Probability distributions EBP038A05 Assignments

Nicky van Foreest
Joost Doornbos, Wietze Koops, Mikael Makonnen, Arpan Rijal

February 21, 2022

CONTENTS

INTRODUCTION

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

Each assignment contains code in Python and R that shows how to implement an example or the solution of an exercises of BH. You have to run the code, read the output, and explain how 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.[1] 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 looking 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 assignment, 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 few sentences, or a number plus some short explanation. The idea of the assignment is to help you study and improve your coding skills by showing good code, 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] (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 extremely tired of this, so I went to using Python as this can do all of this stuff, but within one language.

1 ASSIGNMENT 1

1.1 *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? Amazingly, this turns out to be nearly exponential. To check this fact, we use simulation in this question.

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!).

Python Code

```
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
print(1 / labda)

A = np.sort(np.random.uniform(start, end, size=num))
print(A) #this
print(A[1:])
print(A[:-1])
X = A[1:] - A[:-1]
print(X)

print(X.mean(), X.std())
```

R Code

```r
set.seed(3)


num = 5  # small sample at first, for checking.
start = 0
end = 3
labda = num / (end - start)  # per minute
print(1 / labda)

A = sort(runif(num, min = start, max = end))
print(A) #this
print(A[-1])
print(A[-length(A)])
X = A[-1] - A[-length(A)]
print(X)

print(mean(X))
print(sd(X))
```

Ex 1.1. Explain the result of line P.12 (R.11) marked as 'this'.[2]

Ex 1.2. Compare the result of line P.13 and P.14 (R.12, R.13); explain what is `A[1:]` (`A[-1]`).[3]

Ex 1.3. Compare the result of line P.12 and P.14 (R.11 and R.13); explain what is `A[:-1]` (`A[-length(A)]`).

Ex 1.4. Explain what is `X` in P.15 (R.14).

Ex 1.5. Why do we compare $1/\lambda$ and `X.mean()`?

Ex 1.6. Recall that $\mathsf{E}[X] = \sigma(X)$ when $X \sim \text{Exp}(\lambda)$. Hence, what do you expect to see for `X.std()`?

Ex 1.7. Run the code for a larger sample, e.g. 50, and discuss (very briefly) your results.

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

[3] When two pieces of code are given and one is in parentheses, the first one is the Python code and the one in parentheses is the R code. So if you use Python you have to explain what `A[1:]` does; if you use R you have to explain what `A[-1]` does.

1.2 *BH.5.6.5*

Read this example of BH first. We chop up the exercise in many small steps.

For the code below, run it first for a small number of samples (Here I choose `samples=2`). Then read (and if necessary put in more) print statements to see what results you get. Use the output to answer the questions below.

Python Code

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

np.random.seed(10)

labda = 6
num = 3
samples = 2

X = expon(scale=labda).rvs((samples, num))
print(X) #this
T = np.sort(X, axis=1)
print(T)
print(T.mean(axis=0))

expected = np.array([labda / (num - j) for j in range(num)])
print(expected)
print(expected.cumsum())
```

R Code

```r
set.seed(10)

labda = 6
num = 3
samples = 2

X = matrix(rexp(samples * num, rate = 1 / labda), nrow = samples, ncol = num)
print(X) #this
bigT = X
for (i in 1:samples) {
  bigT[i,] = sort(bigT[i,])
}
print(bigT)
print(colMeans(bigT))

expected = rep(0, num)
for (j in 1:num) {
  expected[j] = labda / (num - (j - 1))
}
print(expected)
print(cumsum(expected))
```

Ex 1.8. In line P.11 (R.8),[4] also marked as 'this', we print the value of `X` in line P.10 (R.7), respectively. What is the meaning of `X`?

Ex 1.9. What is the meaning of `T` in line P.12 (R.11)?

Ex 1.10. What do we print in line P.14 (R.14)?

Ex 1.11. What is meaning of the variable `expected`?

Ex 1.12. What is the `cumsum` of `expected`?

Ex 1.13. 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.

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

1.3 *BH.7.1*

First read and solve BH.7.1. Answers and hints are in the study guide.

Ex 1.14. Sketch, on paper, the 2D area that corresponds to the event that Alice and Bob meet. Just make a photo with your mobile phone and include it. (Like this you learn how to include jpeg or png files in LaTeX.)

Ex 1.15. Write A and B for the rvs corresponding to the arrival times of Alice and Bob. Explain you have to compute the expectation of $I_{A<B+0.25} \cdot I_{B-0.25<A}$ to compute the probability that Alice and Bob meet. Then, explain that it is equal to

$$\int_0^1 \int_0^1 I_{a<b+0.25} \cdot I_{b-0.25<a} \, \mathrm{d}a \, \mathrm{d}b. \tag{1.1}$$

Ex 1.16. Solve the integral by hand to compute the probability that they meet.

Ex 1.17. Explain the code below to compute numerically the probability of overlap. Check the web on what is an anonymous function in Python or R, and explain why we use that here. (Hint: anonymous functions are useful "when it's not worth the effort to give it a name". Do we use elsewhere the anonymous function we define here?)

Python Code

```python
from scipy.integrate import dblquad

area = dblquad(lambda a, b: (b - 0.25 < a) * (a < b + 0.25), 0, 1, 0, 1)
print(area)
```

R Code

```r
library(cubature)

area = adaptIntegrate(function(x){ (x[2] - 0.25 < x[1]) * (x[1] < x[2] + 0.25)},
        lowerLimit = c(0, 0), upperLimit = c(1, 1))
print(area$integral)
print(area$error)
```

Ex 1.18. Explain the output, i.e., the results of the print statements, of this piece of code.

Python Code

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

np.random.seed(3)

n = 5
alice = uniform(0, 1).rvs(n)
bob = uniform(0, 1).rvs(n)
print(alice)
print(bob)
print((alice < bob))
```

R Code

```r
set.seed(3)

n = 5
alice = runif(n,0,1)
bob = runif(n,0,1)

print(alice)
print(bob)
print(alice<bob)
```

It is useful to set the seed of the random number generator so that we always get the same results. This helps a lot when debugging, because the numbers don't change time and again. (You can remove the line that sets the seed, and then rerun the program a few times. You should get different numbers for each run.)

Ex 1.19. Run this piece of code, and then check that you get different output every time.

