + (\lambda - \mu)M(t)h + \theta h + o(h). \tag{33}$$

43 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 43.1 (1). Explain that $\lambda/(\lambda+\mu)$ is the probability an arrival occurs before a departure.

For the moment, assume that $\mathsf{E}[T] < \infty$.

Ex 43.2 (1). Explain that for $n > 0$:

$$\mathsf{E}[T|L(0) = n] = \mathsf{E}[T|L(0) = n+1]\frac{\lambda}{\lambda+\mu} + \mathsf{E}[T|L(0) = n-1]\frac{\mu}{\lambda+\mu} + \frac{1}{\lambda+\mu}. \tag{34}$$

Ex 43.3 (1). Show that $\mathsf{E}[T|L(0) = n] = n/(\mu-\lambda)$.

Define $\rho = \lambda/\mu$. Assume that $L(0) \sim \text{Geo}(1-\rho)$.

Ex 43.4 (1). Find a simple expression for $\mathsf{E}[T]$.

Ex 43.5 (1). Up to now we simply assumed that $\mathsf{E}[T] < \infty$. Motivate intuitively that the condition $\lambda < \mu$ ensures that $\mathsf{E}[T] < \infty$.

44 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 44.1 (1)**.** Explain that $\mathsf{E}[T|X = B] = 3 + \mathsf{E}[T]$.

Ex 44.2 (1)**.** Compute $\mathsf{E}[T]$.

Ex 44.3 (2)**.** Compute $\mathsf{V}[T]$.

Ex 44.4 (1)**.** Why was it actually allowed to assume that $\mathsf{E}[T] < \infty$?

45 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 45.1 (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?

Recall that each element of Y is the mean of k i.i.d. Exp(λ) r.v.s.

Ex 45.2 (1)**.** What is the exact distribution of an element of Y?
Give its name and its parameters, and explain the answer.

Let $(Y_1,\ldots,Y_\ell)$ be the elements of `Y` and let $(Z_1,\ldots,Z_\ell)$ 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 45.3 (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?

Ex 45.4 (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).

Ex 45.5 (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).

46 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 46.1 (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`?

Recall that each element of `Y` is the mean of k i.i.d. Exp(λ) r.v.s.

Ex 46.2 (1)**.** What is the exact distribution of an element of `Y`?
Give its name and its parameters, and explain the answer.

Let $(Y_1, \ldots, Y_\ell)$ be the elements of `Y` and let $(Z_1, \ldots, Z_\ell)$ 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 46.3 (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?

Ex 46.4 (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).

Ex 46.5 (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).

47 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 47.1 (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. $\text{Exp}(\lambda)$ r.v.s. What are k and λ?

(iii) What are the (theoretical) expectation and variance of an element of Y?

Recall that each element of Y is the mean of k i.i.d. $\text{Exp}(\lambda)$ r.v.s.

Ex 47.2 (1)**.** What is the exact distribution of an element of Y?
Give its name and its parameters, and explain the answer.

Let $(Y_1, \ldots, Y_\ell)$ be the elements of `Y` and let $(Z_1, \ldots, Z_\ell)$ be the elements of `Z`. Recall that each Z_i depends on k because Y_i is the mean of k i.i.d. $\text{Exp}(\lambda)$ r.v.s. Let T be the random variable to which Z_1 converges in the limit $k \to \infty$.

Ex 47.3 (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?

Ex 47.4 (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).

Ex 47.5 (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).

48 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 48.1 (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?

Recall that each element of Y is the mean of k i.i.d. Exp(λ) r.v.s.

Ex 48.2 (1). What is the exact distribution of an element of Y?
Give its name and its parameters, and explain the answer.

Let $(Y_1, \ldots, Y_\ell)$ be the elements of `Y` and let $(Z_1, \ldots, Z_\ell)$ be the elements of `Z`. Recall that each Z_i depends on k because Y_i is the mean of k i.i.d. $\mathrm{Exp}(\lambda)$ r.v.s. Let T be the random variable to which Z_1 converges in the limit $k \to \infty$.

Ex 48.3 (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?

Ex 48.4 (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).

Ex 48.5 (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).

49 Solutions

s.1.1. 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

s.1.2. 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

s.1.3. Similar to example 7.2.2., we get that by 2D LOTUS:

$$\begin{aligned}\mathsf{E}[|X-Y|] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |x-y| f(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_0^{\infty}\int_y^{\infty} (x-y)(2e^{-x-2y})\,\mathrm{d}x\,\mathrm{d}y + \int_0^{\infty}\int_0^{y} (y-x)(2e^{-x-2y})\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_0^{\infty}\left([(x-y)\cdot -2e^{-x-2y}]_y^{\infty} - \int_y^{\infty} -2e^{-x-2y}\,\mathrm{d}x\right)\mathrm{d}y \\ &\quad + \int_0^{\infty}\left([(y-x)\cdot -2e^{-x-2y}]_0^{y} - \int_0^{y} 2e^{-x-2y}\,\mathrm{d}x\right)\mathrm{d}y \\ &= \int_0^{\infty} [-2e^{-x-2y}]_y^{\infty}\,\mathrm{d}y + \int_0^{\infty} 2ye^{-2y} - [-2e^{-x-2y}]_0^{y}\,\mathrm{d}y \\ &= \int_0^{\infty} 2e^{-3y}\,\mathrm{d}y + \int_0^{\infty} 2ye^{-2y} + 2e^{-3y} - 2e^{-2y}\,\mathrm{d}y \\ &= \left[-\tfrac{2}{3}e^{-3y}\right]_0^{\infty} + \left[-ye^{-2y} + \tfrac{1}{2}e^{-2y} - \tfrac{2}{3}e^{-3y} + e^{-2y}\right]_0^{\infty} \\ &= \tfrac{2}{3} + \tfrac{1}{2} + \tfrac{2}{3} - 1 = \tfrac{5}{6}.\end{aligned}$$

One point for writing down the integral correctly using LOTUS and splitting it up correctly. Two points for the computations.

s.1.4. 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 $Z^2 \sim \chi_1^2$.
0.5 points for mentioning data is standardized, 0.5 points for mentioning a squared standard normal r.v. is chi-square.

s.2.1. 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.

s.2.2. 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.

s.2.3. 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 $|X-Y|$.

$$\begin{aligned}\mathbb{V}[|X-Y|] &= E(|X-Y|^2) - E(|X-Y|)^2\\ &= \frac{2}{3}-\left(\frac{2}{3}\right)^2\\ &= \frac{2}{9}\end{aligned}$$

Step 4. Find the standard deviation of $|X-Y|$.

$$\begin{aligned}SD(|X-Y|) &= \sqrt{\mathbb{V}[|X-Y|]}\\ &= \sqrt{\frac{2}{9}}\\ &= 0.4714\end{aligned}$$

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.

s.2.4. 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.

s.2.5. 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.

s.3.1. So we can see that both X and Y are uniformly distributed on $(-\sqrt{\pi},\sqrt{\pi})$. Then their joint PDF is simply:

$$
\begin{aligned}
f(x,y) &= \Big(\frac{1}{\sqrt{\pi}-(-\sqrt{\pi}))}\Big)\Big(\frac{1}{(\sqrt{\pi}-(-\sqrt{\pi}))}\Big) \\
&= \frac{1}{4\pi}
\end{aligned}
$$

One mistake, zero points.

s.3.2.

$$
\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy &= \int_{-\sqrt{\pi}}^{\sqrt{\pi}}\int_{-\sqrt{\pi}}^{\sqrt{\pi}} \frac{1}{4\pi}\,dx\,dy \\
&= 1
\end{aligned}
$$

