Probability distributions EBP038A05: 2020-2021
Assignments plus solutions

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

February 28, 2021

CONTENTS

GENERAL INFORMATION

Here we just provide the exercises of the assignments. For information with respect to grading we refer to the course manual.

Each assignment contains several sections. The first section is meant to help you read the book well and become familiar with definitions and concepts of probability theory. These questions are mostly simple checks, not at exam level, but lower. The second section contains some exercises at about the exam level to get you started. Here you have to derive and explain a solution, in mathematical notation. Most of the selected exercises of the book are also at about (or just a bit above) exam level. The third section is about coding skills. We explain the rationale presently. The final section with challenges is for those students that like a challenge; the problems are above exam level.

You have to get used to programming and checking your work with computers, for instance by using simulation. The coding exercises address this skill. You should know that much of programming is 'monkey see, monkey do'. This means that you take code of others, try to understand it, and then adapt it to your needs. For this reason we include the code to answer the question. The idea is that you copy the code, you run it and include the numerical results in your report. You should be able to explain how the code works. For this reason we include questions in which you have explain how the most salient parts of the code works.

We include python and R code, and leave the choice to you what to use. In the exam we will also include both languages in the same problem, so you can stay with the language you like. You should know, however, that many of you will need to learn multiple languages later in life. For instance, when you have to access databases to obtain data about customers, patients, clients, suppliers, inventory, demand, lifetimes (whatever), you often have to use `sql`. Once you have the raw data, you process it with `R` or python to do statistics or make plots. (While I (= NvF) worked at a bank, I used Fortran for numerical work, AWK for string parsing and making tables, excel, SAS to access the database, and matlab for other numerical work, all next to each other. I got tired of this, so I went to using python as it did all of this stuff, but then within one language.) For your interest, based on the statistics here or here, python scores (much) higher than R in popularity; if you opt for a business career, the probability you have to use python is simply higher than to have to use R.

You should become familiar with look up documentation on coding on the web, no matter your programming language of choice. Invest time in understanding the, at times, rather technical and terse, explanations. Once you are used to it, the core documentation is faster to read, i.e., less clutter. In the long run, it pays off.

The rules:

1. For each assigment you have to turn in a pdf document typeset in LaTeX. Include a title, group number, student names and ids, and date.

2. We expect brief answers, just a sentence or so, or a number plus some short explanation. The idea of the assignment is to help you studying, not to turn you in a writer.

3. When you have to turn in a graph, provide decent labels and a legend, ensure the axes have labels too.

1 ASSIGNMENT 1

1.1 *Have you read well?*

Ex 1.1. In your own words, explain what is

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

Ex 1.2. We have two r.v.s $X, Y \in [0,1]^2$ (here $[0,1]^2 = [0,1] \times [0,1]$) with the joint PDF $f_{X,Y}(x,y) = 2 I_{x \leq y}$.

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

Ex 1.3. Correct (that is, is the following claim correct?)? We have two continuous r.v.s X, Y. Even though the joint CDF factors into the product of the marginals, i.e., $F_{X,Y}(x,y) = F_X(x)F_Y(y)$, it is still possible in general that the joint PDF does not factor into a product of marginals PDFs of X and Y, i.e., $f_{X,Y}(x,y) \neq f_X(x)f_Y(y)$.

Ex 1.4. Consider $F_{X,Y}(x,y)/F_X(x)$. Write this expression as a conditional probability. Is this equal to the conditional CDF of X and Y?

Ex 1.5. Let X be uniformly distributed on the set $\{0,1,2\}$ and let $Y \sim \text{Bern}(1/4)$; X and Y are independent.

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

Ex 1.6. Apply the chicken-egg story. A machine makes items on a day. Some items, independent of the other items, are failed (i.e., do not meet the quality requirements). What is N, what is p, what are the 'eggs' in this context, and what is the meaning of 'hatching'? What type of 'hatching' do we have here?

Ex 1.7. Correct? We have two r.v.s X and Y on $\mathbb{R}^+$. It is given that $F_{X,Y}(x,y) = F_X(x)F_Y(y)$ for $x, y \leq 1/3$. Then X and Y are necessarily independent.

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

Ex 1.9. Correct? For any two r.v.s X and Y on $\mathbb{R}^+$ with marginals F_X and F_Y, it holds that $\mathsf{P}\{X \leq x, Y \leq y\} = F_X(x)F_Y(y)$.

Ex 1.10. Theorem 7.1.11. What is the meaning of the notation $X|N = n$?

Ex 1.11. Correct? X, Y are two discrete r.v.s with CDF $F_{X,Y}$. We can compute the PDF as $\partial_x \partial_y F_{X,Y}(x, y)$.

1.2 *Exercise at about exam level*

Ex 1.12. This is about the simplest model for an insurance company that I can think of. We start with an initial capital $I_0 = 2$. The company receives claims and contributions every period, a week say. In the ith period, we receive a contribution X_i uniform on the set $\{1,2,\dots,10\}$ and a claim C_i uniform on $\{0,1,\dots 8\}$.

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

1.3 *Coding skills*

Ex 1.13. Use simulation to estimate the answer of BH.7.1. Run the code below and explain line 9 of python code or line 7 of the R code.

Then run the code for a larger sample, e.g, `num=1000` or so, but remove the prints of `a`, `b`, and `succes`, because that will fill your screen with numbers you don't need. Only for small simulations such output is handy so that you can check the code.

Compare the value of the simulation to the exact value.

```python
import numpy as np

np.random.seed(3)

num = 10

a = np.random.uniform(size=num)
b = np.random.uniform(size=num)
success = np.abs(a - b) < 0.25
print(a)
print(b)
print(success)
print(success.mean(), success.var())
```

```r
set.seed(3)

num <- 10

a <- runif(num)
b <- runif(num)
success <- abs(a-b) < 0.25
a
b
success
paste(mean(success), var(success))
```

Challenge (not obligatory): If you like, you can include a plot of the region (in time) in which Alice and Bob meet, and put marks on the points of the simulation that were 'successful'.

Ex 1.14. Let $X \sim \text{Exp}(3)$. Find a simple expression for $\mathsf{P}\{1 < X \leq 4\}$ and compute the value. Then use simulation to check this value. Finally, use numerical integration to compute this value. What are the numbers? Explain lines 11, 21 and 26 of the python code or lines 7, 17 and 23 of the R code.

```python
1 import numpy as np
2 from scipy.stats import expon
3 from scipy.integrate import quad
4
5 labda = 3
6
```

```python
X = expon(scale = 1 / labda).rvs(1000)
# print(X)
print(X.mean())

success = (X > 1) * (X < 4)
# print(success)
print(success.mean(), success.std())


def F(x):  # CDF
    return 1 - np.exp(-labda * x)


def f(x):  # density
    return labda * np.exp(-labda * x)


print(F(4) - F(1))

I = quad(f, 1, 4)
print(I)
```

```r
labda <- 3

X <- rexp(1000, rate = labda)
# X
mean(X)

success <- (X > 1) * (X < 4)
# print(success)
paste(mean(success), sd(success))


CDF <- function(x) {  # CDF
  return(1 - exp(-labda * x))
}

f <- function(x) {    # density
  return(labda * exp(-labda * x))
}


```

```
CDF(4) - CDF(1)

I = integrate(f, 1, 4)
I
```

1.4 *Challenges, optional*

You are free to chose one of these problems, but of course you can do both if you like.

A UNIQUENESS PROPERTY OF THE POISSON DISTRIBUTION Consider again the chicken-egg story (BH 7.1.9): A chicken lays a random number of eggs N an each egg independently hatches with probability p and fails to hatch with probability $q = 1-p$. Formally, $X|N \sim \text{Bin}(N,p)$. Assume also that $X|N \sim \text{Bin}(N,p)$ and that $N-X$ is independent of X. For $N \sim \text{Pois}(\lambda)$ it is shown in BH 7.1.9 that X and Y are independent. This exercise asks for the converse: showing that the independence of X and Y implies that $N \sim \text{Pois}(\lambda)$ for some λ. Hence, the Poisson distribution is quite special: it is the only distribution for which the number of hatched eggs doesn't tell you anything about the number of unhatched eggs.

Let $0 < p < 1$. Let N be an r.v. taking non-negative integer values with $P(N > 0) > 0$. Assume also that $X|N \sim \text{Bin}(N,p)$ and that $N-X$ is independent of X.

Ex 1.15. Use the assumption that $\mathsf{P}\{N > 0\} > 0$ to prove that N has support $\mathbb{N}$, i.e. $\mathsf{P}\{N = n\} > 0$ for all $n \in \mathbb{N}$. Note: $0 \in \mathbb{N}$.

Ex 1.16. Write $Y = N - X$. Prove that

$$\mathsf{P}\{X = x\}\,\mathsf{P}\{Y = y\} = \binom{x+y}{x} p^x (1-p)^y\,\mathsf{P}\{N = x+y\}. \tag{1.1}$$

Ex 1.17. Prove that N is Poisson distributed.

IMPROPER INTEGRALS AND THE CAUCHY DISTRIBUTION This problem challenges your integration skills and lets you think about the subtleties of integrating a function over an infinite domain. (Such integrals are called improper integrals.)

Assume that X has the Cauchy distribution. Recall that $\mathsf{E}[X]$ does not exist (hence, it is not automatic that the expectation of a some arbitrary r.v. exists).

Ex 1.18. Why does $\mathsf{E}\left[\frac{|X|}{X^2+1}\right]$ exist? Find its value. It is essential that you include your arguments.