Python Code

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

n = 5
alice = uniform(0, 1).rvs(n)
bob = uniform(0, 1).rvs(n)
print(alice)
print(bob)
overlap = (alice < bob) # this
print(overlap)
```

R Code

```
n = 5
alice = runif(n,0,1)
bob = runif(n,0,1)
print(alice)
print(bob)
overlap = alice<bob #this
print(overlap)
```

Explain that the line marked as 'this' implements an indicator function.

Ex 1.20. Run this piece of code, and explain the output.

Python Code

```
import numpy as np
from scipy.stats import uniform

np.random.seed(3)

n = 5
alice = uniform(0, 1).rvs(n)
bob = uniform(0, 1).rvs(n)
overlap = (bob - 0.1 < alice) * (alice < bob + 0.1)
print(alice)
print(bob)
print(overlap)  # this
```

R Code

```
set.seed(3)

n = 5
alice = runif(n,0,1)
bob = runif(n,0,1)
overlap = (bob - 0.1 < alice) * (alice < bob + 0.1)
print(alice)
print(bob)
print(overlap) #this
```

Ex 1.21. To find the probability of overlap, we have to modify the code above the line marked as 'this' to

Python Code

```
print(overlap.mean())
```

R Code

```
print(mean(overlap))
```

Change this in your code and explain what it does.

Ex 1.22. Modify the code of the previous exercise so that you sample 1,000 times an appointment between Alice and Bob, and such that they only meet when 15 minutes apart. Include your code and show the output.

Ex 1.23. Contrary to [1.14] suppose Bob is not prepared to wait longer than 5 minutes, while Alice is still prepared to wait for 15 minutes. Write down the integral that corresponds to the probability that they meet. Solve it by hand. Modify the above code such that you can evaluate it numerically. Include your code in the answers.

Ex 1.24. Contrary to [1.14] suppose Bob is prepared to wait for 20 minutes when he arrives before 1230, but just 10 when he arrives after 1230, while Alice is still prepared to wait for 15 minutes. Write down the integral that corresponds to the probability that they meet. Explain that this is the code to compute this.

Python Code

```
from scipy.integrate import dblquad
area1 = dblquad(lambda a, b: (a < b + 1 / 3.0 ) * (b < a + 0.25), 0, 1, 0, 0.5)
area2 = dblquad(lambda a, b: (a < b + 1 / 6.0) * (b < a + 0.25), 0, 1, 0.5, 1)
print(area1[0] + area2[0])
```

R Code

```
library(cubature)

area1 = adaptIntegrate(function(x){(x[1] - 1 / 3.0 < x[2]) * (x[2] < x[1] + 0.25)},
        lowerLimit = c(0, 0), upperLimit = c(1, 0.5))
area2 = adaptIntegrate(function(x){(x[1] - 1 / 6.0 < x[2]) * (x[2] < x[1] + 0.25)},
        lowerLimit = c(0, 0.5), upperLimit = c(1, 1))
print(area1$integral+ area2$integral)
```

Ex 1.25. Contrary to [1.14] suppose Bob's arrival time is Exp(λ) with $\lambda = 1/2$, while Alice's arrival time is still uniform. Explain the code below.

Python Code

```python
import numpy as np
from scipy.integrate import dblquad
from scipy.stats import uniform, expon

np.random.seed(3)

labda = 1.0 / 2
end = 5

value = dblquad(
    lambda a, b: (b - 0.25 < a) * (a < b + 0.25) * np.exp(-labda * b), 0, 1, 0, end
)
print(value)

area = value[0] * labda
print(area)

n = 100000
alice = uniform(0, 1).rvs(n)
bob = expon(scale=1 / labda).rvs(n)
overlap = (bob - 0.25 < alice) * (alice < bob + 0.25)
print(overlap.mean())
```

R Code

```
library(cubature)
set.seed(3)

labda = 1.0/2
end = 5

value = pcubature(function(x){(x[2] - 0.25 < x[1]) * (x[1] < x[2] + 0.25) *
        exp(-labda * x[2])}, lowerLimit = c(0, 0), upperLimit = c(1, end),
      tol = 0.0001)
print(value)

area = value$integral * labda
print(area)

n = 100000
alice = runif(n, 0, 1)
bob = rexp(n, labda)
overlap = (bob - 0.25 < alice) * (alice < bob + 0.25)
print(mean(overlap))

print(mean(bob))
```

Ex 1.26. Changing the value of the variable `end` to 100 in the code above, and run it again. Then change it to 50, and see what happens. When I do this in python, I get a very strange answer. Then try 1000: in python you should also get some strange value, and R doesn't seem to give an answer at all.

In this exercise you don't have to explain anything; you just have to see that things can go wrong easily. You should memorize from this exercise that numerical integration in multiple dimensions quickly becomes unmanageable. In fact, most 2D (and higher dimensional) ntegrals are approximated with simulation, like we do here.

Hopefully you got by now the point: modeling is an interesting, but quite difficult, activity. Once you have the model, the computer can help solve the problem. However, computers are extremely dumb, and they can live with any answer whatsoever. Mind, we humans can do pretty smart things, and we should not automatically trust the output of computers. That is the reason why we have to include many tests and checks.

1.4 *BH.7.9*

First read and solve BH.7.9. Here we provide a simulation.

Ex 1.27. Run this code, and use the documentation of scipy.stats.geom in Python (rgeom() function in R) to explain why it fails. Perhaps you can also use the wikipedia page on the geometric distribution. (Hint: is this the first success distribution or the geometric distribution?)

Python Code

```python
import numpy as np
from scipy.stats import randint, geom

np.random.seed(3)

p, num = 0.1, 10000
q = 1 - p
rv = geom(p)  # This line
X = rv.rvs(num)

if X.min() != 0:
    raise ValueError("The minimal value of X is not 0")
if not np.isclose(q / p, X.mean(), 0.1):
    raise ValueError("The mean of X is not ok")
```

R Code

```r
library(dplyr)
set.seed(3)

p = 0.1
num = 100000
q = 1 - p
X = rgeom(num, p) + 1 #This line

if (min(X) != 0){
  stop("The minimal value of X is not 0")
}
if (!(near(q/p,mean(X),0.1))){
  stop("The mean of X is not ok")
}
```

Ex 1.28. The code of the previous exercise failed, on purpose. It will work if we change the line marked as “this line”, into this:

Python Code

```python
rv = geom(p, loc=-1)
```

R Code

```
X = rgeom(num, p)
```

Explain the difference, and why it now works. Relate this location-scale transformation to one of the distributions you know from BH.

Memorize from this exercise that with such checks you can build code that fails when it gets unexpected input.

Ex 1.29. Run this code and explain in detail the output. Make, and include, some simple example code to show that you are correct.

Python Code

```
import numpy as np

N = np.array([3, 4, 4, 8, 4])
X = np.array([8, 7, 6, 5, 3])
print(np.argwhere(N == 4))
Xn = X[np.argwhere(N == 4)]
print(Xn)
```

R Code

```
N = c(3, 4, 4, 8, 4)
X = c(8, 7, 6, 5, 3)
print(which(N == 4))
Xn = X[which(N == 4)]
print(Xn)
```

Ex 1.30. Explain the lines marked with 'this line' in the code of Listing 1 or Listing 2. You already explained the other parts so you don't have to do that again.

Ex 1.31. Explain the results. Do they match with the theoretical results?

Python Code

```python
import numpy as np
from scipy.stats import randint, geom
import matplotlib.pyplot as plt

np.random.seed(3)

p, num = 0.1, 100000
q = 1 - p
rv = geom(p, loc=-1)
X = rv.rvs(num)

Y = rv.rvs(num)
N = X + Y

fig, axes = plt.subplots(3, 3, sharey="all", figsize=(6, 3))  # this
n = 7  # gives nice results, after some experimentation
for ax in axes.flatten():  # this
    Xn = X[np.argwhere(N == n)]
    y, bins = np.histogram(Xn, bins=np.linspace(0, n + 1, n + 2)) # this
    ax.plot(range(n + 1), y, 'bo', ms=3, label='N=8')  # this
    ax.vlines(range(n + 1), 0, y, colors='b', lw=2, alpha=0.5) # this
    ax.set_title(f'{n=} ({len(Xn)})')
    ax.set_ylim(0) # this
    n += 1
```

Listing 1: BH.7.9, Python code.

R Code