One mistake, zero points

s.3.3. Note that $R=\sqrt{X^2+X^2}$, so then $R^2=X^2+Y^2$. Using LOTUS:

$$
\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Big(x^2+y^2\Big)f(x,y)\,dx\,dy &= \int_{-\sqrt{\pi}}^{\sqrt{\pi}}\int_{-\sqrt{\pi}}^{\sqrt{\pi}}\Big(x^2+y^2\Big)\Big(\frac{1}{4\pi}\Big)\,dx\,dy \\
&= \int_{-\sqrt{\pi}}^{\sqrt{\pi}}\int_{-\sqrt{\pi}}^{\sqrt{\pi}}\frac{x^2+y^2}{4\pi}\,dx\,dy \\
&= \frac{1}{4\pi}\int_{-\sqrt{\pi}}^{\sqrt{\pi}}\int_{-\sqrt{\pi}}^{\sqrt{\pi}} x^2+y^2\,dx\,dy \\
&= \frac{1}{4\pi}\int_{-\sqrt{\pi}}^{\sqrt{\pi}}\Big[\frac{x^3}{3}+y^2x\Big]_{-\sqrt{\pi}}^{\sqrt{\pi}}\,dy \\
&= \frac{1}{4\pi}\int_{-\sqrt{\pi}}^{\sqrt{\pi}}\Big(\frac{\pi^{\frac{3}{2}}}{3}+\sqrt{\pi}y^2-\frac{(-\pi)^{\frac{3}{2}}}{3}+\sqrt{\pi}y^2\Big)\,dy \\
&= \frac{1}{4\pi}\int_{-\sqrt{\pi}}^{\sqrt{\pi}}\Big(\frac{\pi^{\frac{3}{2}}}{3}+\sqrt{\pi}y^2+\frac{\pi^{\frac{3}{2}}}{3}+\sqrt{\pi}y^2\Big)\,dy \\
&= \frac{1}{4\pi}\int_{-\sqrt{\pi}}^{\sqrt{\pi}}\frac{2\pi^{\frac{3}{2}}}{3}+2\sqrt{\pi}y^2\,dy \\
&= \frac{1}{4\pi}\Big[\frac{2\pi^{\frac{3}{2}}}{3}y+\frac{2\sqrt{\pi}y^3}{3}\Big]_{-\sqrt{\pi}}^{\sqrt{\pi}} \\
&= \frac{1}{4\pi}\Big(\frac{2\pi^2}{3}+\frac{2\pi^2}{3}+\frac{2(-\pi)^2}{3}+\frac{2(-\pi)^2}{3}\Big) \\
&= \frac{1}{4\pi}\frac{8\pi^2}{3}=\frac{2\pi}{3}
\end{aligned}
$$

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.

s.3.4. 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.

s.3.5. 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 $E(XY) = E(X)E(Y)$

s.4.1. So we can see that both X and Y are uniformly distributed on $(-\sqrt{7}, \sqrt{7})$. Then their joint PDF is simply:

$$\begin{aligned}
f(x,y) &= \Big(\frac{1}{\sqrt{7}-(-\sqrt{7}))}\Big)\Big(\frac{1}{(\sqrt{7}-(-\sqrt{7}))}\Big) \\
&= \frac{1}{28}
\end{aligned}$$

One mistake, zero points.

s.4.2.

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy &= \int_{-\sqrt{7}}^{\sqrt{7}}\int_{-\sqrt{7}}^{\sqrt{7}} \frac{1}{28}\,dx\,dy \\
&= 1
\end{aligned}$$

One mistake, zero points.

s.4.3. Note that $S = X^2 + Y^2$. Using LOTUS:

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(x^2+y^2\right) f(x,y)\,dx\,dy &= \int_{-\sqrt{7}}^{\sqrt{7}}\int_{-\sqrt{7}}^{\sqrt{7}} \left(x^2+y^2\right)\Big(\frac{1}{28}\Big)\,dx\,dy \\
&= \int_{-\sqrt{7}}^{\sqrt{7}}\int_{-\sqrt{7}}^{\sqrt{7}} \frac{x^2+y^2}{28}\,dx\,dy \\
&= \frac{1}{28}\int_{-\sqrt{7}}^{\sqrt{7}}\int_{-\sqrt{7}}^{\sqrt{7}} x^2+y^2\,dx\,dy \\
&= \frac{1}{28}\int_{-\sqrt{7}}^{\sqrt{7}} \Big[\frac{x^3}{3}+y^2x\Big]_{-\sqrt{7}}^{\sqrt{7}}\,dy \\
&= \frac{1}{28}\int_{-\sqrt{7}}^{\sqrt{7}} \Big(\frac{7^{\frac{3}{2}}}{3}+\sqrt{7}y^2-\frac{(-7)^{\frac{3}{2}}}{3}+\sqrt{7}y^2\Big)\,dy \\
&= \frac{1}{28}\int_{-\sqrt{7}}^{\sqrt{7}} \Big(\frac{7^{\frac{3}{2}}}{3}+\sqrt{7}y^2+\frac{7^{\frac{3}{2}}}{3}+\sqrt{7}y^2\Big)\,dy \\
&= \frac{1}{28}\int_{-\sqrt{7}}^{\sqrt{7}} \frac{2}{3}7^{\frac{3}{2}}+2\sqrt{7}y^2\,dy \\
&= \frac{1}{28}\Big[\frac{2}{3}7^{\frac{3}{2}}y+\frac{2}{3}\sqrt{7}y^3\Big]_{-\sqrt{7}}^{\sqrt{7}} \\
&= \frac{1}{28}\Big(\frac{2}{3}7^2+\frac{2}{3}7^2+\frac{2}{3}(-7)^2+\frac{2}{3}(-7)^2\Big) \\
&= \frac{1}{28}\frac{8}{3}7^2 = \frac{14}{3} = 4\frac{2}{3}
\end{aligned}$$

One point for finding $S^2 = X^2 + Y^2$ and writing down the integral correctly using LOTUS. Two points for the remaining calculations.

s.4.4. 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.

s.4.5. The mean distance from a randomly selected point in the unit circle to the origin is $\approx \frac{2}{3}$.
Trivial

s.5.1. For $y \in (0,1)$:

$$F_{X_1}(y) = \mathsf{P}\{X_1 \le y\} = \mathsf{P}\left\{(U_1)^{1/a} \le y\right\} = \mathsf{P}\left\{U_1 \le y^a\right\} = F_{U_1}(y^a).$$

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.

s.5.2. $X_1 \sim \text{Beta}(a,1)$ and $X_2 \sim \text{Beta}(b,1)$.

Grading scheme:

- Correct 0.5pt.

s.5.3. For $y \in (0,1)$ we have that

$$F_{1-B}(y) = \mathsf{P}\{1-B \le y\} = \mathsf{P}\{B \ge 1-y\} = 1 - \mathsf{P}\{B \le 1-y\} = 1 - F_B(1-y).$$

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.

s.5.4. 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.

s.6.1. We first find the distribution of Y_1.

$$F_{Y_1}(y) = \mathsf{P}\{Y_1 \le y\} = \mathsf{P}\{2\lambda X_1 \le y\} = \mathsf{P}\left\{X_1 \le \frac{y}{2\lambda}\right\} = F_{X_1}\left(\frac{y}{2\lambda}\right).$$

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.

s.6.2. B.H. show that $\sum_{i=1}^{n} X_i \sim \text{Gamma}(n, \lambda)$, 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)

s.6.3. 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 $\text{Gamma}\left(n+2, \frac{1}{2}\right)$ 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).

s.7.1. We start with finding the CDF of Y.

$$F_Y(y) = \mathsf{P}\{Y \le y\} = \mathsf{P}\left\{e^X \le y\right\} = \mathsf{P}\{X \le \ln y\} = F_X(\ln y).$$

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.

s.7.2. They are independent. The book proves that $X_1 + X_2$ and $X_1 - X_2$ are independent. Then it must be that $e^{X_1+X_2} = Y_1Y_2$ 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)

s.7.3. 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.

s.8.1. 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

$$F_B(y) = \mathsf{P}\{B \le y\} = \mathsf{P}\{|X| \le y\} = \mathsf{P}\{-y \le X \le y\} = F_X(y) - F_X(-y)$$

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.

s.8.2. 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.

s.8.3. 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

$$\begin{aligned}
F_X(y) &= \mathsf{P}\{X \le y\} \\
&= \mathsf{P}\left\{\kappa + \lambda \ln\left(\frac{B}{1-B}\right) \le y\right\} \\
&= \mathsf{P}\left\{\ln\left(\frac{B}{1-B}\right) \le \frac{y-\kappa}{\lambda}\right\} \\
&= \mathsf{P}\left\{B \le \frac{\exp\frac{y-\kappa}{\lambda}}{1+\exp\frac{y-\kappa}{\lambda}}\right\} \\
&= F_B\left(\frac{\exp\frac{y-\kappa}{\lambda}}{1+\exp\frac{y-\kappa}{\lambda}}\right) \\
&= \frac{\exp\frac{y-\kappa}{\lambda}}{1+\exp\frac{y-\kappa}{\lambda}}
\end{aligned}$$

for $y \in \mathbf{R}$.