Ex 1.19. Explain why the previous exercise implies that $\mathsf{E}\left[\frac{X}{X^2+1}\right]$ exists. Then find its value.

2 ASSIGNMENT 2

2.1 *Have you read well?*

Ex 2.1. Example 7.2.2. Write down the integral to compute $\mathsf{E}\left[(X-Y)^2\right]$. You don't have to solve the integral.

Ex 2.2. Give a brief example of a situation where it might be more convenient to employ the correlation instead of the covariance and explain why.

Ex 2.3. In queueing theory the concept of squared coefficient of variance SCV of a rv X is very important. It is defined as $C = \mathsf{V}[X]/(\mathsf{E}[X])^2$. Is the SCV of X equal to Corr(X,X)? Can it happen that $C > 1$?

Ex 2.4. Using the definition of Covariance (Definition 7.3.1) derive the expression $\mathsf{Cov}[X,Y] = \mathsf{E}[XY] - \mathsf{E}[X]\,\mathsf{E}[Y]$. Use this finding to show why independence of X and Y implies their uncorrelatedness (note that the converse does not hold).

Ex 2.5. Let U,V be two r.v.s and let $a,b \in \mathbb{R}$. Express $\mathsf{Cov}[a(U+V),b(U-V)]$ in terms of $\mathsf{V}[U]$, $\mathsf{V}[V]$ and $\mathsf{Cov}[U,V]$ (by using the expression obtained in the previous question).

Ex 2.6. Prove the key properties of covariance 1 to 5 on page 327 of the book (page 338 pdf).

Ex 2.7. Come up with a short illustrative example in which the random vector $\mathbf{X} = (X_1,\ldots,X_6)$ follows a Multinomial Distribution with parameters $n = 10$ and $\mathbf{p} = (\frac{1}{6},\ldots,\frac{1}{6}) \in \mathbb{R}^6$.

Ex 2.8. Is the following claim correct? If the r.v.s X,Y are both normally distributed, then (X,Y) follows a Bivariate Normal distribution.

Ex 2.9. Let (X,Y) follow a Bivariate Normal distribution, with X and Y marginally following $\mathcal{N}(\mu,\sigma^2)$ and with correlation ρ between X and Y.

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

Ex 2.10. Let X,Y,Z be i.i.d. $\mathcal{N}(0,1)$. Determine whether or not the random vector

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

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

2.2 *Exercises at about exam level*

Ex 2.11. Take $X \sim \text{Unif}(\{-2,-1,1,2\})$ and $Y = X^2$. What is the correlation coefficient of X and Y? If we would consider another distribution for X, would that change the correlation?

Ex 2.12. We have a machine that consists of two components. The machine works as long as both components have not failed (in other words, the machine fails when one of the two components fails). Let X_i be the lifetime of component i.

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

Ex 2.13. We have two r.v.s X and Y with the joint PDF $f_{X,Y}(x,y) = \frac{6}{7}(x+y)^2$ for $x, y \in (0,1)$ and 0 else. Also we consider the two r.v.s U and V with the joint PDF $f_{U,V}(u,v) = 2$ for $u, v \in [0,1], u+v \leq 1$ and 0 else.

1. Compute $\mathsf{P}\{X+Y>1\}$.
2. Compute $\mathsf{Cov}[U,V]$.

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

2.3 *Coding skills*

VERIFY THE ANSWERS OF BH.5.6.5 Read this example of BH first. We chop up the exercise in many small exercises..

For the python code below, run it for a small number of samples; here I choose `samples=2`. Read the print statements, and use that to answer the questions below.

```python
import numpy as np
from scipy.stats import expon

np.random.seed(10)

labda = 4
num = 3
samples = 2

```

```
10 X = expon(scale=labda).rvs((samples, num))
11 print(X)
12 T = np.sort(X, axis=1)
13 print(T)
14 print(T.mean(axis=0))
15
16 expected = np.array([labda / (num - j) for j in range(num)])
17 print(expected)
18 print(expected.cumsum())
```

```
1 set.seed(10)
2
3 labda = 4
4 num = 3
5 samples = 2
6
7 X = matrix(rexp(samples * num, rate = 1 / labda), nrow = samples, ncol = num)
8 print(X)
9 bigT = X
10 for (i in 1:samples) {
11   bigT[i,] = sort(bigT[i,])
12 }
13 print(bigT)
14 print(colMeans(bigT))
15
16 expected = rep(0, num)
17 for (j in 1:num) {
18   expected[j] = labda / (num - (j - 1))
19 }
20 print(expected)
21 print(cumsum(expected))
```

Ex 2.14. In line P.11[1] we print the value of `X` in line P.10, R.7 and R.8, respectively. What is the meaning of `X`?

Ex 2.15. What is the meaning of `T` in line P.12 (R.11)?

Ex 2.16. What do we print in line P.14, R.14?

Ex 2.17. What is meaning of the variable `expected`?

Ex 2.18. What is the `cumsum` of `expected`?

Ex 2.19. Now that you understand what is going on, rerun the simulation for a larger number of samples, e.g., 1000, and discuss the results briefly.

[1] Line P.x refers to line x of the Python code. Line R.x refers to line x of the R code.

ON BH.7.48 Read this exercise first and solve it. Then consider the code below.

```python
import numpy as np

np.random.seed(3)


def find_number_of_maxima(X):
    num_max = 0
    M = -np.infty
    for x in X:
        if x > M:
            num_max += 1
            M = x
    return num_max


num = 10
X = np.random.uniform(size=num)
print(X)

print(find_number_of_maxima(X))

samples = 100
Y = np.zeros(samples)
for i in range(samples):
    X = np.random.uniform(size=num)
    Y[i] = find_number_of_maxima(X)

print(Y.mean(), Y.var(), Y.std())
```

```r
set.seed(3)

find_number_of_maxima = function(X) {
  num_max = 0
  M = -Inf
  for (x in X) {
    if(x > M) {
      num_max = num_max + 1
      M = x
    }
  }
```

```
12   return(num_max)
13 }
14
15
16 num = 10
17 X = runif(num, min = 0, max = 1)
18 print(X)
19
20 print(find_number_of_maxima(X))
21
22 samples = 100
23 Y = rep(0, samples)
24 for (i in 1:samples) {
25   X = runif(num, min = 0, max = 1)
26   Y[i] = find_number_of_maxima(X)
27 }
28
29 print(mean(Y))
30 print(var(Y))
31 print(sd(Y))
```

Ex 2.20. Explain how the small function in lines P.6 to P.13 (R.4-R.12) works. (You should know that `x += 1` is an extremely useful abbreviation of the code `x = x + 1`).

Ex 2.21. Explain the code in lines P.25 and P.26 (R.25, R.26).

WHY IS THE EXPONENTIAL DISTRIBUTION SO IMPORTANT? At the Paris metro, a train arrives every 3 minutes on a platform. Suppose that 50 people arrive between the departure of a train and an arrival. It seems entirely reasonable to me to model the arrival times of the individual people as distributed on the interval [0,3]. What is the distribution of the inter-arrival times of these people? It turns out to be exponential!

You might want to compare your final result to Figure BH.13.1 (It is not forbidden to read the book beyond what you have to do for this course!). In this exercise we use simulation to see that clustering of arrival times.

```python
import numpy as np

np.random.seed(3)


num = 5 # small sample at first, for checking.
start, end = 0, 3
labda = num / (end - start)  # per minute
```

```
9  print(1 / labda)
10
11 A = np.sort(np.random.uniform(start, end, size=num))
12 print(A)
13 print(A[1:])
14 print(A[:-1])
15 X = A[1:] - A[:-1]
16 print(X)
17
18 print(X.mean(), X.std())
```

```
1  set.seed(3)
2
3
4  num = 5
5  start = 0
6  end = 3
7  labda = num / (end - start)
8  print(1 / labda)
9
10 A = sort(runif(num, min = start, max = end))
11 print(A)
12 print(A[-1])
13 print(A[-length(A)])
14 X = A[-1] - A[-length(A)]
15 print(X)
16
17 print(mean(X))
18 print(sd(X))
```

Ex 2.22. Explain the result of line P.12 (R.13)

Ex 2.23. Compare the result of line P.13 and P.14 (R.12, R.13); explain what is `A[1:]` (`A[-1]`)

Ex 2.24. Compare the result of line P.12 and P.14 (R.11 and R.13); explain what is `A[:-1]` (`A[-length(A)]`).

Ex 2.25. Explain what is `X` in P.15 (R.14)

Ex 2.26. Why do we compare $1/\lambda$ and `X.mean()`?

Ex 2.27. Recall that $\mathsf{E}[X] = \sigma(X)$ when $X \sim \text{Exp}(\lambda)$. Hence, what do you expect to see for `X.std()`?

Ex 2.28. Run the code for a larger sample, e.g. 50, and discuss (very briefly) your results.

2.4 *Challenges*

This exercise will given an example of how probability theory can pop up in OR problems, in particular in linear programs. It introduces you to the concept of *recourse models*, which you will learn about in the master course Optimization Under Uncertainty. Disclaimer: the story is quite lengthy, but the concepts introduced and questions asked are in fact not very hard. We just added the story to make things more intuitive.

WE CONSIDER A pastry shop that only sells one product: chocolate muffins. Every morning at 5:00 a.m., the shop owner bakes a stock of fresh muffins, which he sells during the rest of the day. Making one muffin comes at a cost of $c = \$1$ per unit. Any leftover muffins must be discarded at the end of the day, so every morning he starts with an empty stock of muffins.

The owner has one question for you: determine the amount x of muffins that he should make in the morning to minimize his production cost. Note that the owner never wants to disappoint any customer, i.e., he requires that $x \geq d$, where d is the daily demand for muffins.