```r
library(ggplot2)
library(gridExtra)

p = 0.1
num = 100000
q = 1 - p
X = rgeom(num, p)

Y = rgeom(num, p)
N = X + Y

pdf(file="bh-7-9.pdf", width = 8, height = 7,
    bg = "white", colormodel = "cmyk", paper = "A4")

plots <-list()
par(oma= c(0,0,0,0))

i = 1
n = 7  # gives nice results, after some experimentation
while (i <= 9){  #this
  Xn = X[which(N == n)]
  df = data.frame(Xn)
  plots[[i]] =  ggplot()+
                geom_histogram(df, mapping = (aes(x = Xn)),
                bins = n + 1, color = "blue")+ # this
                ylim(0,600)+  # this
                labs(title=paste0("n=", n, ", (", length(Xn) ,")"))+
                theme(axis.title.x=element_blank(),
                axis.title.y = element_blank())
  n = n + 1
  i = i + 1
}

do.call("grid.arrange", c(plots, ncol=3))
dev.off()
```

Listing 2: BH.7.9, R code.

1.5 *Challenges*

Select one of either problem if you would like to get a 10. Of course you can try both, but one suffices.

1.5.1 *Challenge: A uniqueness property of the Poisson distribution*

Consider 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 rv. 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 1.32. 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: $0 \in \mathbb{N}$. In this exercise we want to prove that N is Poisson distributed. So you cannot assume this in your solution.

Ex 1.33. Write $Y = N - X$. Prove that

$$\mathsf{P}\{X = x\}\,\mathsf{P}\{Y = y\} = \binom{x+y}{x} p^x (1-p)^y\,\mathsf{P}\{N = x + y\}. \tag{1.2}$$

Ex 1.34. Prove that N is Poisson distributed. Use the relation of the previous exercise to show that

$$P(N = n+1) = \frac{\lambda}{1+n} P(N = n). \tag{1.3}$$

Bigger hint: Fill in $y = 0$ in the LHS and RHS of (1.2); 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

$$\lambda = \frac{\mathsf{P}\{Y = 1\}}{(1-p)\,\mathsf{P}\{Y = 0\}}. \tag{1.4}$$

1.5.2 *Challenge: 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.[5]

[5] Such integrals are called improper Riemann 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 rv. exists).

Ex 1.35. Why does $\mathsf{E}\left[\frac{|X|}{X^2+1}\right]$ exist? Find its value. It is essential that you include your arguments.

Ex 1.36. 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 *BH.7.48*

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

Python Code

```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 Code

```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
    }
  }
  return(num_max)
}


num = 10
X = runif(num, min = 0, max = 1)
print(X)

print(find_number_of_maxima(X))

samples = 100
Y = rep(0, samples)
for (i in 1:samples) {
  X = runif(num, min = 0, max = 1)
  Y[i] = find_number_of_maxima(X)
}

print(mean(Y))
print(var(Y))
print(sd(Y))
```

Ex 2.1. 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`. Unfortunately, R does not have such increment operators.)

Ex 2.2. Explain the code in lines P.25 and P.26 (R.25, R.26).

2.2 *BH.7.53*

Ex 2.3. Read and solve BH.7.53, so that you understand what the simulation is about. (You don't have to provide an answer for this exercise.)

Ex 2.4. The code in Listing 3 or Listing 4 simulates this exercise. Run the code. (To improve your understanding, just change some parameters here and there, and check the output.) Explain the lines marked with `#*`. (In particular, why do we do a `runs` number of runs to estimate the covariance?) Finally, explain the output.

Ex 2.5. Modify the code such that the drunkard makes steps of size 2 when moving left or right; the stepsizes up or down remain the same. What happens with $\mathsf{E}\left[R_n^2\right]$?

Ex 2.6. Optional: Add a third dimension.

2.3 *BH.2.3.9*

In line with Exercise BH.7.86, we are now going to analyze the effect on $\mathsf{P}\{D \mid T\}$ when the sensitivity is not known exactly. So, why is this interesting? In Example 2.3.9 the sensitivity is given, but in fact, in 'real' experiments, this is not always known as accurately as assumed in this example. For example, in this paper: False-positive COVID-19 results: hidden problems and costs it is claimed that 'The current rate of operational false-positive swab tests in the UK is unknown; preliminary estimates show it could be somewhere between 0·8% and 4·0%.'. Hence, even though it is claimed that PCR tests 'have analytical sensitivity and specificity of greater than 95%', it may be 4% lower. Simply put, the specificity and sensitivity are not precisely known, hence this must affect $\mathsf{P}\{D \mid T\}$.

I write `p_D_g_T` for $\mathsf{P}\{D \mid T\}$. Here is how Example 2.3.9 can be implemented.

Python Code

```
sensitivity = 0.95
specificity = 0.95
p_D = 0.01

p_T = sensitivity * p_D + (1-specificity)*(1-p_D)
p_D_g_T  = sensitivity * p_D/p_T
p_D_g_T
```

Python Code

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

np.random.seed(3)


num = 4
steps = randint(0, 4).rvs(num)  # *
print(steps)
X = (2 * steps - 1) * (steps < 2)  # *
Y = (2 * steps - 5) * (steps >= 2)  # *
print(X)
print(Y)
R2 = (X.sum()) ** 2 + (Y.sum()) ** 2
print(f"{num = }, {R2 = }")


def cov(X, Y):
    return (X * Y).mean() - X.mean() * Y.mean()  # *


num = 10
runs = 300
Ss = np.zeros(runs)  # *
Ts = np.zeros(runs)
R2s = np.zeros(runs)
for i in range(runs):  # *
    steps = randint(0, 4).rvs(num)
    X = (2 * steps - 1) * (steps < 2)
    Y = (2 * steps - 5) * (steps >= 2)
    Ss[i] = X.sum()  # *
    Ts[i] = Y.sum()  # *
    R2s[i] = (X.sum()) ** 2 + (Y.sum()) ** 2   # *

print(f"{cov(Ss,Ts) = }")  # *
print(f"{R2s.mean() = }")  # *
```

Listing 3: BH.7.53, python code.

R Code

```r
sensitivity = 0.95
specificity = 0.95
p_D = 0.01

p_T = sensitivity * p_D + (1-specificity)*(1-p_D)
p_D_g_T  = sensitivity * p_D/p_T
p_D_g_T
```

R Code