Grading scheme:

- CDF technique derivation 0.5pt.
- No mistakes 0.5pt.

s.8.4. 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.

s.9.1. Since all of the first t days are equally likely to have the lowest return, by symmetry, $\mathsf{P}\{A_t\} = \frac{1}{t}$.

s.9.2. 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, $\mathsf{P}\{A_t \cap A_{t+1}\} = \frac{(t-1)!}{(t+1)!} = \frac{1}{t(t+1)}$.

s.9.3. 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:

$$
\begin{aligned}
\mathsf{P}\{A_s \cap A_t\} &= \frac{\binom{t-1}{t-s-1}(t-s-1)!(s-1)!}{t!} \\
&= \frac{(s-1)!(t-1)!}{s!t!} \\
&= \frac{1}{st} \\
&= \mathsf{P}\{A_s\}\,\mathsf{P}\{A_t\}.
\end{aligned}
$$

s.9.4. 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 $N = \sum_{k=1}^{t} I_k$,

$$
\begin{aligned}
\mathsf{E}[N] &= \mathsf{E}\left[\sum_{k=1}^{t} I_k\right] \\
&= \sum_{k=1}^{t} \mathsf{E}[I_k] \\
&= \sum_{k=1}^{t} \frac{1}{k}
\end{aligned}
$$

$$
\begin{aligned}
\mathsf{E}[NI_t] &= \mathsf{E}[\mathsf{E}[NI_t \mid I_t]] \\
&= \mathsf{E}\left[\sum_{k=1}^{t} I_k I_t \,\middle|\, I_t = 1\right] \mathsf{P}\{I_t = 1\} \\
&= \mathsf{E}\left[\sum_{k=1}^{t-1} I_k + 1\right] \mathsf{P}\{I_t = 1\} \\
&= \frac{1}{t}\sum_{k=1}^{t-1}\frac{1}{k} + \frac{1}{t}
\end{aligned}
$$

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,

$$
\begin{aligned}
\mathsf{Cov}[N,I_t] &= \mathsf{Cov}\left[\sum_{k=1}^{t} I_k, I_t\right] \\
&= \sum_{k=1}^{t} \mathsf{Cov}[I_k, I_t] \\
&= \mathsf{Cov}[I_t, I_t] = \mathsf{V}[I_t] = \mathsf{E}\left[I_t^2\right] - \mathsf{E}[I_t]^2 \\
&= \frac{1}{t} - \frac{1}{t^2} = \frac{t-1}{t^2}.
\end{aligned}
$$

s.10.1. Since the current month and all previous 11 months are equally likely to have the lowest return, by symmetry, $\mathsf{P}\{A_t\} = \frac{1}{12}$.

s.10.2. 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:

$$
\begin{aligned}
\mathsf{Cov}[N,I_t] &= \mathsf{E}[NI_t] - \mathsf{E}[N]\,\mathsf{E}[I_t] \\
&= \frac{t}{144} - \frac{t-11}{12}\frac{1}{12} = \frac{11}{144}.
\end{aligned}
$$

s.10.3. 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.

s.10.4. 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, $\mathsf{P}\{B_t \cap B_{t+1}\} = \frac{(t-1)!}{(t+1)!} = \frac{1}{t(t+1)}$.

s.11.1. 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 $\mathsf{E}\left[\Sigma_{i=3}^{6} I_i\right] = \frac{1}{3} \cdot 4 = \frac{4}{3}$.

s.11.2. In order to answer this question, we want to know $P(X_i - X_1 > 1)$, for $i = 3,...,6$. 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 $0.16\cdot 4 = 0.64$.

s.11.3. $\mathsf{Cov}[X_1 - cX_3, X_3] = \mathsf{Cov}[X_1 - cX_3, X_3] = \mathsf{Cov}[X_1, X_3] - c\,\mathsf{V}[X_3] = \frac{1}{2} - c$, so for $c = \frac{1}{2}$, we have that $X_1 - cX_3$ and X_3 are uncorrelated. Since $(X_1,...,X_6)$ has the multivariate normal distribution, it follows that $X_1 - cX_3$ and X_3 are independent.

s.12.1. 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 $\mathsf{E}\left[\Sigma_{i=4}^{5} I_i\right] = \frac{1}{4}\cdot 2 = \frac{1}{2}$.

s.12.2. 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 $0.16\cdot 2 = 0.32$.

s.12.3. $\mathsf{Cov}[X_1 - cX_4, X_4] = \mathsf{Cov}[X_1 - cX_4, X_4] = \mathsf{Cov}[X_1, X_4] - c\,\mathsf{V}[X_4] = \frac{1}{2} - c$, so for $c = \frac{1}{2}$, we have that $X_1 - cX_4$ and X_4 are uncorrelated. Since $(X_1,...,X_5)$ has the multivariate normal distribution, it follows that $X_1 - cX_4$ and X_4 are independent.

s.13.1. By Bayes' rule we have

$$\begin{aligned}f_1(\lambda|X_1 = x_1) &= \frac{f_{X_1|\lambda}(x_1|\lambda) f_0(\lambda)}{f_{X_1}(x_1)} && (35)\\ &\propto f_{X_1|\lambda}(x_1|\lambda) f_0(\lambda) && (36)\\ &= \frac{b^a}{\Gamma(a)}\lambda^{a-1}e^{-b\lambda}\lambda e^{-\lambda x_1} && (37)\\ &\propto \lambda^a e^{-(b+x_1)\lambda}, && (38)\end{aligned}$$

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.

s.13.2. 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.

s.13.3. 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.

s.14.1. By Bayes' rule we have

$$\begin{aligned} f_1(\lambda|Y_1 = y_1) &= \frac{f_{Y_1|\lambda}(y_1|\lambda) f_0(\lambda)}{f_{Y_1}(y_1)} && (39)\\ &\propto f_{Y_1|\lambda}(y_1|\lambda) f_0(\lambda) && (40)\\ &= \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda} \lambda e^{-\lambda x_1} && (41)\\ &\propto \lambda^a e^{-(b+x_1)\lambda}, && (42) \end{aligned}$$

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.

s.14.2. 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 $Y_1 = y_1, \ldots, Y_n = y_n$ 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.

s.14.3. 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.

s.15.1. We have

$$\begin{aligned} \mathsf{P}\{X \leq 1\} &= \int_0^1 \frac{3}{4} x(2-x) dx && (43)\\ &= \frac{3}{4}\left[x^2 - \frac{1}{3}x^3\right]_0^1 dx && (44)\\ &= \frac{3}{4}\left(1 - \frac{1}{3}\right) && (45)\\ &= 1/2. && (46) \end{aligned}$$

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.