The problem can be formulated as a linear program (LP):

$$\min_{x \geq 0} \{cx \ : \ x \geq d\}. \tag{2.1}$$

For simplicity, we ignore the fact that x should be integer-valued.

Ex 2.29. Determine the optimal value x^* for x and the corresponding objective value in case d is deterministic.

Of course, in practice d is not deterministic. Instead, d is a random variable with some distribution. However, note that the LP above is ill-defined if d is a random variable. We cannot guarantee that $x \geq d$ if we do not know the value of d.

You explained the issue to the shop owner and he replies: “Of course, you’re right! You know, whenever I’ve run out of muffins and a customer asks for one, I make one on the spot. I never disappoint a customer, you know! It does cost me 50% more money to produce them on the spot, though, you know.”

Mathematically speaking, the shop owner just gave you all the (mathematical) ingredients to build a so-called *recourse model*. We introduce a *recourse variable* y in our model, representing the amount of muffins produced on the spot. Production comes at a unit cost of $q = 1.5c = \$1.5$. Assuming that we know the distribution of d, we can then minimize the *expected total cost*:

$$\min_{x \geq 0} \left\{cx + \mathsf{E}\left[v(d,x)\right]\right\}, \tag{2.2}$$

where $v(d,x)$ is the optimal value of another LP, namely the *recourse problem*:

$$v(d,x) := \min_{y \geq 0} \{qy \ : \ x + y \geq d\}, \tag{2.3}$$

for given values of d and x. The recourse problem can easily be solved explicitly: we get $y = d - x$ if $d \geq x$ and $y = 0$ if $d < x$. So we obtain

$$v(d,x) = q(d-x)^+, \tag{2.4}$$

where the operator $(\cdot)^+$ represents the *positive value* operator, defined as

$$(s)^+ = \begin{cases} s & \text{if } s \geq 0, \\ 0 & \text{if } s < 0. \end{cases} \tag{2.5}$$

Ex 2.30. To get some more insight into the model, suppose (for now) that $d \sim U\{10,20\}$. Solve the model, i.e., find the optimal amount x^*. *Hint: First, compute the value of* $\mathsf{E}[v(d,x)]$ *as a function of x. Then find the optimal value of x.*

Ex 2.31. What is the expected recourse cost (expected cost of on-the-spot production) at the optimal solution x^*, i.e., compute $\mathsf{E}[v(d,x^*)]$?

To solve the model correctly, we need the true distribution of d. We learn the following from the shop owner: “My granddaughter, who’s always running around in my shop, is a bit data-crazy, you know, so she’s been collecting some data. I remember her saying that ‘the demand from male and female customers are both approximately normally distributed, with mean values both equal to 10 and standard deviations of 5’. She also mentioned something about correlation, but I don’t remember exactly, you know. It was either almost 1 or almost −1. I hope this helps!”

Mathematically, we’ve learned that $d = d_m + d_f$, with $(d_m, d_f) \sim \mathcal{N}(\mu, \Sigma)$, where $\mu = (\mu_m, \mu_f) = (10,10)$ and $\Sigma_{11} = \sigma_m^2 = \Sigma_{22} = \sigma_f^2 = 5^2 = 25$. Finally, $\Sigma_{12} = \Sigma_{21} = \mathsf{Cov}\left[d_m, d_f\right] = \rho\sigma_m\sigma_f = 25\rho$. Also, we know that either $\rho \approx 1$ or $\rho \approx -1$.

Ex 2.32. Calculate x^* and the corresponding objective value for the case $\rho = -1$. (Do not read $\rho = 1$, this case is not simple.)

Ex 2.33. Consider the two extreme cases $\rho = 1$ and $\rho = -1$. In which case will the shop owner have lower expected total costs? Provide a short, intuitive explanation. *Hint: you don’t have to compute x^* for the case where $\rho = 1$ (this is not easy!).*

3 ASSIGNMENT 3

3.1 *Have you read well?*

Ex 3.1. Let X have the discrete uniform distribution on the set $\{0,1,2,3,4,5\}$. Derive the PMF and the CDF of $Z = 3X$. Explicitly specify the domain.

Ex 3.2. Let $X \sim \text{Unif}(0,5)$. Using the one dimensional change of variables theorem (BH.8.1.1), derive the PDF and the CDF of $Z = 3X$. Explicitly specify the domain.

Ex 3.3. Let $X \sim \text{Norm}(\mu,\sigma^2)$. Using the one dimensional change of variables theorem BH.8.1.1, show that $Z = \frac{X-\mu}{\sigma}$ follows the standard normal distribution.

Ex 3.4. Let $X \sim \text{Exp}(1)$. Derive the PDF of e^{-X}.

Ex 3.5. Let X,Y be i.i.d. standard normal. Using the n-dimensional change of variables theorem, derive the joint PDF of $(X+Y, X-Y)$.

Check your final answer using BH.7.5.8.

Ex 3.6. Specify the domain of the new random variable for the following transformations; this important aspect of the change of variables is often overlooked. Let U, V, W, X, X_1, X_2, Y and Z be r.v.s and let a, b and c be arbitrary constants.

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

Ex 3.7. To find the distribution of a convolution through the change of variables formula, we seem to need to add a 'redundant' equality? But why is that? What would be the problem if we do not add this? Explain in your own words.

Ex 3.8. When adding a different equality, we need to be careful to not create a functional relationship between our two new variables U,V, for example $U = X+Y$ and $V = \sin(X+Y)$, or $U = \frac{X}{Y}$ and $V = \frac{Y}{X}$ for conforming X,Y. What would happen to the determinant of the Jacobian matrix if we did? Why would this happen? Explain in your own words.

Ex 3.9. In this exercise, we combine what we learned in BH.8.1.4 and BH.8.1.9. Let S be the sum of two i.i.d. chi-square distributed variables. Using just these two examples, show that $S \sim \text{Exp}(2)$.

Ex 3.10. A student has obtained an i.i.d. random sample of size 2 from a Cauchy distribution. Let the r.v.s X and Y model the values of the first and second sample. Since s/he does not know what the mean of a Cauchy distribution is, s/he wants to average the sample to obtain what she thinks is a good estimate of the true mean.

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

1. Find an expression for $f_W(w)$ in the form of an integral, but do not solve it.
2. It turns out that if we solve the integral, we get that $f_W(w) = f_X(w)$. The distribution of our sample mean is still Cauchy; we did not obtain a better estimate of the Cauchy mean by calculating the sample mean!

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

3.2 *About exam level*

Ex 3.11. Let X, Y iid $\sim \text{Unif}([0,1])$.

1. Write a computer program in python or R to estimate $\mathsf{P}\{X + Y \leq 1, XY \leq 2/9\}$.

Ex 3.12. Let X, Y be continuous r.v.s with CDF $F_{X,Y}(x, y) = (x-1)^2(y-2)/8$ for $x \in (1,3)$.

a. Explain that $y \in (2,4)$ for F to be a proper CDF.

b. What is $F(3,7)$?

c. Determine the PDF.

d. Compute $\mathsf{P}\{2 < X < 3\}$

e. Compute $\mathsf{P}\{2 < X < 3, 2 < Y < 3\}$.

f. Compute $\mathsf{P}\{Y < 2X\}$.

g. Compute $\mathsf{P}\{Y \leq 2X\}$.

h. Compute $\mathsf{P}\{Y < 2X, Y + 2X > 6\}$.

Ex 3.13. Consider the general case where we are given the relationship $U = V^4$ between the random variables U and V for $V \in (-3,2)$.

Explain why we cannot simply invoke the change of variables theorem.

Now imagine V following a uniform distribution on the given interval. Consider the given transformation on the intervals $(-3,0)$ and $(0,2)$ separately. Explain why this allows you to employ the change of variables theorem and find the distribution of U on these intervals. Finally combine these results (using indicator functions) and state the PDF of U (remember to adjust the domain for the indicator functions according to the transformations).

Ex 3.14. Let $U \sim \text{Unif}(0,\pi)$. Use BH.8.1.9 to show that $X = \tan(U)$ has the Cauchy distribution. Compare this exercise to BH.8.1.5.

3.3 *Coding skills*

PING PONG BALLS How many ping pong balls fit into an Airbus Beluga? One way to answer this is as follows. According to this wiki-page the cargo volume V of this airplane is $1500\,\text{m}^3$. But this number is based on the physical dimensions that is available to store containers, tanks, and so on. So, I estimate the volume as about twice that amount, i.e., $V = 2500\,\text{m}^3$. The volume of a ping pong ball is $v = 4\pi r^3/3 = 33.4933333333333\,\text{cm}^3$ with $r = 2$ cm. A plain division gives 74.6268656716418 ping pong balls. Note, I left out the 10^6 conversion from meters to cm, and I do not take into the sphere packing factor. Besides that, I hope you agree with me that providing an result with the precision as given here is plain ridiculous. (But from reason incomprehensible to me, even professional econometricians like to report results with 10 digits or more, without questioning the precision.)

However, I know that the volumes of an air plane and a ping pong ball is an estimate, rather than a precise number as assumed above. It seems to be better to approximate V and v as rvs. Let's assume that

$$V \sim N(2500, 500^2), \qquad v \sim N(33.5, 0.5^2),$$

where the variances express my trust in my guess work. What is now the mean of $N = V/v$ and its std? In fact, finding the closed form expression for the distribution of N is not entirely simple. However, with simulation it's easy to get an estimate.

Ex 3.15. How does this exercise relate to BH.8.11 and BH.8.12? What is similar, what are crucial differences?