```r
set.seed(3)

num = 4
steps = sample(0:3, num, replace = TRUE)  # *
print(steps)
X = (2 * steps - 1) * (steps < 2)  # *
Y = (2 * steps - 5) * (steps >= 2)  # *
print(X)
print(Y)
R2 = sum(X)^2 + sum(Y)^2
print(paste0("num = ", num, " R2 = ", R2))

cov = function(X, Y) {
  return (mean(X * Y) - mean(X) * mean(Y))  # *
}

num = 10
runs = 300
Ss = rep(0, runs)  # *
Ts = rep(0, runs)
R2s = rep(0, runs)
for (i in 1:runs) {  # *
  steps = sample(0:3, num, replace = TRUE)
  X = (2 * steps - 1) * (steps < 2)
  Y = (2 * steps - 5) * (steps >= 2)
  Ss[i] = sum(X)  # *
  Ts[i] = sum(Y)  # *
  R2s[i] = sum(X)^2 + sum(Y)^2  # *
}

print(paste0("cov(Ss, Ts) = ", cov(Ss, Ts)))  # *
print(paste0("R2 = ", mean(R2s)))  # *
```

Listing 4: BH.7.53, R code.

Ex 2.7. 1. Make a plot of $\mathsf{P}\{D \mid T\}$ in which you vary the sensitivity from 0.9 to 0.99. Explain what you see.

2. Make a plot of $\mathsf{P}\{D \mid T\}$ in which you vary the specificity from 0.9 to 0.99.

3. Make a plot of $\mathsf{P}\{D \mid T\}$ in which you vary $\mathsf{P}\{D\}$ from 0.01 to 0.5. Explain what you see.

2.4 *Challenges*

Try this challenge if you would like to get a 10.

2.4.1 *Challenge: Proof about independence of normal rvs*

Consider two iid rvs X, Y such that $X+Y$ and $X-Y$ are independent. In BH 7.5.8, it is claimed that this implies that X and Y are normally distributed.

This challenge asks to give a proof of this claim. Throughout this problem, you may assume that X and Y have a MGF that is defined for all $t \in \mathbb{R}$.[6] You may also use without proof the fact that MGFs that are defined everywhere are infinitely often continuously differentiable.

Ex 2.8. Let M_X be the MGF of X (and hence of Y).

1. Prove that $M_X(2t) = (M_X(t))^3(M_X(-t))$.
2. Define $f(t) = \log M_X(t)$. Prove that $8f'''(2t) = 3f'''(t) - f'''(-t)$.
3. Let $R > 0$ be arbitrary. Use Weierstrass' theorem to prove that f''' attains a minimum m and a maximum M on the interval $[-R, R]$, and then prove that $m = M = 0$.
4. Prove that X is normally distributed.

---

[6] This may seem like a big restriction, but this argument can easily be adapted to work with the *characteristic function* instead of the MGF, and the characteristic function does always exist. You will learn the characteristic function in the second year courses Statistical Inference and Linear Models in Statistics.

3 ASSIGNMENT 3

3.1 *BH.8.4 Figure*

The challenge is to make Figure BH.8.4 for a specific case.

We throw a coin a number of times. We take some beta distribution as prior for the probability p that Heads appears. Then we update the PDF of p according to outcomes of the throws.

Ex 3.1. Run the follow code and explain the output.

We define the variables H and T as these outcomes are easier to memorize than 1 and 0. Like this, there is less possibility for confusion, hence less bugs.

Python Code

```python
H, T = 1, 0
outcomes = [H, T]

h, t = 1, 1  # this
h += outcomes[0] == H
t += outcomes[0] == T
print(f"{h=}, {t=}")

h += outcomes[1] == H
t += outcomes[1] == T
print(f"{h=}, {t=}")
```

R Code

```r
H = 1
T = 0
outcomes = c(H,T)

h = 1 #this
t = 1 #this
h = h + as.numeric(outcomes[1] == H)
t = t + as.numeric(outcomes[1]== T)
print(paste0("h=", h, ", t=", t))

h = h + as.numeric(outcomes[2] == H)
t = t + as.numeric(outcomes[2]== T)
print(paste0("h=", h, ", t=", t))
```

Ex 3.2. Explain the code below. Why does the line marked as 'this' relate to the prior distribution? What is that prior distribution?

Python Code

```python
import numpy as np
from scipy.stats import beta
import matplotlib.pyplot as plt

support = np.arange(0, 1, 0.03)

H, T = 1, 0
outcomes = [H, H, T, H, H, H, T, H, H]
h, t = 1, 1  # this

fig, axes = plt.subplots(3, 3, sharey="all", figsize=(6, 3))  # this
for i, ax in enumerate(axes.flatten()):  # this
    h += outcomes[i] == H
    t += outcomes[i] == T
    X = beta(h, t).pdf(support)
    ax.plot(support, X)
    ax.set_title(f'h={h-1}, t={t-1}')

plt.tight_layout()
plt.savefig("beta.pdf")
```

R Code

```r
library(ggplot2)
library(gridExtra)

support = seq(0, 1, 0.03)

H = 1
T = 0
outcomes = c(H, H, T, H, H, H, T, H, H)
h = 1 #this
t = 1 #this

pdf(file="beta.pdf", width = 8, height = 7,
    bg = "white", colormodel = "cmyk", paper = "A4")

plots <-list()
par(oma= c(0,0,0,0))

i = 1
while (i <= length(outcomes)){    #this
  h = h + (outcomes[i] == H)
  t = t + (outcomes[i] == T)
  X = dbeta(support, h, t)
  df = data.frame(support,X)
  plots[[i]] = ggplot(data=df, aes(x=support, y=X, group=1)) +
               geom_line()+
               ylim(0,3.5)+
               labs(title=paste0("h=", h-1, ", t=", t-1),x=" ", y = " ")
  i = i+1
}

do.call("grid.arrange", c(plots, ncol=3))

dev.off()
```

Ex 3.3. What do the PDFs actually model? Explain why the PDFs behave as they do. For instance, why does the location of the peak move from right to left to right?

Ex 3.4. Which conjugacy relation between prior and data do we use here to update the prior to the posterior?

Ex 3.5. Adapt the code above such that the prior is $a = 3.5, b = 1.5$, the outcomes are T, H, T, H, T, H, and the output is a figure with 3 x 2 subfigures. Besides the explanation, include the figure.

3.2 *BH-8-4-3*

Let us investigate BH.8.4.3. The rvs X and Y are independent, $X \sim \text{Exp}(\lambda)$ and $Y \sim \text{Exp}(\mu)$. We want to compare $X + Y$ to Gamma$(2, \mu)$ and Gamma$(2, \lambda)$, just to see what happens.

Ex 3.6. Run the code below. Explain the lines marked as 'this'. In particular, why do we simulate outcomes for X and Y? Then, why are the shape and scale variables as they are? Then explain the three graphs.

Python Code

```python
import numpy as np
from scipy.stats import expon, gamma
import matplotlib.pyplot as plt

labda, mu = 3, 5

n = 1000
X = expon(scale=1 / labda)
Y = expon(scale=1 / mu)
Z = X.rvs(n) + Y.rvs(n)
G1 = gamma(2, scale=1 / labda)
G2 = gamma(2, scale=1 / mu)

support = np.arange(0, 5, 0.03)

plt.hist(Z, label="Z", density=True)
plt.plot(support, G1.pdf(support), label="G1")
plt.plot(support, G2.pdf(support), label="G2")
plt.legend()
plt.tight_layout()
plt.savefig("figures/gamma1.pdf")
```

R Code

```r
library(ggplot2)

labda = 3
mu = 5

n = 1000
X = rexp(n, labda)
Y = rexp(n, mu)
Z = X + Y

support = seq(0, 5, 0.03)