s.15.2. Let Y denote the payout in thousands of euros. Then, $Y = \min\{X, 1\}$. We find

$$
\begin{aligned}
\mathsf{E}[Y|X \le 1] &= \mathsf{E}[X|X \le 1] && (47)\\
&= \int_0^1 x f_X(x|X \le 1) dx && (48)\\
&= \int_0^1 x \frac{f_X(x)}{\mathsf{P}\{X \le 1\}} dx && (49)\\
&= 2\int_0^1 x \frac{3}{4} x(2-x) dx && (50)\\
&= \frac{3}{2}\int_0^1 (2x^2 - x^3) dx && (51)\\
&= \frac{3}{2}\left[\frac{2}{3}x^3 - \frac{1}{4}x^4\right]_0^1 && (52)\\
&= \frac{3}{2}\left(\frac{2}{3} - \frac{1}{4}\right) && (53)\\
&= \frac{5}{8}. && (54)
\end{aligned}
$$

Writing down $\mathsf{E}[X|X \le 1]$: 0.5 point.
Correct derivation and solution: 1 point.

s.15.3. By the law of total expectation,

$$
\begin{aligned}
\mathsf{E}[Y] &= \mathsf{P}\{X \le 1\}\,\mathsf{E}[Y|X \le 1] + \mathsf{P}\{X > 1\}\,\mathsf{E}[Y|X > 1] && (55)\\
&= \frac{1}{2}\frac{5}{8} + \frac{1}{2}\cdot 1 && (56)\\
&= \frac{13}{16}. && (57)
\end{aligned}
$$

Mentioning law of total expectation: 0.5 point.
Correctly using law of total expectation: 0.5 point.
Correct solution: 0.5 point.

s.15.4. By Adam's law,

$$
\begin{aligned}
\mathsf{E}[T] &= \mathsf{E}[\mathsf{E}[T|Y]] && (58)\\
&= \mathsf{E}\left[\frac{3}{2}Y\right] && (59)\\
&= \frac{3}{2}\mathsf{E}[Y] && (60)\\
&= \frac{3}{2}\cdot\frac{13}{16} = \frac{39}{32}. && (61)
\end{aligned}
$$

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.

s.16.1. We have

$$
\begin{aligned}
\mathsf{P}\{X \le 1\} &= \int_0^1 \frac{3}{4} x(2-x) dx && (62)\\
&= \frac{3}{4}\left[x^2 - \frac{1}{3}x^3\right]_0^1 dx && (63)\\
&= \frac{3}{4}\left(1 - \frac{1}{3}\right) && (64)\\
&= 1/2. && (65)
\end{aligned}
$$

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.

s.16.2. Note that $Y = \min\{X, 1\}$. We find

$$\begin{aligned}
\mathsf{E}[Y|X \le 1] &= \mathsf{E}[X|X \le 1] && (66)\\
&= \int_0^1 x f_X(x|X \le 1)dx && (67)\\
&= \int_0^1 x \frac{f_X(x)}{\mathsf{P}\{X \le 1\}} dx && (68)\\
&= 2\int_0^1 x\frac{3}{4}x(2-x)dx && (69)\\
&= \frac{3}{2}\int_0^1 (2x^2 - x^3)dx && (70)\\
&= \frac{3}{2}\left[\frac{2}{3}x^3 - \frac{1}{4}x^4\right]_0^1 && (71)\\
&= \frac{3}{2}\left(\frac{2}{3} - \frac{1}{4}\right) && (72)\\
&= \frac{5}{8}. && (73)
\end{aligned}$$

Writing down $\mathsf{E}[X|X \le 1]$: 0.5 point.
Correct derivation and solution: 1 point.

s.16.3. By the law of total expectation,

$$\begin{aligned}
\mathsf{E}[Y] &= \mathsf{P}\{X \le 1\}\,\mathsf{E}[Y|X \le 1] + \mathsf{P}\{X > 1\}\,\mathsf{E}[Y|X > 1] && (74)\\
&= \frac{1}{2}\frac{5}{8} + \frac{1}{2}\cdot 1 && (75)\\
&= \frac{13}{16}. && (76)
\end{aligned}$$

Mentioning law of total expectation: 0.5 point.
Correctly using law of total expectation: 0.5 point.
Correct solution: 0.5 point.

s.16.4. By Adam's law,

$$\begin{aligned}
\mathsf{E}[Z] &= \mathsf{E}[\mathsf{E}[Z|Y]] && (77)\\
&= \mathsf{E}\left[\frac{3}{2}Y\right] && (78)\\
&= \frac{3}{2}\mathsf{E}[Y] && (79)\\
&= \frac{3}{2}\cdot\frac{13}{16} = \frac{39}{32}. && (80)
\end{aligned}$$

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.

s.17.1. 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.

s.17.2. 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.)

s.17.3. 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.

s.18.1. 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.

s.18.2.

$$\mathsf{E}[S] = \mathsf{E}[T] + \mathsf{E}[N(T)]\,\mathsf{E}[R] = 1/\mu + r\lambda\,\mathsf{E}[T] = 1/\mu + r\lambda/\mu.$$

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.)

s.18.3. 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.

s.19.1. 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.

s.19.2. 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:

$$R \neq rT \tag{81}$$
$$\mathsf{E}[R] \neq ra/2 \tag{82}$$
$$\tag{83}$$

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.)

s.19.3. 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.

s.20.1. 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.

s.20.2. 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.

$$\mathsf{E}[NR\,|\,N] \neq \mathsf{E}[NR] = \mathsf{E}[N]\,\mathsf{E}[R] \tag{84}$$

$$\mathsf{E}[S] \neq \mathsf{E}[T] + N(t)\,\mathsf{E}[R]. \tag{85}$$

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.)

s.20.3. 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.

s.21.1. (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, $\mathsf{P}\{X \geq a\} = \mathsf{P}\{f(X) \geq f(a)\}$.
(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 $\mathsf{P}\{X \geq a\} \leq \mathsf{P}\{f(X) \geq f(a)\}$, 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.

s.21.2. 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.

s.21.3. 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 $\mathsf{P}\{|Z| > 2\} \le e^{-3/2}(1-3/4)^{-1/2} \approx 0.446$.

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).

s.22.1. Let $Y = Z^2$, then $Y \sim \chi_1^2$, so $Y \sim \text{Gamma}(\frac{1}{2}, \frac{1}{2})$. We have

$$\begin{aligned}
\mathsf{E}\left[Z^{2n+2}\right] &= \mathsf{E}\left[Y^{n+1}\right] = \frac{\Gamma(n+3/2)}{(1/2)^{n+1}\Gamma(1/2)} = \frac{2^{n+1}\Gamma(n+3/2)}{\Gamma(1/2)} \\
\mathsf{E}\left[Z^{2n}\right] &= \mathsf{E}\left[Y^{n}\right] = \frac{\Gamma(n+1/2)}{(1/2)^{n}\Gamma(1/2)} = \frac{2^{n}\Gamma(n+1/2)}{\Gamma(1/2)}.
\end{aligned}$$

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.

s.22.2. By using the previous exercise with $n = 1$, we get $\mathsf{E}\left[Z^4\right] = 3\,\mathsf{E}\left[Z^2\right] = 3$.

s.22.3. 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).

s.22.4. By Markov's inequality,

$$\mathsf{P}\{|Z| > 3\} = \mathsf{P}\left\{Z^{2n} > 9^n\right\} \le \frac{\mathsf{E}\left[Z^{2n}\right]}{9^n}.$$

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,

$$\mathsf{P}\{|Z| > 3\} \le \frac{105}{9^4} \approx 0.016.$$

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 $\mathsf{P}\{|Z| > 3\} \le \frac{\mathsf{E}[Z^{2n}]}{9^n}$, 1 for showing that the best bound is obtained for $n = 4$ and $n = 5$, and 0.5 for calculating the resulting bound.