Ex 3.16. Use the documentation of the `norm` (`rnorm`) function of python (R) to explain why we set the scale as we do. Relate this to location-scale discussion in BH.

Ex 3.17. Explain lines 8 and 13 of the python code or lines 5 and 8 of the R code. (Optional: There is a conceptual difference between the two languages here. If you are interested in both languages, also comment on the difference)

Ex 3.18. Use the code below to provide the estimates.

Ex 3.19. Contrary to BH.7.1.25 if you run the code below, you'll see that $\mathsf{E}[N] < \infty$, and, in fact, very near to the deterministic answer. But isn't this strange? We divide two normal random variables, just like BH.7.1.25, but there the expectation is infinite. Comment on the difference.

The numerical results suggest the interesting guess $\mathsf{V}[N] \approx \mathsf{V}[V] * \mathsf{V}[v]$, but is this true more generally? In Section 3.4 we study this problem in more detail.

```python
import numpy as np
from scipy.stats import norm

num = 500

np.random.seed(3)

V = norm(loc=2500, scale=500)
v = norm(loc=33.5, scale=0.5)

print(V.mean(), V.std()) # just a check

N = V.rvs(num) / v.rvs(num)
print(N.mean(), N.std())

print(2500/33.5)
print(np.sqrt(500*0.5))
```

```r
num <- 500

set.seed(3)

V = rnorm(num, 2500, 500)
v = rnorm(num, 33.5, 0.5)

N = V / v
paste(mean(N), sd(N))

2500/33.5
sqrt(500*0.5)
```

SUMS OF RVS We start from BH.8.27 (which you have to read now). We are interested in the difference between the distribution of $X + Y + Z$ and the normal distribution. But why the normal distribution? As it turns out, the central limit law, see BH.10, states that the distribution of sums of r.v.s converge to the normal distribution (in a specific sense)

Here some code to simulate.

```python
import numpy as np
from scipy.stats import norm
```

```python

import matplotlib.pylab as plt
import seaborn as sns

sns.set()

np.random.seed(3)

k = 3
Zexact = norm(loc=k / 2, scale=np.sqrt(k / 12))
X = np.arange(0, 3, 0.1)

XYZ = np.random.uniform(size=(4000, k))
# print(XYZ)  # if you want to see it.
Z = XYZ.sum(axis=1)
sns.distplot(Z)
plt.plot(X, Zexact.pdf(X))
plt.show()
```

```r
set.seed(3)

k = 3
X <- seq(0, 3, by = 0.1)
Zexact <- dnorm(X, mean = k / 2, sd = sqrt(k / 12))


XYZ <- matrix(NA, 4000, k)
for (i in 1:k) {
  XYZ[,i] <- runif(4000, min = 0, max = 1)
}
Z <- rowSums(XYZ)

par()
hist(Z, prob = TRUE, breaks = 31)
lines(X, Zexact, type = "l", col = "orange")
lines(density(Z), col = "blue")
```

Ex 3.20. What is the shape of XYZ in the code above, i.e., how many rows and columns does it have? If you don't know, run the code, and print it.

Ex 3.21. What is the shape (rows and columns) of Z?

Ex 3.22. Explain the values for `loc` and `shape` in `Zexact`. (Read the documentation of scipy.stats.norm on the web is necessary.) To which definition in BH does this loc-scale transformation relate?

Ex 3.23. Change the seed to your student id, or any other number you like, run the code, and include the graph produced by your simulation. Explain what you see.

Now we do an exact computation.

```python
import numpy as np
from scipy.stats import norm

import matplotlib.pylab as plt
import seaborn as sns

sns.set()

N = 200
x = np.linspace(0, 2, 2 * N)
fx = np.ones(N) / N
f2 = np.convolve(fx, fx)
f3 = np.convolve(f2, fx)

k = 3

x = np.linspace(0, k, len(f3))
Zexact = norm(loc=k / 2, scale=np.sqrt(k / 12))


plt.plot(x, N * f3, label="conv")
plt.plot(x, Zexact.pdf(x))
plt.legend()
plt.show()
```

```r
N = 200
x = seq(0, 2, length.out = 2 * N)
fx = rep(1, N) / N
f2 = convolve(fx, fx, type = "open")
f3 = convolve(f2, fx, type = "open")

k = 3

```

```r
x = seq(0, k, length.out = length(f3))
Zexact = dnorm(x, mean = k/2, sd = sqrt(k / 12))

par()
plot(x, N * f3, col = "blue", type = "l", ylim = c(0, 0.8))
lines(x, Zexact, type = "l", col = "orange")
legend("topright", legend = "conv", bty = "n",
       lwd = 2, cex = 1.2, col = "blue", lty = 1)
```

Ex 3.24. Read the documentation of np.convolve . Why is it called like this?

Ex 3.25. In the code, what is `f2`?

Ex 3.26. What is `f3`?

Ex 3.27. Why do we set `k=3`?

Ex 3.28. A bit harder, why do we plot `N*f3`, i.e., why do have to multiply with `N`? Relate this to the meaning of $f(x)\,dx$, where f the density of some random variable. (To understand why is very important. Think hard, and read the solution when it becomes available.)

Ex 3.29. Yet a tiny bit harder, consider `f4 = np.convolve(f3, fx)` and `g4 = np.convolve(f2, f2)`. Why are they, numerically speaking, equal?

Ex 3.30. When you would compute the maximum of `np.abs(f4 -g4)` you would see that this is about 10^{-10}, or so. Hence, a small number. This is not equal to 0, but we know that this is due to rounding effects.

How can we use the function `np.isclose()` to get around this problem? (You should memorize from this question that you should take care when testing on whether floating point numbers are the same or not.)

3.4 *Challenges*

This challenge is a continuation of the Beluga case of Section 3.3, and we discuss some ways to check whether $\mathsf{V}[N] \approx \mathsf{V}[V]\,\mathsf{V}[v]$ holds in general, and then we try to find a better approximation. We chopped up the challenge into many exercises, to help you organize the ideas.

Recall that in Section 3.3 we have been a bit sloppy about the units, measuring the volumes of the airplane in m^3 and a ping pong ball in cm^3, so actually N is in millions of ping pong balls. Note that using different units can easily lead to confusion; as a take-away , choose one unit.

One way to check the correctness of $\mathsf{V}[N] \approx \mathsf{V}[V]\,\mathsf{V}[v]$ is to change the scale. In fact, memorize that changing scale is an easy way to check laws.

Ex 3.31. Suppose we instead measure the size of a ping pong ball in meters and the size of the airplane in hectometers. Explain that N is still in millions of ping pong balls. What happens to $\mathsf{V}[N]$ and what happens to $\mathsf{V}[V]\,\mathsf{V}[v]$ (theoretically)?

Another way to check a statement is to consider some extreme cases.

Ex 3.32. Suppose that we would know the size of a ping pong ball very accurately, i.e. we consider the extreme case where $\mathsf{V}[v] \to 0$. Explain that the approximation is not a good approximation in this limit.

Ex 3.33. Which of these two checks convinces you most that something is wrong with this approximation, and why?

We now turn to the task of trying to find a good approximation.

Ex 3.34. Assume that X and Y are independent. Show that

$$\mathsf{V}[XY] = \mathsf{V}[X]\,\mathsf{V}[Y] + \mathsf{V}[X]\,\mathsf{E}[Y]^2 + \mathsf{E}[X]^2\,\mathsf{V}[Y].$$

Ex 3.35. Assume in addition that we know at least one of X and Y quite precisely. Argue that the following is then a good approximation:

$$\mathsf{V}[XY] \approx \mathsf{V}[X]\,\mathsf{E}[Y]^2 + \mathsf{E}[X]^2\,\mathsf{V}[Y].$$

So far we have only considered the variance of a product, but we would like to know the variance of a ratio. For this we can use Taylor expansions to make accurate approximations.

Ex 3.36. Find the first order Taylor expansion of $\frac{1}{Z}$ around $a = \mathsf{E}[Z]$. By taking the expectation and the variance of this expansion, show that

$$\mathsf{E}\left[\frac{1}{Z}\right] \approx \frac{1}{\mathsf{E}[Z]}, \qquad \mathsf{V}\left[\frac{1}{Z}\right] \approx \frac{\mathsf{V}[Z]}{\mathsf{E}[Z]^4}.$$

Ex 3.37. Combine all of the above to derive the following approximation for the variance of the ratio of two independent random variables X and Z:

$$\mathsf{V}\left[\frac{X}{Z}\right] \approx \frac{\mathsf{V}[X]}{\mathsf{E}[Z]^2} + \mathsf{E}[X]^2\,\frac{\mathsf{V}[Z]}{\mathsf{E}[Z]^4}.$$

Ex 3.38. Check this approximation in the ways of the first two exercises.

After doing all this work, we would of course like to know how well this approximation does. When comparing the approximation to the sample standard deviation found in [3.15] for `num=500`, the result may be a bit disappointing. However, this is just because the sample standard deviation is also an estimate of the actual standard deviation of N, so by chance the result may be closer to $\mathsf{V}[V]\,\mathsf{V}[v]$ than to our new approximation.

In Chapter 10, you will learn something about the distribution of the sample variance. For now, just increase `num`. We know this decreases the variance of the sample mean and it also decreases the variance of the sample variance, so we get a more accurate estimate.

Ex 3.39. Use the result of the previous exercise to compute an approximation for $\mathsf{V}[N] = \mathsf{V}[V/v]$. Also use the code with a (much) higher value of `num`, to show that the approximation $\mathsf{V}[N] \approx \mathsf{V}[V]\,\mathsf{V}[v]$ is likely to be worse, even in the setting of [3.15] where it was quite good.