G1 = dgamma(support, 2, scale=1 / labda)
G2 = dgamma(support, 2, scale=1 / mu)

df1 = data.frame(Z)
df2 = data.frame(support,G1, G2)

pdf(file="figures/gamma1.pdf", width = 8, height = 7,
    bg = "white", colormodel = "cmyk", paper = "A4")

ggplot()+
geom_histogram(df1, mapping = (aes(x = Z, y = ..density.., fill = "Z")), bins = 20)+
geom_line(df2, mapping = (aes(x = support, y = G1, color="G1")))+
geom_line(df2, mapping = (aes(x = support, y = G2, color = "G2")))+
scale_fill_manual(name = " ", values = c("Z" = "blue"))  +
scale_color_manual(name = " ", values = c( "G1" = "orange", "G2" = "green"))+
theme(axis.title.x=element_blank(),axis.title.y = element_blank())

dev.off()
```

Ex 3.7. Modify the above code such that X and Y are both Exp(λ). Then make a histogram of a simulation of $X+Y$ and compare it to the PDF of the correct gamma distribution. Discuss your result, and include the figure of course.

3.3 *BH.8.1*

Ex 3.8. Here is the code to solve BH.8.1. Explain the relevant steps, i.e., the lines marked as 'this'. Is the result according to your expectation?

Python Code

```python
import numpy as np
from scipy.stats import expon
import matplotlib.pyplot as plt

X = expon(scale=1) # this
Y = np.exp(-X.rvs(1000)) # this
plt.hist(Y, density=True, label="sim")

x = np.linspace(0, 1, 30)
dxdy = np.exp(x)  # this
pdf_y = X.pdf(x) * dxdy # this
plt.plot(x, pdf_y, label="f_Y")
plt.legend()
plt.tight_layout()
plt.savefig("figures/bh-8-1-fig.pdf")
```

R Code

```r
library(ggplot2)

X = rexp(1000, 1) #this
Y = exp(-X) #this
df1 = data.frame(Y)

x = seq(0,1,length.out = 30)
dxdy = exp(x) #this
pdf_y = dexp(x,1) * dxdy #this
df2 = data.frame(x,pdf_y)

pdf(file="figures/bh-8-1-fig.pdf", width = 8, height = 7,
    bg = "white", colormodel = "cmyk", paper = "A4")

ggplot()+
geom_histogram(df1, mapping = (aes(x = Y, y = ..density.., fill = "sim")), bins = 20)+
geom_line(df2, mapping =(aes(x=x,y=pdf_y, color = "f_y")))+
scale_fill_manual(name = " ", values = c("sim" = "blue"))  +
scale_color_manual(name = " ", values = c( "f_y" = "orange"))+
theme(axis.title.x=element_blank(),axis.title.y = element_blank())

dev.off()
```

Ex 3.9. What error would we make if we would use this code: `dxdy = np.exp(-x)` (`dxdy = exp(-x)`)?

3.4 *BH.8.18*

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 $1500\,\mathrm{m}^3$. 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., $V = 2500\,\mathrm{m}^3$. The volume of a ping pong ball is $v = 4\pi r^3/3 = 33.49333333333333\,\mathrm{cm}^3$ 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 account the sphere packing factor. Besides that, I hope you agree with me that providing a result with the precision as given here is plain ridiculous.[7]

In fact, I know that the volumes of an airplane 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

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

where the variances express my trust in my guess work. What is now the mean of $N = V/v$ and its std? Can we get the CDF? From BH.8.18 you know that finding the closed form expression for the distribution of N is not entirely simple. However, with simulation it's easy to get an estimates for the mean and the standard deviation.

Ex 3.10. Explain lines marked as 'this'.

[7] For reasons incomprehensible to me, even professional econometricians sometimes report results with 10 digits or more, without questioning the precision.

Python Code

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

num = 500

np.random.seed(3)

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

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

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

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

R Code

```r
num <- 500

set.seed(3)

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

N = V / v  # this
print(paste(mean(N), sd(N)))

```

Ex 3.11. Contrary to BH.7.1.25 if you run the code above, you'll see that $\mathsf{E}[N] < \infty$, and is 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. Explain why it works for the Beluga case, but not for the case discussed in BH.7.1.25.

Ex 3.12. Include code to make a histogram of the PDF of N.

Remark 3.13. The numerical results suggest the interesting guess $\mathsf{V}[N] \approx \mathsf{V}[V] * \mathsf{V}[v]$. Whether this is true more generally will be discussed in the Challenge of week 4.

3.5 *Challenges*

Try this challenge if you would like to get a 10.

3.5.1 *Challenge: Recourse models*

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):

$$\min_{x \geq 0} \{cx \,:\, x \geq d\}. \tag{3.1}$$

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

Ex 3.14. 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*:

$$\min_{x \geq 0} \big\{cx + \mathsf{E}\,[v(d,x)]\big\}, \tag{3.2}$$

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

$$v(d,x) := \min_{y \geq 0} \{qy \,:\, x + y \geq d\}, \tag{3.3}$$

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

$$v(d,x) = q(d-x)^+, \tag{3.4}$$

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

$$(s)^+ = \begin{cases} s & \text{if } s \geq 0, \\ 0 & \text{if } s < 0. \end{cases} \tag{3.5}$$

Ex 3.15. 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^*. First, compute the value of $\mathsf{E}[v(d,x)]$ as a function of x. Then find the optimal value of x.

Ex 3.16. 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 3.17. Calculate x^* and the corresponding objective value for the case $\rho = -1$. (Do not read $\rho = 1$, this case is not simple.)

Ex 3.18. 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. You don't have to compute x^* for the case where $\rho = 1$; this is not easy!

4 ASSIGNMENT 4

4.1 *BH.8.27*

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 is some code to simulate.

Python Code

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

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 Code

```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)
# print(XYZ)  # if you want to see it.
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 4.1. 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 4.2. What is the shape (rows and columns) of Z?

Ex 4.3. Explain the values for `loc` and `shape` in `Zexact` (and corresponding `mean` and `sd` in R). (Read the documentation of scipy.stats.norm (or norm() for R) on the web if necessary.) To which definition in BH does this loc-scale transformation relate?

Ex 4.4. 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 Code

```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 Code

```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

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 4.5. Read the documentation of np.convolve (or convolve() in R). Why is it called like this?

Ex 4.6. In the code, what is `f2`?

Ex 4.7. What is `f3`?

Ex 4.8. Why do we set `k=3`?

Ex 4.9. A bit harder, why do we plot `N*f3`, i.e., why do we have to multiply with `N`? Relate this to the meaning of $f(x)\,\mathrm{d}x$, where f the density of some random variable. (Think hard about this problem; the ideas are important to understand how numerical integration works.) 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, $f_3(x_i) = f3[i]/\Delta x = Nf3[i]$.

So, why should we scale with $\Delta x = 1/N$? (Forgetting this step is a common error when dealing with densities.)

To memorize, $f(x)\,\mathrm{d}x$ is the area of a block of height $f(x)$ and length $\mathrm{d}x$.

Ex 4.10. Yet a tiny bit harder, consider `f4 = np.convolve(f3, fx)` (`f4 = convolve(f3, fx, type = "open")` in R) and `g4 = np.convolve(f2, f2)` (`g4 = convolve(f2, f2, type = "open")` in R). Why are they, numerically speaking, equal?

Ex 4.11. When you would compute the maximum of `np.abs(f4 -g4)` (`abs(f4 - g4)` in R) 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()` (`all.equal()` in R) 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.)