s.23.1. (i). Since g is increasing, we have $Z \ge a$ if and only if $g(Z) \ge g(a)$, so $\{Z \ge a\}$ and $\{g(Z) \ge g(a)\}$ are the same event. Hence, $\mathsf{P}\{Z \ge a\} = \mathsf{P}\{g(Z) \ge g(a)\}$.
(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 $g(Z) \ge g(a)$ if $Z \ge a$, then you are only proving that $\mathsf{P}\{Z \ge a\} \le \mathsf{P}\{g(X) \ge g(a)\}$, because if you just say "$g(Z) \ge g(a)$ if $Z \ge a$", $g(Z) \ge g(a)$ could still be true in cases where $Z \ge a$ is not, and hence $g(Z) \ge g(a)$ 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.

s.23.2. 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

$$\mathsf{P}\{|Y| > 3\} \le e^{-9t}\,\mathsf{E}\left[e^{t|Y|^2}\right] = e^{-9t}\,\mathsf{E}\left[e^{tY^2}\right].$$

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:

$$\mathsf{P}\{|Z| > 3\} = \mathsf{P}\left\{Z^2 > 9\right\} \le e^{-9t}\,\mathsf{E}\left[e^{t|Z|^2}\right] = e^{-9t}\,\mathsf{E}\left[e^{tZ^2}\right],$$

  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.

s.23.3. 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 $\mathsf{P}\{|Y| > 3\} \le e^{-4}(1-8/9)^{-1/2} \approx 0.0549$.

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).

s.24.1. 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 $\mathsf{E}[Y^{2n+2}] = \mathsf{E}[Y^2]\,\mathsf{E}[Y^{2n}]$. 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 $\mathsf{E}[Y^{2n+2}]$ and $\mathsf{E}[Y^{2n}]$,
0.5 for using a property of the Gamma function to finish the answer.

s.24.2. By applying the previous exercise with $n = 1$, we obtain that $\mathsf{E}[Y^4] = 3\,\mathsf{E}[Y^2] = 3$. By applying the previous exercise with $n = 3$ and $n = 2$, we obtain that $\mathsf{E}[Y^8] = 7\,\mathsf{E}[Y^6] = 7\cdot 5\,\mathsf{E}[Y^4] = 7\cdot 5\cdot 3 = 105$.

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 $\mathsf{E}[Y^2] = 1$ for $Y \sim \text{Norm}(0,1)$.

- Grading: 0.5 for a correct solution for $\mathsf{E}[Y^4]$ and 0.5 for a correct solution for $\mathsf{E}[Y^8]$.

s.24.3. We have $\mathsf{P}\{|Y| > 4\} = \mathsf{P}\{Y^4 > 256\}$ 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).

s.24.4. By Markov's inequality,

$$\mathsf{P}\{|Y| > 4\} = \mathsf{P}\{Y^{2n} > 16^n\} \leq \frac{\mathsf{E}[Y^{2n}]}{16^n}.$$

From the formula for $\mathsf{E}[Y^{2n}]$ we see that $\mathsf{E}[Y^{2(n+1)}] = (2n+1)\mathsf{E}[Y^{2n}]$. 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 $n = 8$.

s.25.1. 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

s.25.2. 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.

s.25.3. 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.*

s.26.1. 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.

s.26.2. 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.

s.26.3. 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.

s.26.4. 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.

s.27.1. Since (X,Y) is uniformly distributed on $(-\sqrt{e},\sqrt{e})^2$, their joint PDF is simply:

$$
\begin{aligned}
f(x,y) &= \left(\frac{1}{\sqrt{e}-(-\sqrt{e}))}\right)\left(\frac{1}{(\sqrt{e}-(-\sqrt{e}))}\right) \\
&= \frac{1}{4e},
\end{aligned}
$$

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.*

s.27.2. Note that $S = X^2 + Y^2$. Using LOTUS and symmetry in x and y:

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(x^2+y^2\right)f(x,y)\,dx\,dy &= \int_{-\sqrt{e}}^{\sqrt{e}}\int_{-\sqrt{e}}^{\sqrt{e}}\left(x^2+y^2\right)\left(\frac{1}{4e}\right)dx\,dy \\
&= \frac{1}{4e}\int_{-\sqrt{e}}^{\sqrt{e}}\int_{-\sqrt{e}}^{\sqrt{e}}(x^2+y^2)\,dx\,dy \\
&= \frac{1}{4e}\int_{-\sqrt{e}}^{\sqrt{e}}\int_{-\sqrt{e}}^{\sqrt{e}}x^2\,dx\,dy + \frac{1}{4e}\int_{-\sqrt{e}}^{\sqrt{e}}\int_{-\sqrt{e}}^{\sqrt{e}}y^2\,dx\,dy + \\
&= \frac{2}{4e}\int_{-\sqrt{e}}^{\sqrt{e}}\int_{-\sqrt{e}}^{\sqrt{e}}x^2\,dx\,dy \\
&= \frac{1}{2e}\int_{-\sqrt{e}}^{\sqrt{e}}[\frac{1}{3}x^3]_{\sqrt{e}}^{\sqrt{e}}\,dy \\
&= \frac{1}{6e}\int_{-\sqrt{e}}^{\sqrt{e}}(e^{3/2}+e^{3/2})\,dy \\
&= \frac{2}{6e}2\sqrt{e}e^{3/2} \\
&= \frac{4e^2}{6e} = \frac{2}{3}e.
\end{aligned}$$

One point for finding $S^2 = X^2 + Y^2$ and writing down the integral correctly using LOTUS. Two points for the remaining calculations.

s.27.3. 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.

s.28.1. The random vector (X, Y) is uniformly distributed on $(-\pi, \pi)^2$. Hence, the joint pdf is given by

$$\begin{aligned}
f(x,y) &= \left(\frac{1}{\pi-(-\pi))}\right)\left(\frac{1}{(\pi-(-\pi))}\right) \\
&= \frac{1}{4\pi^2}
\end{aligned}$$

for $-\pi < x < \pi$ and $-\pi < x < \pi$. *0.5 points for the correct solution, 0.5 points for the boundaries.*

s.28.2. Note that $N^2 = c^2(X^2+Y^2)$. Using LOTUS:

$$\begin{aligned}
\mathsf{E}\left[N^2\right] &= \mathsf{E}\left[c^2(X^2+Y^2)\right] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} c^2\left(x^2+y^2\right) f(x,y)\,dx\,dy \\
&= c^2\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left(x^2+y^2\right)\left(\frac{1}{4\pi^2}\right)dx\,dy \\
&= c^2\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\frac{x^2+y^2}{4\pi^2}\,dx\,dy \\
&= \frac{c^2}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} x^2+y^2\,dx\,dy \\
&= \frac{c^2}{4\pi^2}\int_{-\pi}^{\pi}\left[\frac{x^3}{3}+y^2x\right]_{-\pi}^{\pi}dy \\
&= \frac{c^2}{4\pi^2}\int_{-\pi}^{\pi}\left(\left(\frac{\pi^3}{3}+y^2\pi\right)-\left(\frac{(-\pi)^3}{3}-y^2\pi\right)\right)dy \\
&= \frac{c^2}{4\pi^2}\int_{-\pi}^{\pi}\left(\frac{2\pi^3}{3}+2y^2\pi\right)dy \\
&= \frac{c^2}{4\pi^2}\int_{-\pi}^{\pi}\frac{2}{3}\pi^3+2\pi y^2\,dy \\
&= \frac{c^2}{4\pi^2}\left[\frac{2}{3}\pi^3 y+\frac{2}{3}\pi y^3\right]_{-\pi}^{\pi} \\
&= \frac{c^2}{4\pi^2}\left(\frac{2}{3}\pi^4+\frac{2}{3}\pi^4+\frac{2}{3}\pi^4+\frac{2}{3}\pi^4\right) \\
&= \frac{c^2}{4\pi^2}\frac{8}{3}\pi^4 = c^2\frac{2\pi^2}{3}
\end{aligned}$$

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.

s.28.3. 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.

s.29.1. 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

$$F_Y(y) = \mathsf{P}\{Y \le y\} = \begin{cases} \mathsf{P}\{X \le 0\} + \mathsf{P}\left\{X \ge \frac{1}{y}\right\} & \text{if } y > 0 \\ \mathsf{P}\left\{\frac{1}{y} \le X \le 0\right\} & \text{if } y < 0 \end{cases}.$$

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.

s.29.2.