The following two exercises are really optional, but I found them very neat and insightful.

Ex 3.40. Recall that for a non-negative random variable X with finite variance, we define the squared coefficient of variation as $\text{SCV}(X) = \mathsf{V}[X]/\mathsf{E}[X]^2$. Using the SCV, show that the approximations of [3.35] and [3.36] can be rewritten in the following neat way:

$$\begin{aligned}
&\text{SCV}(XY) \approx \text{SCV}(X) + \text{SCV}(Y).\\
&\text{SCV}(1/Z) \approx \text{SCV}(Z).
\end{aligned}$$

In BH.10, you will learn Jensen's inequality, which implies that $\mathsf{E}\left[\frac{1}{Z}\right] \geq \frac{1}{\mathsf{E}[Z]}$ for all positive random variables Z. In the following exercise, we reflect on this by finding a more accurate approximation based on the second order Taylor expansion.

Ex 3.41. Find the second order Taylor expansion of $\frac{1}{Z}$ around $a = \mathsf{E}[Z]$. By taking the expectation, show that

$$\mathsf{E}\left[\frac{1}{Z}\right] \approx \frac{1}{\mathsf{E}[Z]} + \frac{2\mathsf{V}[Z]}{\mathsf{E}[Z]^3}.$$

Note that this is always at least $\dfrac{1}{\mathsf{E}[Z]}$.

4 HINTS

h.1.8. From this exercise you should memorize this: **independence is a property of the joint CDF, not of the rvs**.

h.1.15. In this exercise we want to prove that N is Poisson distributed. So you cannot assume this in your solution.

h.1.17. Use the relation of the previous exercise to show that

$$P(N=n+1) = \frac{\lambda}{1+n} P(N=n). \tag{4.1}$$

Bigger hint: Fill in $y=0$ in the LHS and RHS of (1.1); call this expression 1. Then fill in $y=1$ to a obtain a second expression. Divide these two expressions and note that $\mathsf{P}\{X=x\}$ cancels. Finally, define

$$\lambda = \frac{\mathsf{P}\{Y=1\}}{(1-p)\mathsf{P}\{Y=0\}}. \tag{4.2}$$

h.3.9. Let X, Y be i.i.d. standard normal. Since the square of a standard normal r.v. is chi-square distributed, we can write S as $S = X^2+Y^2$ (here we use BH.8.1.4).

h.3.13. What is the domain of V on each of the intervals $(-3,0)$ and $[0,2)$? For the final part, combining the results into one PDF: Use LOTP, conditioning on $U \geq 0$.

5 SOLUTIONS

Compare your answers very carefully against ours. You should spend time thinking about the definition and notation we use. For instance, there is conceptual huge difference between X and x. More generally, good notation and good understanding correlate (positively).

s.1.1. Check the definitions of the book.

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

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

s.1.2.

$$f_X(x) = \int_0^1 f_{X,Y}(x,y)\,\mathrm{d}y = 2\int_0^1 I_{x\leq y}\,\mathrm{d}y = 2\int_x^1 \mathrm{d}y = 2(1-x) \tag{5.1}$$

$$f_Y(y) = \int_0^1 f_{X,Y}(x,y)\,\mathrm{d}x = 2\int_0^1 I_{x\leq y}\,\mathrm{d}x = 2\int_0^y \mathrm{d}y = 2y. \tag{5.2}$$

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

$$F_{X,Y}(x,y) = \int_0^x \int_0^y f_{X,Y}(u,v)\,\mathrm{d}v\,\mathrm{d}u \tag{5.3}$$

$$= 2\int_0^x \int_0^y I_{u\leq v}\,\mathrm{d}v\,\mathrm{d}u \tag{5.4}$$

$$= 2\int_0^x \int I_{u\leq v} I_{0\leq v\leq y}\,\mathrm{d}v\,\mathrm{d}u \tag{5.5}$$

$$= 2\int_0^x \int I_{u\leq v\leq y}\,\mathrm{d}v\,\mathrm{d}u \tag{5.6}$$

$$= 2\int_0^x [y-u]^+\,\mathrm{d}u, \tag{5.7}$$

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

$$2\int_0^x [y-u]^+\,\mathrm{d}u = 2\int_0^x (y-u)\,\mathrm{d}u = 2yx - x^2, \tag{5.8}$$

while if $y \leq x$,

$$2\int_0^x [y-u]^+\,\mathrm{d}u = 2\int_0^y (y-u)\,\mathrm{d}u = 2y^2 - y^2 = y^2 \tag{5.9}$$

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

s.1.3.

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

s.1.4.

$$\frac{F_{X,Y}(x,y)}{F_X(x)} = \frac{\mathsf{P}\{X \le x, Y \le y\}}{\mathsf{P}\{X \le x\}} \tag{5.10}$$

In the notes we define the conditional CDF as the function $F_{X|Y}(x|y) = \mathsf{P}\{X \le x | Y = y\}$. This is not the same as the function above.

Mistake: $F_{X,Y}(x,y) \neq \mathsf{P}\{X = x, Y = y\}$. If you wrote this, recheck BH. for the conditional CDF, you do not condition on e.g. $X \le x$. Compare your answer to what is written in the notes or the solution manual. Good notation and good understanding are positively correlated :).

s.1.5. $\mathsf{P}\{X = 0, Y = 0\} = 1/3 \cdot 3/4$, $\mathsf{P}\{X = 0, Y = 1\} = 1/3 \cdot 1/4$, and so on.

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

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

s.1.6. The number of produced items (laid eggs) is N. The probability of hatching is p, that is, an item is ok. The hatched eggs are the good items.

s.1.7. For X, Y to be independent, it is necessary that $F_{X,Y}(x,y) = F_X(x)F_Y(y)$ for all x, y, not just one particular choice. (This is an example that satisfying a necessary condition is not necessarily sufficient.)

s.1.8. Many answers are possible here, depending on extra assumptions you make. Here is one. Suppose, just by change, the fraction of taller guys in the street is a bit higher the population fraction. Assuming that taller (shorter) people prefer taller (shorter) spouses, there must be dependence between the height of the men and the woman. This is because when selecting a man, I can also select his wife.

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

s.1.9. Only when X, Y are independent.

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

s.1.10. Given $N = n$, the random variable X has a certain distribution, binomial for instance.

s.1.11. This claim is incorrect, because X, Y are discrete, hence they have a PMF, not a PDF.

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

s.1.12. This question tests your modeling skills too.

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

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

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

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

$$\mathsf{P}\{I_0 <= 0\} = \mathsf{P}\{2 + X_1 - C_1 < 0\} = \frac{1}{10}\sum_{i=1}^{10} \mathsf{P}\{C_1 > 2 + i\} = \frac{1}{10}\sum_{i=1}^{5} \mathsf{P}\{C_1 > 2 + i\} \tag{5.11}$$

$$= \frac{1}{10}\sum_{i=1}^{5} \frac{6-i}{9}. \tag{5.12}$$

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

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

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



I0 = 2

I1 = {}
for k, p in X.items():
    for l, q in C.items():
        i = I0 + k - l
```

```
        I1[i] = I1.get(i, 0) + p * q

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


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


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

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

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

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

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

s.1.14. Mistakes: Simulation and numerical integration are not the same. Formulate your answers precisely: it is not simulation that yields exactly the same value!

s.2.1. We have

$$\begin{aligned} \mathsf{E}\left[(X-Y)^2\right] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-y)^2 f_{X,Y}(x,y)dxdy && (5.13)\\ &= \int_0^1\int_0^1 (x-y)^2 dxdy && (5.14)\end{aligned}$$

s.2.2. Any situation in which the units of measurement might be distracting. Correlation is usually easier to interpret.

s.2.3. Answers: no and yes.

We have

$$C = \frac{\mathsf{V}[X]}{(\mathsf{E}[X])^2}, \qquad (5.15)$$

which does not equal

$$\operatorname{Corr}(X,X) = \frac{\mathsf{Cov}[X,X]}{\sqrt{\mathsf{V}[X]\,\mathsf{V}[X]}} = 1 \tag{5.16}$$

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

$$C = 100/(1^2) = 100 > 1. \tag{5.17}$$

s.2.4. We have

$$\begin{aligned}
\mathsf{Cov}[X,Y] &= \mathsf{E}[(X - \mathsf{E}[X])(Y - \mathsf{E}[Y])] && (5.18)\\
&= \mathsf{E}[XY - X\,\mathsf{E}[Y] - Y\,\mathsf{E}[X] + \mathsf{E}[X]\,\mathsf{E}[Y]] && (5.19)\\
&= \mathsf{E}[XY] - \mathsf{E}[X]\,\mathsf{E}[Y] - \mathsf{E}[Y]\,\mathsf{E}[X] + \mathsf{E}[X]\,\mathsf{E}[Y] && (5.20)\\
&= \mathsf{E}[XY] - \mathsf{E}[X]\,\mathsf{E}[Y]. && (5.21)
\end{aligned}$$

s.2.5. By linearity of the covariance wea have

$$\begin{aligned}
\mathsf{Cov}[a(U+V), b(U-V)] &= a\Big(\mathsf{Cov}[U, b(U-V)] + \mathsf{Cov}[V, b(U-V)]\Big) && (5.22)\\
&= a\Big(b\big(\mathsf{Cov}[U,U] - \mathsf{Cov}[U,V]\big) + b\big(\mathsf{Cov}[V,U] - \mathsf{Cov}[V,V]\Big) && (5.23)\\
&= a\Big(b\big(\mathsf{Cov}[U,U] - \mathsf{Cov}[U,V]\big) + b\big(\mathsf{Cov}[V,U] - \mathsf{Cov}[V,V]\Big) && (5.24)\\
&= ab\Big(\mathsf{V}[U] - \mathsf{Cov}[U,V] + \mathsf{Cov}[V,U] - \mathsf{V}[V]\Big) && (5.25)\\
&= ab\Big(\mathsf{V}[U] - \mathsf{V}[V]\Big). && (5.26)
\end{aligned}$$

s.2.6. 1. We have

$$\mathsf{Cov}[X,X] = \mathsf{E}[XX] - \mathsf{E}[X]\,\mathsf{E}[X] = \mathsf{E}\left[X^2\right] - \mathsf{E}[X]^2 = \mathsf{V}[X]. \tag{5.27}$$