4.2 *BH.8.36*

We simulate the post office part of BH.8.36. Read it now, i.e., before reading the text below. Then read the code below. In the questions, we ask you to explain what the code does. There are lots of print statements that have been commented out, but we left them in for you to include while experimenting with the code to see how the code works. (I often use print statements of intermediate results when writing a program, just to see whether I am still on track. Once I checked, I remove them, because they clutter the looks of the code.)

Python Code

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

np.random.seed(3)

labdas = np.array([3, 4])

N = 10

T1 = expon(scale=1 / labdas[0]).rvs(N)
# print(T1.mean())
T2 = expon(scale=1 / labdas[1]).rvs(N)
T = np.zeros([2, N])
T[0, :] = T1
T[1, :] = T2
# print(T)
server = np.argmin(T, axis=0)
# print(server)

# BH.8.36.b
print((server == 1).mean())

# BH.8.36.c
# print(labdas[server])
S = expon(scale=1).rvs(N) / labdas[server] # this
# print(S)

D = np.min(T, axis=0) + S
# print(D)

print(D.mean(), D.std())
```

R Code

```r
set.seed(3)

labdas = 3:4

N = 10

T1 = rexp(N, rate = labdas[1])
T2 = rexp(N, rate = labdas[2])
#print(mean(T1)

bigT = matrix(0, nrow = 2, ncol = N)
bigT[1,] = T1
bigT[2,] = T2
#print(T)

server = rep(0, ncol(bigT))

for (i in 1:ncol(bigT)) {
  server[i] = which.min(bigT[,i])

#print(server)

#BH.8.36.b
print(mean((server == 1)))

#BH.8.36.c
#print(labdas[server])
S = rexp(N, rate = 1) / labdas[server] #this
#print(S)

D = apply(bigT, MARGIN = 2, FUN = min) + S
#print(D)

print(mean(D))
print(sd(D))
```

Ex 4.12. Explain why `T1` corresponds to a number of service times of the first server.

Ex 4.13. To what do the rows of `T` (`bigT`) correspond?

Ex 4.14. What is the content of `server`? Why do we compute this?

Ex 4.15. Explain how we use the fundamental bridge in line P.21 (R.24) to answer BH.8.36.b.

Ex 4.16. Alice is taken into service by the server that finishes first. We need to simulate the service time that Alice needs at that server. Explain how we do this in line P.25 (R.28). Hint, reread BH.5.5 on the exponential. BTW, this is a good time to reread BH.5.3.

Ex 4.17. Why is `np.min(T, axis=0)` in the Python code (`apply(bigT, MARGIN=2, FUN=min)` in R) the time Alice spends waiting in queue, i.e., the time Alice spends in the post office before her service starts?

Ex 4.18. Why is `D` the departure time of Alice, i.e., the time Alice spends in the post office?

Ex 4.19. Set `N` to 1000 or so, or any other large number to your liking, but not so large that your computer will keep simulating for a month... Compare the values of the simulation to the theoretical result that you have to compute in BH.8.36.c.

Ex 4.20. Run the code for $\lambda_1 = 1$ fixed, and take λ_2 equal to $0.5, 1, 1.5$ and 2 successively. Compute the mean time waiting time and mean sojourn time of Alice, and put your results in a table. Compare the results of the simulation to the theoretical values.

There is an important lesson to learn here. With simulation it is, often, reasonably simple to get numerical answers, but it requires many simulations to see a pattern in the numbers. For patterns, we can better use theory, as theory gives us formulas that show how the output of some model depends on the input.

4.3 *BH.8.54*

Read and solve BH.8.54. Perhaps you should read the hints in the study guide on this problem too. First we consider the case with $r = 1$, and we need to tackle some technical details.

Ex 4.21. Just as in the problem, let $Y = pX/q$. Why is $\mathsf{E}[Y] = 1$?

Ex 4.22. If $Z \sim \text{Exp}(1)$, then explain that $M_Z(s) = 1/(1-s)$ for $s < 1$.

Ex 4.23. Use the mathematical solution of the problem to explain that $M_Y(s) \to M_Z(s)$ as $p \to 0$.

Ex 4.24. Here is code to compare $M_Y(s)$ to $M_Z(s)$. Explain what the function $M(s)$ does. Then explain why in the line marked as 'this' we multiply with $1 - s_i$. Explain why we limit s to the interval $[0, 0.8]$. Finally, explain why the graph goes through the point $x = 0, y = 1$.

Python Code

```python
import numpy as np
from scipy.stats import geom, gamma
import matplotlib.pyplot as plt

np.random.seed(3)

p = 0.1
q = 1 - p
n = 10000


def M(s):
    X = geom(p).rvs(n)
    Y = p * X / q
    return np.exp(s * Y).mean()


num = 30
m = np.zeros(num)
s = np.linspace(0, 0.8, num)
for i in range(num):
    m[i] = M(s[i]) * (1 - s[i])  # this

plt.plot(s, m, label="MY")
plt.tight_layout()
plt.savefig("figures/moments.pdf")
```

R Code

```r
set.seed(3)

p = 0.1
q = 1 - p
n = 100000

M = function(s){
    X = rgeom(n,p) - 1
    Y = p * X / q
    return (mean((exp(s * Y))))
}


num = 30
m = rep(0,num)
s = seq(0,0.8,length.out = num)
for (i in 1:num){
    m[i] = M(s[i]) * (1 - s[i]) # this
}

pdf(file="figures/moments.pdf",
    width = 8, height = 7, bg = "white",
    colormodel = "cmyk", paper = "A4")

plot(s,m, type = "l", ann = FALSE, col = "blue")

dev.off()
```

Ex 4.25. The match is not terrric when $p = 0.1$. Rerun the code, but now with $p = 1/1000$. Which line to you have to change for this? Include your graph. Is the match better now?

Ex 4.26. The problem asks for the general r. What lines of code have to change to deal with general r? Modify the code appropriately, and explain why your code is correct.

4.4 *BH.9.1*

Here is the code. (BTW, for this assignment, you don't have to read all of BH.9, reading the problem suffices. Hence, it is not a typo that this assignment appears in week 4, not week 5.)

Python Code

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

mus = np.array([2, 5, 8])
stds = np.array([0.1, 0.5, 1])


m2 = stds ** 2 + mus ** 2 # this
v = m2.mean() - (mus.mean()) ** 2 # this
print(mus.mean(), v)

routes = []
for i in range(3):
    routes.append(norm(mus[i], stds[i])) # this


num = 200
options = randint(0, 3).rvs(num) # this

times = np.zeros(num)  # this
for i in range(num):
    times[i] = routes[options[i]].rvs() # this

print(times.mean(), times.var())
```

R Code

```r
mus = c(2, 5, 8)
stds = c(0.1, 0.5, 1)

m2 = stds ** 2 + mus ** 2 #this
v = mean(m2) - (mean(mus)) ** 2 #this
print(paste(mean(mus), v))

num = 200
options = sample(1:3, num, replace = TRUE) #this

times = rep(0, num)
for(i in 1:num){
  times[i] = rnorm(1, mus[options[i]],stds[options[i]]) # this
}

print(paste(mean(times),var(times)))
```

Ex 4.27. Explain the lines marked as 'this', i.e., how does the program work?