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) = \begin{cases} \frac{\mathrm{d}}{\mathrm{d}y}(\frac{1}{2} + \mathsf{P}\left\{X \geq \frac{1}{y}\right\}) & \text{if } y > 0 \\ \frac{\mathrm{d}}{\mathrm{d}y}(\frac{1}{2} - \mathsf{P}\left\{X \leq \frac{1}{y}\right\}) & \text{if } y < 0 \end{cases}$$
$$= \begin{cases} -\frac{\mathrm{d}}{\mathrm{d}y}\mathsf{P}\left\{X \leq \frac{1}{y}\right\} & \text{if } y > 0 \\ -\frac{\mathrm{d}}{\mathrm{d}y}\mathsf{P}\left\{X \leq \frac{1}{y}\right\} & \text{if } y < 0 \end{cases}.$$

We can write, by definition:

$$\mathsf{P}\left\{X \leq \frac{1}{y}\right\} = \int_{-\infty}^{\frac{1}{y}} f_X(s)\,\mathrm{d}s,$$

such that by the FTC,

$$-\frac{\mathrm{d}}{\mathrm{d}y}\mathsf{P}\left\{X \leq \frac{1}{y}\right\} = \frac{1}{y^2} f_X(\frac{1}{y})$$

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.

s.29.3. When $v = 1$, X follows a Cauchy distribution. Then, Y must also be Cauchy.

Grading scheme:

- Correct 1pt.

s.29.4. It converges to the standard normal distribution.

Grading scheme:

- Correct 0.5pt.

s.29.5. For $v > 1$, we have that

$$f_Y(y) \propto \left(y + \frac{1}{vy}\right)^{\frac{v}{2}+\frac{1}{2}}.$$

The FOC tells us that the mode is at

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}y}\left(y + \frac{1}{vy}\right)^{\frac{v}{2}+\frac{1}{2}} = \left(y + \frac{1}{vy}\right)^{\frac{v}{2}-\frac{1}{2}}\left(1 - \frac{1}{vy^2}\right) = 0 \implies \\ &1 - \frac{1}{vy^2} = 0 \implies \\ &y = \pm\frac{\sqrt{v}}{v}. \end{aligned}$$

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.

s.30.1. By the book, we know that $\min\{X_1, X_2, \ldots, X_n\} \sim \text{Expo}(n\lambda)$.

Grading scheme:

- Correct 0.5pt.

s.30.2. 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.

s.30.3. 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.

s.30.4. 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.

s.31.1. 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.

s.31.2. 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.

s.31.3. Clearly, $W_k = \frac{k}{l} W_l$, and thus we can see that

$$
\begin{aligned}
W_k W_l &= \frac{k}{l} W_k^2, \\
\frac{W_k}{W_l} &= \frac{k}{l}.
\end{aligned}
$$

These are independent, since one is a constant.

Grading scheme:

- Correct probability 0.5pt.
- Correct conclusion 0.5pt.

s.31.4. 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,

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi\sqrt{x_1x_2}} \exp\left(-\frac{x_1+x_2}{2}\right)$$

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.

s.32.1. Since we are dealing with discrete uniform, we see that $\mathsf{P}\{B = k\} = \mathsf{P}\{U = k\} + \mathsf{P}\{U = -k\}$ for $k = 1, 2, \ldots, n$, 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.

s.32.2. First, note that $Z = a + (b-a)X$. Then, we know that

$$F_Z(y) = \mathsf{P}\{X(b-a)+a \leq y\} = \mathsf{P}\left\{X \leq \frac{y-a}{b-a}\right\} = \int_0^{\frac{y-a}{b-a}} f_X(s)\mathrm{d}s.$$

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.

s.32.3. We start as usual by considering the CDF of $Q = |Z|$. This shows us that

$$F_Q(y) = \mathsf{P}\{Q \leq y\} = \mathsf{P}\{|Z| \leq y\} = \mathsf{P}\{-y \leq Z \leq y\} = F_Z(y) - F_Z(-y).$$

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.

s.33.1.

$$\begin{aligned}
&\mathsf{Cov}[2N_1 + N_2, 2N_1 - N_2] \\
=&\mathsf{Cov}[2N_1, 2N_1] - \mathsf{Cov}[2N_1, N_2] + \mathsf{Cov}[N_2, 2N_1] - \mathsf{Cov}[N_2, N_2] \cdots \text{(0.5 point)} \\
=&4 \cdot Var(N_1) - Var(N_2) \\
=&4\lambda_1 - \lambda_2 \neq 0. \cdots \text{(0.5 point)}
\end{aligned}$$

s.33.2. 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)

s.33.3. 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)

$$\begin{aligned}
\mathsf{Cov}[N, X] &= \mathsf{Cov}[X + Y, X] \\
&= \mathsf{Cov}[X, X] + \mathsf{Cov}[Y, X] \\
&= Var(X) \cdots \text{(0.5 point)} \\
&= (\lambda_1 + \lambda_2)p \cdots \text{(0.5 point)}
\end{aligned}$$

Then it follows that

$$\begin{aligned}
\rho_{X,N} &= \frac{\mathsf{Cov}[N, X]}{sd(N) \cdot sd(X)} \\
&= \frac{(\lambda_1 + \lambda_2)p}{\sqrt{\lambda_1 + \lambda_2} \cdot \sqrt{(\lambda_1 + \lambda_2)p}} \\
&= \sqrt{p}. \cdots \text{(0.5 point)}
\end{aligned}$$

s.33.4. 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.)

s.34.1.

$$\begin{aligned}
&\mathsf{Cov}[N_1 + N_2, N_2 - 2N_3] \\
=&\mathsf{Cov}[N_1, N_2] - \mathsf{Cov}[N_1, 2N_3] + \mathsf{Cov}[N_2, N_2] - \mathsf{Cov}[N_2, 2N_3] \cdots \text{(0.5 point)} \\
=&Var(N_2) \\
=&\lambda_2 > 0. \cdots \text{(0.5 point)}
\end{aligned}$$

s.34.2. 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)

s.34.3. 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)

$$\begin{aligned}
\mathsf{Cov}[N, X] &= \mathsf{Cov}[X + Y, X] \\
&= \mathsf{Cov}[X, X] + \mathsf{Cov}[Y, X] \\
&= Var(X) \cdots \text{(0.5 point)} \\
&= (\lambda_1 + \lambda_2 + \lambda_3)p \cdots \text{(0.5 point)}
\end{aligned}$$

Then it follows that

$$\begin{aligned}
\rho_{X,N} &= \frac{\mathsf{Cov}[N, X]}{sd(N) \cdot sd(X)} \\
&= \frac{(\lambda_1 + \lambda_2 + \lambda_3)p}{\sqrt{\lambda_1 + \lambda_2 + \lambda_3} \cdot \sqrt{(\lambda_1 + \lambda_2 + \lambda_3)p}} \\
&= \sqrt{p}. \cdots \text{(0.5 point)}
\end{aligned}$$

s.34.4. 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.)

s.35.1. 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

$$\begin{aligned}
&\mathsf{Cov}[X_A, X]\\
=&\mathsf{Cov}[X_A, X_A + X_B]\\
=&\mathsf{Cov}[X_A, X_A] + \mathsf{Cov}[X_A, X_B]\\
=&Var(X_A)\\
=&\lambda p \cdots \text{ (0.5 points)}
\end{aligned}$$