2. We have

$$\mathsf{Cov}[X,Y] = \mathsf{E}[XY] - \mathsf{E}[X]\,\mathsf{E}[Y] = \mathsf{E}[YX] - \mathsf{E}[Y]\,\mathsf{E}[X] = \mathsf{Cov}[Y,X]. \tag{5.28}$$

3. We have

$$\mathsf{Cov}[X,c] = \mathsf{E}[Xc] - \mathsf{E}[X]\,\mathsf{E}[c] = c\,\mathsf{E}[X] - c\,\mathsf{E}[X] = 0. \tag{5.29}$$

4. We have

$$\mathsf{Cov}[aX,Y] = \mathsf{E}[aXY] - \mathsf{E}[aX]\,\mathsf{E}[Y] = a\big(\mathsf{E}[XY] - \mathsf{E}[X]\,\mathsf{E}[Y]\big) = a\,\mathsf{Cov}[X,Y]. \tag{5.30}$$

5. We have

$$\begin{aligned}\mathsf{Cov}[X+Y,Z] &= \mathsf{E}[(X+Y)Z] - \mathsf{E}[X+Y]\,\mathsf{E}[Z] &(5.31)\\ &= \mathsf{E}[XZ+YZ] - \big(\mathsf{E}[X]+\mathsf{E}[Y]\big)\mathsf{E}[Z] &(5.32)\\ &= \mathsf{E}[XZ] - \mathsf{E}[X]\,\mathsf{E}[Z] + \mathsf{E}[YZ] - \mathsf{E}[Y]\,\mathsf{E}[Z] &(5.33)\\ &= \mathsf{Cov}[X,Z] + \mathsf{Cov}[Y,Z]. &(5.34)\end{aligned}$$

s.2.7. We throw 10 fair dice. X_i denotes the number of dice that show the number i, $i = 1,\dots,6$.

s.2.8. No, this does not always hold. It does hold when X and Y are independent, though.

s.2.9. In hindsight, this question was more an exam-level question.

1. Since (X,Y) are bivariate normally distributed, every linear combination of X and Y is normally distributed. Note that every linear combination of $(X+Y)$ and $(X-Y)$ can be written as a linear combination of X and Y. Hence, every linear combination of $(X+Y)$ and $(X-Y)$ is normally distributed. Hence, $(X+Y, X-Y)$ is bivariate normally distributed.

2. By the story above, both X and Y are normally distributed. We have

$$\mathsf{E}[X+Y] = \mathsf{E}[X] + \mathsf{E}[Y] = \mu + \mu = 2\mu, \tag{5.35}$$

and

$$\mathsf{E}[X-Y] = \mathsf{E}[X] - \mathsf{E}[Y] = \mu - \mu = 0. \tag{5.36}$$

Moreover,

$$\mathsf{V}[X+Y] = \mathsf{V}[X] + \mathsf{V}[Y] + 2\,\mathsf{Cov}[X,Y] = 2\sigma^2 + 2\rho\sigma^2 = 2(1+\rho)\sigma^2. \tag{5.37}$$

Simlarly,

$$\begin{aligned}\mathsf{V}[X-Y] = \mathsf{V}[X] + \mathsf{V}[-Y] + 2\,\mathsf{Cov}[X,-Y] &= \mathsf{V}[X] + \mathsf{V}[Y] - 2\,\mathsf{Cov}[X,Y] &(5.38)\\ &= 2\sigma^2 - 2\rho\sigma^2 = 2(1-\rho)\sigma^2. &(5.39)\end{aligned}$$

So we have found that $X+Y \sim N(2\mu, 2(1+\rho)\sigma^2)$ and $X-Y \sim N(0, 2(1-\rho)\sigma^2$.

3. We have

$$\mathsf{Cov}[X+Y, X-Y] = \mathsf{Cov}[X,X] - \mathsf{Cov}[X,Y] + \mathsf{Cov}[Y,X] - \mathsf{Cov}[Y,Y] = \mathsf{V}[X] - \mathsf{V}[Y] = \sigma^2 - \sigma^2 = 0. \tag{5.40}$$

Write $U = X+Y$, $V = X-Y$. Plugging all the parameters into the formula for the joint pdf of a bivariate normal distribution (see https://en.wikipedia.org/wiki/Multivariate_normal_distribution#Bivariate_case), we obtain

$$f_{U,V}(u,v) = \frac{1}{2\pi\sqrt{2(1+\rho)\sigma^2 2(1-\rho)\sigma^2}} \exp\left(-\frac{1}{2}\left[\frac{(u-2\mu)^2}{2(1+\rho)\sigma^2} + \frac{v^2}{2(1-\rho)\sigma^2}\right]\right). \tag{5.41}$$

s.2.10. Since X, Y, Z are independent normally distributed variables, (X, Y, Z) is multivariate normally distributed. Hence, every linear combination of X, Y, Z is normally distributed. Note that every linear combination of the elements of W can be written as a linear combination of X, Y, Z. Hence, every linear combination of the elements of W is normally distributed. Hence, W is multivariate normally distributed.

s.2.11. We have

$$\mathsf{Cov}[X,Y] = \mathsf{Cov}\left[X,X^2\right] = \mathsf{E}\left[XX^2\right] - \mathsf{E}[X]\,\mathsf{E}\left[X^2\right] = 0 - 0\cdot 2.5 = 0. \tag{5.42}$$

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

$$\mathsf{Cov}[X,Y] = \mathsf{E}\left[XX^2\right] - \mathsf{E}[X]\,\mathsf{E}\left[X^2\right] = 0.5 - 0.5\cdot 0.5 = 0.25. \tag{5.43}$$

s.2.12.

1. The interpretation is: the time until the first component fails. That is, the time until the machine stops working.

2. Let $\lambda = 10$. We have

$$\begin{aligned}
\mathsf{P}\{\text{machine not failed at time } T\} &= \mathsf{P}\{\min\{X_1,X_2\} > T\} && (5.44)\\
&= \mathsf{P}\{X_1 > T, X_2 > T\} && (5.45)\\
&= \mathsf{P}\{X_1 > T\}\,\mathsf{P}\{X_2 > T\} && (5.46)\\
&= e^{-\lambda T}\cdot e^{-\lambda T} && (5.47)\\
&= e^{-(2\lambda)T} && (5.48)\\
&= e^{-20T} && (5.49)\\
& && (5.50)
\end{aligned}$$

3. Note that

$$\mathsf{P}\{\min\{X_1,X_2\} \le T\} = 1 - \mathsf{P}\{\min\{X_1,X_2\} > T\} = 1 - e^{-20T}. \tag{5.51}$$

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

4. The expected time until the machine fails is

$$\mathsf{E}[\min\{X_1,X_2\}] = 1/20, \tag{5.52}$$

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

s.2.13. 1. We have

$$\begin{aligned}
\mathsf{P}\{X+Y>1\} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_{X+Y>1} f_{X,Y}(x,y)\,dy\,dx && (5.53)\\
&= \int_0^1 \int_{1-x}^1 \frac{6}{7}(x+y)^2\,dy\,dx && (5.54)\\
&= \frac{6}{7}\int_0^1 \left[\frac{1}{3}(x+y)^3\right]_{y=1-x}^1 dx && (5.55)\\
&= \frac{2}{7}\int_0^1 \Big((x+1)^3-(x+1-x)^3\Big)dx && (5.56)\\
&= \frac{2}{7}\int_0^1 \Big((x+1)^3-1\Big)dx && (5.57)\\
&= \frac{2}{7}\left[\frac{1}{4}(x+1)^4-x\right]_{x=0}^1 && (5.58)\\
&= \frac{1}{14}\left[(x+1)^4-4x\right]_{x=0}^1 && (5.59)\\
&= \frac{1}{14}\Big(\big((1+1)^4-4\big)-\big((0+1)^4-0\big)\Big) && (5.60)\\
&= \frac{1}{14}\Big(16-4-1\Big) && (5.61)\\
&= \frac{11}{14}. && (5.62)
\end{aligned}$$

2. We have

$$\mathsf{Cov}[U,V] = \mathsf{E}[UV] - \mathsf{E}[U]\,\mathsf{E}[V]. \qquad (5.63)$$

First, we compute

$$\begin{aligned}
\mathsf{E}[UV] &= \int_0^1 \int_0^{1-u} 2uv\,dv\,du && (5.64)\\
&= \int_0^1 [uv^2]_{v=0}^{1-u}\,du && (5.65)\\
&= \int_0^1 \big(u(1-u)^2-0\big)du && (5.66)\\
&= \int_0^1 u(1-2u+u^2)\,du && (5.67)\\
&= \int_0^1 (u-2u^2+u^3)\,du && (5.68)\\
&= \left[\frac{1}{2}u^2-\frac{2}{3}u^3+\frac{1}{4}u^4\right]_{u=0}^1 && (5.69)\\
&= \frac{1}{2}-\frac{2}{3}+\frac{1}{4} && (5.70)\\
&= \frac{1}{12}. && (5.71)
\end{aligned}$$