Ex 4.28. Run the code and explain the results. Are the simulated results in line with the theory?

Ex 4.29. Modify the code so that you can plot the histogram of the `times` vector. Include your code and a figure. Set also a seed equal to, e.g., the day of the month in which you do the assignment.

Ex 4.30. Set μ_3 (the third element in `mus`) to 80 instead of 8. Rerun the code and explain what happens to the mean and the variance. Then change to 800 and include a histogram of the PDF of the times.

Ex 4.31. Change σ_1 from 0.1 to 10. Run the code and describe what you get. What's wrong with such a large standard deviation? Are you sure that the route lengths remain positive?

4.5 *Challenges*

Try this challenge if you would like to get a 10.

4.5.1 *Challenge: Ping pong balls ina Beluga*

This challenge is a continuation of the simulation we did for the Beluga case, 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 earlier 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 4.32. 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 4.33. 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 4.34. 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 4.35. 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 4.36. 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 4.37. 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 4.38. 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 4.39. 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 **??** 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 4.40. 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 **??** where it was quite good.

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

Ex 4.41. Recall that for a non-negative random variable X with finite variance, we define the squared coefficient of variation as $\mathrm{SCV}(X) = \mathsf{V}[X]/\mathsf{E}[X]^2$. Using the SCV, show that the approximations of [4.36] and [4.37] can be rewritten in the following neat way:

$$\mathrm{SCV}(XY) \approx \mathrm{SCV}(X) + \mathrm{SCV}(Y).$$
$$\mathrm{SCV}(1/Z) \approx \mathrm{SCV}(Z).$$

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 4.42. 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 $\frac{1}{\mathsf{E}[Z]}$.

5 ASSIGNMENT 5

5.1 *BH.9.1.7*

The aim is to use a simulator to analyze the expected profit for the game of BH.9.1.7 but for a different prior than the uniform. First, however, we build BH.9.1.7 with the uniform prior.

First read and solve BH.9.1.7.

Ex 5.1. Run the code below. What are the values of `accepted`? What does it represent? How is the gain computed? Why do we take the mean?

Python Code

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

np.random.seed(3)

alpha = 2 / 3
n = 100
V = uniform(0, 1).rvs(n)

bid = 0.5
accepted = bid > alpha * V
gain = (V - bid) * accepted
print(gain.mean())
```

R Code

```r
set.seed(3)

alpha = 2 / 3
n = 100
V = runif(n, 0, 1)

bid = 0.5
accepted =  bid > alpha * V
gain = (V - bid) * accepted
print(mean(gain))
```

Ex 5.2. Here is some code to compute the gain for multiple bids. Explain how the function f works. In other words, how does it handles the formatting of floating point numbers? The rest of the code must be clear to you, so you don't have to explain that.

Printing pretty floating points is something you really have to learn. If you unquote the line marked as 'unquote this line' and rerun the program, you'll see why people like to see formatted numbers.

Hint, search the web (Python or R documentation) on formatting floating point numbers.

Python Code

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


def f(x):  # format floating point numbers
    # return f"{x}"  # unquote this line
    return f"{x:3.3f}"


np.random.seed(3)


alpha = 2 / 3
n = 100
V = uniform(0, 1).rvs(n)

for b in np.arange(0.1, 1, 0.1):
    accepted = b > alpha * V
    gain = (V - b) * accepted
    print(f"{f(b)}, {f(gain.mean())}")
```

R Code

```r
f = function(x) # format floating point numbers
    # return(x) # unquote this line
    return(signif(x, digits = 3))

set.seed(3)

alpha = 2 / 3
n = 100
V = runif(n, 0, 1)

for (b in seq(0.1, 0.9, 0.1)){
    accepted = b > alpha * V
    gain = (V - b) * accepted
    print(paste0(f(b), ", " ,f(mean(gain))))
}
```

Ex 5.3. Use the code of the previous exercise as a starting point to make a graph of the gain as a function of the bid. Compute the bids for 0.05, 0.1, 0.15, etc, i.e., in steps of 0.05. Include your code, a figure, and comment on the graph. Use your student id as seed.

Ex 5.4. It is very easy to compute the expected gain by means of numerical integration. Run the code below and explain how the `res` variable is computed. Is the result similar to the simulation? Once again (so that you really remember what an anonymous function is!), explain very briefly why we use an anonymous function here.

Python Code

```python
import numpy as np
from scipy.integrate import quad

alpha = 2 / 3
n = 100
b = 0.5

res = quad(lambda v: (v - b) * (b > alpha * v), 0, 1)
print(res)
```

R Code

```r
library(cubature)

alpha = 2 / 3
n = 100
b = 0.5

res = adaptIntegrate(function(v){(v - b) * (b > alpha * v)},
      lowerLimit = c(0), upperLimit = c(1))
print(paste(res$integral, res$error))
```

Ex 5.5. Modify the numerical integrator such that the prior on V is Beta(8,2). Include your code and give the result for a bid of 1/2.

5.2 *BH.9.1.9*

This is a nice variation of one-step analysis of the coin throwing example of BH.9.1.9.

We have a mouse that sits on one corner of a cube whose edges are made of wire, and diagonally opposite the mouse there is some cheese. The mouse chooses at random (uniform)

any edge and moves to the next corner. How many steps does it take, on average, for the mouse to reach the cheese (assuming the cheese does not move)?

For ease, we label all corners with the minimal number of edges that need to be traversed from the corner to reach the cheese into states. Thus, the starting corner will be called state 3, then we have a state 2, and 1, and finally state 0 is the corner with the cheese.

Ex 5.6. Explain in the code below how the while loop simulates the moves of the mouse. To understand the line marked as 'this', read what is a, so-called, ternary operator.

Python Code

```python
import numpy as np
from scipy.stats import uniform
import matplotlib.pyplot as plt

move = uniform()


def do_run():
    state = 3
    n_steps = 0
    while state != 0:
        n_steps += 1
        p = move.rvs()
        if state == 3:
            state = 2
        elif state == 2:
            state = 1 if p < 2 / 3 else 3 # this
        elif state == 1:
            state = 2 if p < 2 / 3 else 0
    return n_steps


n = 200
steps = np.zeros(n)
for i in range(n):
    steps[i] = do_run()

print(steps.mean(), steps.std())
plt.hist(steps, density=True)
plt.tight_layout()
plt.savefig("figures/mouse.pdf")
```

R Code

```r
library(ggplot2)

do_run = function(){
  state = 3
  n_steps = 0
  while (state != 0){
    n_steps = n_steps + 1
    p = runif(1, 0, 1)
    if (state == 3){
      state = 2
    }
    else if (state == 2){
      state = ifelse(p < 2 / 3, 1, 3) # this
    }
    else if (state == 1){
      state = ifelse(p < 2 / 3, 3, 0) # this
    }
  }
  return (n_steps)
}

n = 200
steps = rep(0, n)

for (i in 1:n){
  steps[i] = do_run()
}

print(paste(mean(steps),sd(steps)))

pdf(file="figures/mouse.pdf",
    width = 8, height = 7, bg = "white",
    colormodel = "cmyk", paper = "A4")

df1 = data.frame(steps)
ggplot()+
geom_histogram(df1, mapping = (aes(x = steps, y = ..density.., fill = "steps")))+
scale_fill_manual(name = " ", values = c("steps" = "blue"))  +
theme(axis.title.x=element_blank(),axis.title.y = element_blank())

dev.off()
```

5.3 *BH.9.6.1*

Read BH.9.6.1 first. We will redo the computations and check the results with simulation. We will also cover many other topics as we proceed.