Then

$$\rho_{X_A,X} = \frac{\mathsf{Cov}[X_A,X]}{sd(X_A)sd(X)} = \frac{\lambda p}{\sqrt{\lambda p}\sqrt{\lambda}} = \sqrt{p} \cdots \text{ (0.5 points)}$$

s.35.2. 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,

$$\mathsf{Cov}[X_A, X_C] = -\frac{1000}{3^2}, \cdots \text{ (0.5 points)}$$

$$\rho_{X_A,X_C} = \frac{\mathsf{Cov}[X_A,X_C]}{sd(X_A)sd(X_C)} = -\frac{1000/3^2}{1000(1/3)(2/3)} = -\frac{1}{2}. \cdots \text{ (0.5 points)}$$

s.35.3. 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.)

s.36.1. 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

$$
\begin{aligned}
&\mathsf{Cov}[X_B, N] \\
=&\mathsf{Cov}[X_B, X_A + X_B] \\
=&\mathsf{Cov}[X_B, X_A] + \mathsf{Cov}[X_B, X_B] \\
=&Var(X_B) \\
=&\lambda(1-p)\cdots \text{ (0.5 points)}
\end{aligned}
$$

Then

$$\rho_{X_B,N} = \frac{\mathsf{Cov}[X_B, N]}{sd(X_B)sd(N)} = \frac{\lambda(1-p)}{\sqrt{\lambda(1-p)}\sqrt{\lambda}} = \sqrt{1-p}\cdots \text{ (0.5 points)}$$

s.36.2. 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,

$$\mathsf{Cov}[X_B, X_C] = -\frac{500}{4^2}, \cdots \text{ (0.5 points)}$$

$$\rho_{X_B,X_C} = \frac{\mathsf{Cov}[X_B, X_C]}{sd(X_B)sd(X_C)} = -\frac{500/4^2}{500(1/4)(3/4)} = -\frac{1}{3}. \cdots \text{ (0.5 points)}$$

s.36.3. 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.)

s.37.1. Using Bayes' rule we have

$$
\begin{aligned}
f_1(p|X_1 = x_1) &= \frac{\mathsf{P}\{X_1 = x_1|p\} f_0(p)}{\mathsf{P}\{X_1 = x_1\}} && (86)\\
&\propto \mathsf{P}\{X_1 = x_1|p\} f_0(p) && (87)\\
&= \binom{x_1 + r - 1}{x_1}(1-p)^r p^{x_1} \frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1} && (88)\\
&\propto (1-p)^{r+b-1} p^{a+x_1-1}, && (89)
\end{aligned}
$$

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.

s.37.2. Yes. The posterior distribution is in the same family of distributions (Beta) as the prior. Hence, Amy has a conjugate prior.

s.37.3. 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 $X_1 = x_1, \ldots, X_n = x_n$ is a Beta($a + \sum_{i=1}^n x_i, b + rn$) distribution.

s.38.1. Using Bayes' rule we have

$$
\begin{aligned}
f_1(p|Y_1 = y_1) &= \frac{\mathsf{P}\{Y_1 = y_1|p\} f_0(p)}{\mathsf{P}\{Y_1 = y_1\}} && (90)\\
&\propto \mathsf{P}\{Y_1 = y_1|p\} f_0(p) && (91)\\
&= \binom{y_1 + r - 1}{y_1}(1-p)^r p^{y_1} \frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1} && (92)\\
&\propto (1-p)^{r+b-1} p^{a+y_1-1}, && (93)
\end{aligned}
$$

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.

s.38.2. Yes. The posterior distribution is in the same family of distributions (Beta) as the prior. Hence, Amy has a conjugate prior.

s.38.3. 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.

s.39.1. 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

$$
\begin{aligned}
\mathsf{P}\{X = x, Y = y|X > Y\} &= \frac{\mathsf{P}\{X = x, Y = y, X > Y\}}{\mathsf{P}\{X > Y\}} && (94)\\
&= \frac{1/9}{1/3} = 1/3, && (95)
\end{aligned}
$$

for $(x, y) \in \{(2,1), (3,1), (3,2)\}$.

s.39.2. We have

$$
A = \begin{cases} 0 & X < Y \\ 1/2 & X = Y \\ X - Y & X > Y. \end{cases} \qquad (96)
$$

Hence,

$$
\begin{aligned}
\mathsf{E}[A|X > Y] &= \mathsf{E}[X - Y|X > Y] && (97)\\
&= \frac{1}{3}(2-1) + \frac{1}{3}(3-1) + \frac{1}{3}(3-2) && (98)\\
&= 4/3. && (99)
\end{aligned}
$$

s.39.3. We have, by the law of total expectation,

$$
\begin{aligned}
\mathsf{E}[A] &= \mathsf{P}\{X < Y\}\,\mathsf{E}[A|X < Y] + \mathsf{P}\{X = Y\}\,\mathsf{E}[A|X = Y] + \mathsf{P}\{X > Y\}\,\mathsf{E}[A|X > Y] && (100)\\
&= 0 + \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3}\frac{4}{3} && (101)\\
&= 11/18. && (102)
\end{aligned}
$$

s.39.4. By Adam's law,

$$\begin{aligned}
\mathsf{E}[T] &= \mathsf{E}[\mathsf{E}[T|A]] && (103)\\
&= \mathsf{E}\left[\frac{3}{2}A\right] && (104)\\
&= \frac{3}{2}\mathsf{E}[A] && (105)\\
&= \frac{3}{2}\frac{11}{18} && (106)\\
&= \frac{33}{36}. && (107)
\end{aligned}$$

s.40.1. 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

$$\begin{aligned}
\mathsf{P}\{X = x, Y = y|X > Y\} &= \frac{\mathsf{P}\{X = x, Y = y, X > Y\}}{\mathsf{P}\{X > Y\}} && (108)\\
&= \frac{1/9}{1/3} = 1/3, && (109)
\end{aligned}$$

for $(x, y) \in \{(2,1),(3,1),(3,2)\}$.

s.40.2. We have

$$C = \begin{cases} 0 & X < Y \\ 1/2 & X = Y \\ X - Y & X > Y. \end{cases} \qquad (110)$$

Hence,

$$\begin{aligned}
\mathsf{E}[C|X > Y] &= \mathsf{E}[X - Y|X > Y] && (111)\\
&= \frac{1}{3}(2-1) + \frac{1}{3}(3-1) + \frac{1}{3}(3-2) && (112)\\
&= 4/3. && (113)
\end{aligned}$$

s.40.3. We have, by the law of total expectation,

$$\begin{aligned}
\mathsf{E}[C] &= \mathsf{P}\{X < Y\}\,\mathsf{E}[C|X < Y] + \mathsf{P}\{X = Y\}\,\mathsf{E}[C|X = Y] + \mathsf{P}\{X > Y\}\,\mathsf{E}[C|X > Y] && (114)\\
&= 0 + \frac{1}{3}\cdot\frac{1}{2} + \frac{1}{3}\frac{4}{3} && (115)\\
&= 11/18. && (116)
\end{aligned}$$

s.40.4. By Adam's law,

$$\begin{aligned}
\mathsf{E}[S] &= \mathsf{E}[\mathsf{E}[S|C]] && (117)\\
&= \mathsf{E}\left[\frac{3}{2}C\right] && (118)\\
&= \frac{3}{2}\mathsf{E}[C] && (119)\\
&= \frac{3}{2}\frac{11}{18} && (120)\\
&= \frac{33}{36}. && (121)
\end{aligned}$$

s.41.1. 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.

s.41.2.

$$\mathsf{P}\{B(h)=0, D(h)=1|L(0)=n\} = e^{-\lambda h}\mu n h e^{-\mu n h} + o(h() \tag{122}$$
$$= (1-\lambda h)\mu n h(1-\mu n h) + o(h) = \mu n h + o(h). \tag{123}$$

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.