Next,

$$\begin{aligned}
\mathsf{E}[U] &= \int_0^1 \int_0^{1-u} 2u\,dv\,du && (5.72)\\
&= \int_0^1 2u \int_0^{1-u} 1\,dv\,du && (5.73)\\
&= \int_0^1 2u(1-u)\,du && (5.74)\\
&= 2\int_0^1 (u-u^2)\,du && (5.75)\\
&= 2\Big[\frac{1}{2}u^2 - \frac{1}{3}u^3\Big]_{u=0}^1 && (5.76)\\
&= 2\Big(\frac{1}{2}-\frac{1}{3}\Big) && (5.77)\\
&= \frac{1}{3} && (5.78)
\end{aligned}$$

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

$$\begin{aligned}
\mathsf{Cov}[U,V] &= \mathsf{E}[UV] - \mathsf{E}[U]\,\mathsf{E}[V] && (5.79)\\
&= \frac{1}{12} - \frac{1}{3}\frac{1}{3} && (5.80)\\
&= \frac{1}{12} - \frac{1}{9} && (5.81)\\
&= -\frac{1}{36}. && (5.82)
\end{aligned}$$

s.2.14. X is a matrix of i.i.d. draws from an exponential distribution with parameter λ.

s.2.15. T is a sorted version of X, where we sort each row increasingly.

s.2.16. We print the mean value of each column of T.

s.2.17. This is an array with expected values of the ith order statistic $X_{(i)}$ (see B.H.5.6.5 afor a proof of this result).

s.2.18. The cumsum is the cumulative sum up to and including the current index. So the final entry indicates the expected value of the sum of all three entries of `expected`.

s.2.19. The result of `print(T.mean(axis=0))` should be close to that of `print(expected.cumsum())`.

s.2.20. It iterates through the elements of X and checks how often the current value is larger than any of the previous values.

s.2.21. We draw a sample of a $U[0,1]$ distribution of size `num` and compute the corresponding number of maxima (or “records”).

s.2.22. This are the arrival times of 5 passengers within the time interval of 3 minutes (sorted increasingly).

s.2.23. `A[1:]` is an array of all elements of `A` except the first one.

s.2.24. `A[:-1]` is an array of all elements of `A` except the last one.

s.2.25. `X` consists of the interarrival times.

s.2.26. $1/\lambda$ is the expected interarrival time. `X.mean()` is the sample average of the interarrival times.

s.2.27. For `X.std()` we expect to see $1/\lambda = 0.6$ too (if X is indeed exponentially distributed with parameter λ).

s.2.28. For a sample of size 50, we expect an average interarrival time of 0.06 and an equal standard deviation if the distribution of the interarrival times is indeed exponential. We indeed observe a sample mean and sample average that are very close to this value.

s.2.29. If d is deterministic and known then $x^* = d$, since cost in increasing in x.

s.2.30. We know that at least 10 muffins will be needed so we stock at least 10 muffins. If we stock less, then we know for sure that we need to bake an additional muffin on the spot which increases cost. Also, we never need more than 20 muffins. So $10 \le x^* \le 20$. Note that

$$\begin{aligned} E[v(d,x)] &= E\left[q(d-x)^+\right] = q\int_{10}^{20}(d-x)^+\tfrac{1}{10}\mathrm{d}d = q\int_{x}^{20}(d-x)\tfrac{1}{10}\mathrm{d}d \\ &= \tfrac{q}{10}\left[\tfrac{1}{2}(d-x)^2\right]_x^{20} = \tfrac{q}{20}(20-x)^2. \end{aligned}$$

Hence, $cx + E[v(d,x)] = cx + \frac{q}{20}(20-x)^2 = x + 0.075(20-x)^2 = 30 - 2x + 0.075x^2$.
Setting the derivative $-2 + 0.15x$ to 0 yields $x^* = \frac{2}{0.15} = \frac{40}{3}$.

s.2.31. $E[v(d,x^*)] = \frac{q}{20}(20-x^*)^2 = \frac{\$1.5}{20}\left(\frac{20}{3}\right)^2 = \$\frac{10}{3}$.

s.2.32. The sum $d_m + d_f$ is again normally distributed with mean $\mathsf{E}\left[d_m + d_f\right] = \mathsf{E}[d_m] + \mathsf{E}\left[d_f\right] = 20$ and variance $\mathsf{V}\left[d_m + d_f\right] = \mathsf{V}[d_m] + 2\,\mathsf{Cov}\left[d_m, d_f\right] + \mathsf{V}\left[d_f\right] = 25(2+2\rho) = 0$.
So actually, demand is deterministic; so $x^* = 20$, with cost $cx^* + 0 = \$20$.

s.2.33. For $\rho = -1$, the expected total cost will be lower, since in this case d is actually deterministic and hence in the optimal policy, all muffins will be produced at cost \$1.
For $\rho = 1$, the expected number of muffins is still the same but either some muffins will be wasted or some muffins will be produced at cost \$1.5 instead.

s.3.1.

$$X \in \{0,\dots,5\} \implies Z \in \{0,3,6,9,12,15\}, \quad \text{and not in } \{0,\dots,15\}, \tag{5.83}$$
$$z = g(x) = 3x, \tag{5.84}$$
$$p_Z(z) = \sum_{x:g(x)=z} p_X(x) = \frac{1}{6} I_{z\in\{0,3,6,9,12,15\}}, \tag{5.85}$$
$$F_Z(z) = \frac{1}{6}\sum_{x=0}^{z} I_{x\in\{0,3,6,9,12,15\}}. \tag{5.86}$$

s.3.2.

$$X \in [0,5] \implies Z \in [0,15], \tag{5.87}$$
$$z = 3x = g(x) \implies x = z/3, \tag{5.88}$$
$$f_Z(z) = f_X(x)\frac{\mathrm{d}x}{\mathrm{d}z}, \tag{5.89}$$
$$\frac{\mathrm{d}z}{\mathrm{d}x} = 3, \tag{5.90}$$
$$f_Z(z) = f_X(z/3)\frac{1}{3}. \tag{5.91}$$

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

$$F_Z(u) = \int_0^u f_X(z/3)\frac{1}{3}\,\mathrm{d}z = \frac{1}{5}\int_0^u I_{0\leq z/3\leq 5}\frac{1}{3}\,\mathrm{d}z \tag{5.92}$$
$$= \frac{1}{5}\int_0^u I_{0\leq z\leq 15}\frac{1}{3}\,\mathrm{d}z = \frac{u}{15}. \tag{5.93}$$

s.3.3.

$$z = g(x) = (x-\mu)/\sigma, \implies x = \sigma z + \mu \tag{5.94}$$
$$f_Z(z) = f_X(x)\frac{\mathrm{d}x}{\mathrm{d}z}, \tag{5.95}$$
$$\frac{\mathrm{d}z}{\mathrm{d}x} = \frac{1}{\sigma}, \tag{5.96}$$
$$f_Z(z) = f_X(x)\sigma = \sigma f_X(\sigma z + \mu) \tag{5.97}$$

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

$$f_Z(z) = \frac{1}{\sqrt{2\pi}\sigma}e^{-(\sigma z+\mu-\mu)^2/2\sigma^2}\sigma = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}. \tag{5.98}$$

s.3.4.

$$z = g(x) = e^{-x} \implies x = -\log z, \tag{5.99}$$
$$x \in (0,\infty) \implies z \in (0,1), \tag{5.100}$$
$$f_Z(z) = f_X(x)\frac{\mathrm{d}x}{\mathrm{d}z}, \tag{5.101}$$
$$\frac{\mathrm{d}z}{\mathrm{d}x} = -e^{-x}, \quad \text{Don't forget to take the abs value next,} \tag{5.102}$$
$$f_Z(z) = f_X(x)e^{x} = e^{-x}e^{x} = 1 I_{0<z<1}, \tag{5.103}$$

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

s.3.5.

$$(u,v) = (x+y, x-y) = g(x,y) \implies (x,y) = ((u+v)/2, (u-v)/2), \tag{5.104}$$

$$\frac{\partial(u,v)}{\partial(x,y)} = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2 \implies |-2| = 2, \tag{5.105}$$

$$f_{U,V}(u,v) = f_{X,Y}(x,y)\frac{\partial(x,y)}{\partial(u,v)} = f_{X,Y}((u+v)/2, (u-v)/2)/2 \tag{5.106}$$

$$= \frac{1}{4\pi} e^{-((u+v)/2)^2/2} e^{-((u-v)/2)^2/2} \tag{5.107}$$

$$= \frac{1}{4\pi} e^{-u^2/4 - v^2/4}, \tag{5.108}$$

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

s.3.6.

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

s.3.7. If we would not add this extra variable, we cannot use the change of variables theorem. We also need a function to deal with the scaling. In the change of variables theorem, this is the Jacobian.