s.41.3. Use conditional expectation and the above results to see that

$$\mathsf{E}[L(t+h)|L(t)=n] = n\,\mathsf{P}\{B(h)=0, D(h)=0\} + (n+1)\,\mathsf{P}\{B(h)=1, D(h)=0\} \tag{124}$$
$$+ (n-1)\,\mathsf{P}\{B(h)=0, D(h)=1\} + o(h) \tag{125}$$
$$= ne^{-\lambda h}e^{-\mu n h} + (n+1)\lambda h + (n-1)\mu n h + o(h) \tag{126}$$
$$= n(1-\lambda h)(1-\mu n h) + (n+1)\lambda h + (n-1)\mu n h + o(h) \tag{127}$$
$$= n - n(\lambda + \mu n)h + (n+1)\lambda h + (n-1)\mu n h + o(h) \tag{128}$$
$$= n + (\lambda - \mu n)h + o(h). \tag{129}$$

s.41.4. Replace n by $L(t)$ in $\mathsf{E}[L(t+h)|L(t)]$ to see that

$$\mathsf{E}[L(t+h)|L(t)] = L(t) + (\lambda - \mu L(t))h + o(h). \tag{130}$$

Take expectations left and right and use Adam's law.

s.42.1. 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.

s.42.2.

$$\mathsf{P}\{B(h)=0, D(h)=1|X(0)=n\} = e^{-(\lambda n+\theta)h}\mu n h e^{-\mu n h} = (1-(\lambda n+\theta)h)\mu n h(1-\mu n h) + o(h) = \mu n h + o(h). \tag{131}$$

Grading:

- Skipping the algebra: -1/2.

s.42.3.

$$\mathsf{E}[X(t+h)|X(t)=n] = n\,\mathsf{P}\{B(h)=0, D(h)=0\} + (n+1)\,\mathsf{P}\{B(h)=1, D(h)=0\} \tag{132}$$
$$+ (n-1)\,\mathsf{P}\{B(h)=0, D(h)=1\} + o(h) \tag{133}$$
$$= ne^{-(\lambda n+\theta)h}e^{-\mu n h} + (n+1)(\lambda n+\theta)h + (n-1)\mu n h + o(h) \tag{134}$$
$$= n(1-(\lambda n+\theta)h)(1-\mu n h) + (n+1)(\lambda n+\theta)h + (n-1)\mu n h + o(h) \tag{135}$$
$$= n + (\lambda n + \theta - \mu n)h + o(h). \tag{136}$$

Grading:

- Show also how to simplify the results of the first question. If not, -1/2.

s.42.4. Replace n by $X(t)$ in $\mathsf{E}[X(t+h)|X(t)]$ to see that

$$\mathsf{E}[X(t+h)|X(t)] = X(t) + (\lambda-\mu)X(t)h + \theta h + o(h). \tag{137}$$

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.

s.43.1. Standard consequence of exponential rvs.

Grading:

- The time to the next arrival is not λ. -0.5.

s.43.2. 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.

s.43.3. Just fill in the expression in the previous exercise and check that the RHS and LHS match.

s.43.4. Use conditioning on L.

$$\mathsf{E}[T] = \mathsf{E}[\mathsf{E}[T|L(0)]] = \mathsf{E}\left[L(0)/(\mu-\lambda)\right] = \frac{\rho}{1-\rho}\frac{1}{\mu-\lambda} = \frac{\lambda}{\mu^2(1-\rho)^2}.$$

s.43.5. 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.

s.44.1. After selecting tunnel B, which takes 3 minutes to travel, the mouse is back in the pit again, and the process starts over again.

s.44.2.

$$\mathsf{E}[T] = \mathsf{E}[T|X=A]/3 + \mathsf{E}[T|X=B]/3 + \mathsf{E}[T|X=C]/3. \tag{138}$$

$$\mathsf{E}[T|X=A] = 2 \tag{139}$$

$$\mathsf{E}[T|X=B] = 3 + \mathsf{E}[T] \tag{140}$$

$$\mathsf{E}[T|X=C] = 4 + \mathsf{E}[T]. \tag{141}$$

Solving gives $\mathsf{E}[T] = 9$.

Grading

- Not using the result of subquestion 1: no points.

s.44.3.

$$\mathsf{V}[T|X=A] = 0 \tag{142}$$

$$\mathsf{V}[T|X=B] = \mathsf{V}[T] \tag{143}$$

$$\mathsf{V}[T|X=C] = \mathsf{V}[T] \tag{144}$$

$$\mathsf{E}[\mathsf{V}[T|X]] = \mathsf{V}[T]2/3. \tag{145}$$

$$\mathsf{E}[T|X] = 2I_{X=A} + (3+\mathsf{E}[T])I_{X=B} + (4+\mathsf{E}[T])I_{X=C} \tag{146}$$

$$= 2I_{X=A} + 12I_{X=B} + 13I_{X=C} \tag{147}$$

$$\mathsf{V}[\mathsf{E}[T|X]] = 4\cdot 2/9 + 144\cdot 2/9 + 169\cdot 2/9 =: \alpha \tag{148}$$

$$\mathsf{V}[T] = \mathsf{V}[T]2/3 + \alpha \quad \text{EVE} \tag{149}$$

$$\mathsf{V}[T] = 3\alpha. \tag{150}$$

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: $\mathsf{E}[T^2] = \ldots + (3+\mathsf{E}[])]^2\, 1/3 + \ldots$. This is not correct of course.

s.44.4. 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.

s.45.1. 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.

s.45.2. 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

s.45.3. By the CLT, $Z_1 \dot{\sim} \text{Norm}(0,1)$. Hence, $Y_1 = \mu + \sigma/\sqrt{n} Z_1 \dot{\sim} \text{Norm}(0.5, 0.25/n)$.

Grading scheme:

- 0.5 for mentioning CLT and the distribution of Z_1;
- 0.5 for the approximate distribution of Y_1.

s.45.4. 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).

s.45.5. By LLN, S does converge to a constant as $\ell \to \infty$, however, it converges to $\mathsf{E}[Z_1^{37}]$ for that fixed value of k. By symmetry, we have $\mathsf{E}[Z_1^{37}] = 0$. However, the gamma distribution is right-skewed, which implies $\mathsf{E}[T^{37}] > 0$. Hence, it does not converge to $\mathsf{E}[T^{37}]$.

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 $\mathsf{E}[T^{37}]$.

s.46.1. 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.

s.46.2. 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

s.46.3. By the CLT, $Z_1 \dot{\sim} \text{Norm}(0,1)$. Hence, $Y_1 = \mu + \sigma/\sqrt{n} Z_1 \dot{\sim} \text{Norm}(2, 4/n)$.
Grading scheme:

- 0.5 for mentioning CLT and the distribution of Z_1;
- 0.5 for the approximate distribution of Y_1.

s.46.4. 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).

s.46.5. 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 $\mathsf{E}\left[T^{29}\right]$.

s.47.1. 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.

s.47.2. 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

s.47.3. By the CLT, $Z_1 \dot{\sim} \text{Norm}(0,1)$. Hence, $Y_1 = \mu + \sigma/\sqrt{n} Z_1 \dot{\sim} \text{Norm}(0.25, 0.0625/n)$.
Grading scheme:

- 0.5 for mentioning CLT and the distribution of Z_1;

- 0.5 for the approximate distribution of Y_1.

s.47.4. 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).

s.47.5. 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 $\mathsf{E}\left[T^{53}\right]$.

s.48.1. 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.

s.48.2. 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

s.48.3. By the CLT, $Z_1 \mathrel{\dot\sim} \text{Norm}(0,1)$. Hence, $Y_1 = \mu + \sigma/\sqrt{n} Z_1 \mathrel{\dot\sim} \text{Norm}(3, 9/n)$.
Grading scheme:

- 0.5 for mentioning CLT and the distribution of Z_1;
- 0.5 for the approximate distribution of Y_1.

s.48.4. 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).

s.48.5. 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 $\mathsf{E}\left[T^{71}\right]$.