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

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

s.3.8. When the variables become dependent, the Jacobian becomes zero. For instance, in the latter case,

$$\frac{\partial(u,v)}{\partial(x,y)} = \begin{vmatrix} 1/y & -x/y^2 \\ -y/x^2 & 1/x \end{vmatrix} = \frac{1}{xy} - \frac{x}{y^2}\frac{y}{x^2} = 0. \tag{5.109}$$

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

s.3.9. From BH.8.1.4: Z chi-square $\implies X = \sqrt{Z} \sim \text{Norm}(0,1)$. Then, from BH.8.1.9,

$$X^2 + Y^2 = (\sqrt{2T}\cos U)^2 + (\sqrt{2T}\sin U)^2 = 2T\left(\cos^2 U + \sin^2 U\right) = 2T \sim \text{Exp}(2), \tag{5.110}$$

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

s.3.10. Take $g(x,y) = (x,w) = (x,(x+y)/2)$. Then, $y = 2w - x$.

$$\frac{\partial(x,w)}{\partial(x,y)} = \begin{vmatrix} 1 & 0 \\ 1/2 & 1/2 \end{vmatrix} = 1/2, \tag{5.111}$$

$$f_{X,W}(x,w) = f_{X,Y}(x,y)\frac{\partial(x,y)}{\partial(x,w)} = \frac{1}{\pi(1+x^2)}\frac{1}{\pi(1+(2w-x)^2)}2, \tag{5.112}$$

$$f_W(w) = \int_{-\infty}^{\infty} f_{X,W}(x,w)\,\mathrm{d}x = \frac{2}{\pi^2}\int_{-\infty}^{\infty}\frac{1}{1+x^2}\frac{1}{1+(2w-x)^2}\,\mathrm{d}x. \tag{5.113}$$

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

s.3.11. The idea is this. We generate a bunch of uniform random deviates (simulated values). Then we count how often the set $\{x + y \leq 1, xy \leq 2/9\}$ is hit.

Here is the code.

```python
import numpy as np

np.random.seed(3)

num = 100

X = np.random.uniform(0, 1, num)
Y = np.random.uniform(0, 1, num)
U = X + Y
V = X * Y
success = (U <= 1) * (V <= 2 / 9)
print(sum(success) / num)
```

s.3.12. a.

$$F \geq 0 \implies 2 < y \tag{5.114}$$

$$F \leq 1 \implies F(3,y) \leq 1 \implies F(3,4) = 1 \tag{5.115}$$

b. $F(3,7) = 1$.

c. $f(x,y) = \partial_x \partial_y F(x,y) = (x-1)/4$ for $x \in (1,3), y \in (2,4)$ and 0 elsewhere.

d.

$$\begin{aligned} \mathsf{P}\{2 < X < 3\} &= F_X(3) - F_X(2) &(5.116)\\ &= F_{X,Y}(3,4) - F_{X,Y}(2,4) = 1 - 1 \cdot 2/8 = 3/4. &(5.117)\end{aligned}$$

e. Make a drawing of the rectangle $[2,3] \times [2,4]$. Then check what parts of this are covered by $F_{X,Y}$.

$$\mathsf{P}\{2 < X < 3, 2 < Y < 3\} = F_{X,Y}(3,3) - F_{X,Y}(2,3) - F_{X,Y}(3,2) + F_{X,Y}(2,2). \tag{5.118}$$

The rest is just number plugging.

f. Use the fundamental bridge and c.

$$\begin{aligned} \mathsf{P}\{Y < 2X\} &= \mathsf{E}[I_{Y<2X}] &(5.119)\\ &= \iint I_{y<2x} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y &(5.120)\\ &= \frac{1}{4}\iint I_{y<2x} I_{2<y<4} I_{1<x<3}(x-1)\,\mathrm{d}x\,\mathrm{d}y &(5.121)\\ &= \frac{1}{4}\int_1^3 (x-1)\int I_{2<y<\min\{2x,4\}}\,\mathrm{d}y\,\mathrm{d}x &(5.122)\\ &= \frac{1}{4}\int_1^3 (x-1)(\min\{2x,4\}-2)\,\mathrm{d}x &(5.123)\\ &= \frac{1}{4}\int_1^2 (x-1)(2x-2)\,\mathrm{d}x + \frac{1}{4}\int_2^3 (x-1)(4-2)\,\mathrm{d}x. &(5.124)\end{aligned}$$

Finishing the computation must be easy for you now (and if not, practice real hard).

g. As X, Y continuous, the answer is equal to that of f.

h. Similar to f. but a bit more involved.

$$\begin{aligned} \mathsf{P}\{Y < 2X, Y + 2X > 6\} &= \mathsf{E}\left[I_{Y<2X, Y>6-2X}\right] &(5.125)\\ &= \iint I_{y<2x, y>6-2x} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y &(5.126)\\ &= \frac{1}{4}\iint I_{y<2x,y>6-2x} I_{2<y<4} I_{1<x<3}(x-1)\,\mathrm{d}x\,\mathrm{d}y &(5.127)\\ &= \frac{1}{4}\int_1^3 (x-1)\int I_{\max\{2,6-2x\}<y<\min\{2x,4\}}\,\mathrm{d}y\,\mathrm{d}x &(5.128)\\ &= \frac{1}{4}\int_1^3 (x-1)[\min\{2x,4\}-\max\{2,6-2x\}]^+\,\mathrm{d}x, &(5.129)\end{aligned}$$

where we need the $[\cdot]^+$ to ensure the positivity of $\min\{2x,4\}-\max\{2,6-2x\}$. To see this, make a graph of the function $\min\{2x,4\}-\max\{2,6-2x\}$. Also, from this graph,

$$= \frac{1}{4}\int_{3/2}^{2}(x-1)(2x-6+2x)\,\mathrm{d}x + \frac{1}{4}\int_{2}^{3}(x-1)(4-2)\,\mathrm{d}x. \tag{5.130}$$

The rest is for you.

s.3.13. The function $g(x) = x^4$ is not one-to-one on $\mathbb{R}$. It is, however, locally, one-to-one, around the roots of U. (In this course we don't deal with complex numbers, for your interest, it can be proven that the equation $x^4 - y$ has, in general, four roots in the complex plane.)

We need to be bit careful with applying the change of variables formula, but we are OK if we apply it locally around the roots $U^{1/4}$ and $-U^{1/4}$. However, mind that we also should take care of the domain of V, so it might be that these roots don't lie in the domain of V.

With all this, let's first tackle the Jacobian, and then get the domain right with indicators.

$$u = g(v) = v^4 \implies v = \pm u^{1/4}, \tag{5.131}$$

$$f_U(u)\,\mathrm{d}u = f_V(v)\,\mathrm{d}v \implies f_U(u) = f_V(v)\frac{\mathrm{d}v}{\mathrm{d}u}, \tag{5.132}$$

$$\frac{\mathrm{d}u}{\mathrm{d}v} = 4v^3 = 4u^{3/4}I_{v\geq 0} - 4u^{3/4}I_{v<0}, \tag{5.133}$$

$$f_U(u) = \frac{f_V(-u^{1/4})}{4(-u)^{3/4}}I_{-u^{1/4}\in(-3,0)} + \frac{f_V(u^{1/4})}{4(u)^{3/4}}I_{u^{1/4}\in[0,2)} \tag{5.134}$$

$$= \frac{f_V(-u^{1/4})}{4(-u)^{3/4}}I_{u\in(0,81)} + \frac{f_V(u^{1/4})}{4(u)^{3/4}}I_{u\in[0,16)}. \tag{5.135}$$

If V has the uniform distribution, then $f_V(v) = \frac{1}{5}$ for $v \in (-3,2)$, so

$$f_U(u) = \frac{1}{20(-u)^{3/4}}I_{u\in(0,81)} + \frac{1}{20(u)^{3/4}}I_{u\in[0,16)}. \tag{5.136}$$

s.3.14. Here is a direct approach.

$$x = \tan u = g(u) \implies u = \arctan x \tag{5.137}$$

$$\frac{\mathrm{d}x}{\mathrm{d}u} = \frac{1}{\cos^2 u} = \frac{\sin^2 u + \cos^2 u}{\cos^2 u} = \tan^2 u + 1 = x^2 + 1, \tag{5.138}$$

$$f_X(x) = f_U(u)\frac{\mathrm{d}u}{\mathrm{d}x} = \frac{1}{\pi}I_{u\in(0,\pi)}\frac{1}{1+x^2} \tag{5.139}$$

$$= \frac{1}{\pi(1+x^2)}I_{\arctan x\in(0,\pi)} = \frac{1}{\pi(1+x^2)}. \tag{5.140}$$

In the last equation we just shifted the tan from $(-\pi/2,\pi/2]$ to the interval $(0,\pi)$. The tan has also a proper inverse in $(0,\pi)$ (make a drawing of tan to see this), hence all is well-defined.

s.3.19. If we divide two $\sim \text{Norm}(0,1)$ r.v.s we obtain a Cauchy distributed r.v. But in our Beluga case, the normally distributed r.v.s have positive expectation.

s.3.28. Suppose we chop up the area under some arbitrary function g in blocks of height $g(x)$ and length Δx. Then the area of such a block is $g(x)\Delta x$.

In our case, we chop up the interval in parts with length $\Delta x = 1/N$. The elements of $f3$ are such that $f3[i] = f_3(x_i)\Delta x$, where x_i lies in the ith interval and f_3 is the density of the sum of the three r.v.s. But then, $f_3(x_i) = f3[i]/\Delta x = Nf3[i]$.

Forgetting to scale with $\Delta x = 1/N$ is a common error when dealing with densities. Hence, recall that, notationally, $f(x)\,\mathrm{d}x$ means a block of height $f(x)$ and length $\mathrm{d}x$. Don't forget to deal with the $\mathrm{d}x$!

s.3.29. Take four r.v.s U, V, X, Y. Then

$$(U+V+X)+Y = (U+V)+(X+Y). \tag{5.141}$$

Thus the density of $(U+V+X)+Y$ must be the same as the density of $(U+V)+(X+Y)$.