Ex 5.7. We need to model the distribution of an individual demand X. As an example, we can use the beta distribution. In the example, μ and σ are assumed as given, but for the beta distribution we need an a and a b. One way is to solve by hand for a and b for given μ and σ. However, we can also be lazy and let the computer do the work. Here is the code to do that.

Explain how this code works. (You can use the relevant documentation on the web on how the function fsolve works.) Why do I use $\mu = 1/2$ and $\sigma^2 = 1/12$ as test case? Are the a and b that comes out of the solver correct? What does the final check do?

Python Code

```
from scipy.stats import beta
from scipy.optimize import fsolve

mu, var = 1 / 2, 1 / 12  # why this test case?


def func(x):
    a, b = x[0], x[1]
    m = a / (a + b)
    v = m * (1 - m) / (a + b + 1)
    return [m - mu, v - var]


a, b = fsolve(func, [2, 2])
print(a, b)
X = beta(a, b)
print(X.stats(moments='mv')) # final check.
```

R Code

```r
library(pracma)

mu = 1 / 2
var = 1 / 12  # why this test case?

func = function(x){
  a = x[1]
  b = x[2]
  m = a / (a + b)
  v = m * (1 - m) / (a + b + 1)
  return (c(m - mu, v - var))
}

output = fsolve(func, c(2,2))$x
a = output[1]
b = output[2]
print(paste(a, b))

X_mean = a / (a + b)
X_var = (X_mean * (1-X_mean )) / (a + b + 1)
print(paste(X_mean, X_var)) # final check

```

Ex 5.8. What do you expect to get when you would choose $\mu = 2$? If it fails, why is that?

Ex 5.9. What do you expect to get when you would choose $\sigma > 1/2$? If it fails, why is that? Hint, when an rv has a support on $[0,1]$, why is the variance at most $1/4$?

Remark 5.10. When setting up simulations and choosing variables, make sure that the variables you choose actually make any sense.

Now we know how to find a and b for given μ and σ, we henceforth assume that $a = 1$ and $b = 2$, as this is easier to work with.

Ex 5.11. Now that we know how to simulate individual demands, we can compute the mean and the variance with the formulas of the book and compare this to simulation.

Explain the lines marked as 'this' of the code. (The rest must be clear for you. If not, then you have to be seriously concerned about your understanding of programming.)

Python Code

```python
import numpy as np
from scipy.stats import poisson, beta
from scipy.optimize import fsolve

np.random.seed(3)

labda = 3
N = poisson(labda)
# print(N.rvs(100).mean()) # why this line?

X = beta(a=1, b=2)
mu, var = X.stats(moments='mv')

days = 1000
demands = np.zeros(days)
for i in range(days):
    demands[i] = X.rvs(N.rvs()).sum()  # this

print(demands.mean(), demands.var())
print(labda * mu, var * labda + mu * mu * labda) # this
```

R Code

```r
set.seed(3)

labda = 3
# print(mean(rpois(100,labda))) # why this line?

days = 1000
demands = rep(0, days)

a = 1
b = 2
mu = a / (a + b)
var = (mu * (1 - mu)) / (a + b + 1)

for (i in 1:days){
  demands[i] = sum(rbeta(rpois(1,labda), 1, 2)) # this
}

print(paste(mean(demands), var(demands)))
print(paste(labda * mu, var * labda + mu * mu * labda)) # this
```

Ex 5.12. Here is the code to compute the mean and variance with Adam and Eve's laws. Explain how it works. In particular:

1. explain the lines marked as 'this';
2. what are the contents of the demand matrix;
3. what is the P;
4. how are EX, EV and VE computed;
5. Note in particular the difference between EXn and EXN, and similarly for the variance. Use BH.9.2.1 to explain why I use a small n and a large N in the naming of the variables.

Remark: I realize that this is a hard piece of code, but your understanding of Eve's law will increase by a lot. Some subtle points about the code that you *don't* have to explain:

1. Suppose the largest amount of customers (jobs) that we saw during the simulation is 7. Then on any day we can have $0, \ldots, 7$ demands, i.e, 8 possible outcomes. Thus, per row in the demands matrix we need to have 8 columns.
2. The numpy function (mean() function in R) that computes the mean gives a warning when it is applied to an empty array. (Just try it to see how it works.) We prevent this by computing the histogram (i.e., the empirical PDF) for $n = 0$ separately. The same reasoning applies to the case when idx has zero length.

Python Code

```python
import numpy as np
from scipy.stats import poisson, beta, rv_discrete

np.random.seed(3)

labda = 3
N = poisson(labda)

X = beta(a=1, b=2)
mu, var = X.stats(moments='mv')

days = 1000
n_jobs = N.rvs(days)
n_max = n_jobs.max() + 1  # see the comments
demands = np.zeros((days, n_max))
for day in range(days):
    n = n_jobs[day]
    demands[day, :n] = X.rvs(n)  # this


P = np.zeros(n_max)
EXn = np.zeros(n_max)
VXn = np.zeros(n_max)
P[0] = np.argwhere(n_jobs == 0).size / days
for n in range(1, n_max):
    idx = np.argwhere(n_jobs == n)
    if idx.size == 0:
        continue
    P[n] = idx.size / days  # this
    EXn[n] = demands[idx, :n].mean() * n  # this
    VXn[n] = demands[idx, :n].var() * n  # this

EXN = rv_discrete(values=(EXn, P))  # this

print("mean:", labda * mu, EXN.mean())

VXN = rv_discrete(values=(VXn, P))  # this
print("EV: ", var * labda, VXN.mean())

VE = EXN.var()
print("VE: ", mu * mu * labda, VE)
```

R Code

```r
library(discreteRV)
set.seed(3)

labda = 3

a = 1
b = 2
mu = a / (a + b)
var = (mu * (1 - mu)) / (a + b + 1)

days = 1000
n_jobs = rpois(days,labda)
n_max = max(n_jobs) + 1 # see the comments
demands = matrix(0,days, n_max) # this
for (day in 1:days){
  n = n_jobs[day]
  if(n == 0) next
  demands[day, 1:n] = rbeta(n, 1, 2) # this
}

P = rep(0, n_max)
EXn = rep(0, n_max)
VXn = rep(0, n_max)
P[1] = length(n_jobs[which(n_jobs == 0)])/days
for(n in 1:n_max){
  idx = which(n_jobs == n - 1)
  if (length(idx) == 0){
    next
  }
  P[n] = length(idx) / days # this
  m =
  EXn[n] = mean(demands[idx, c(1:n)]) * (n - 1)  #this
  VXn[n] = sd(demands[idx, c(1:n)]) * sd(demands[idx, c(1:n)])* (n - 1) #this
}

EXN = RV(outcomes = EXn, probs = P) # this
print(paste("mean:", labda * mu, E(EXN) ))

VXN = RV(outcomes = VXn, probs = P) # this
print(paste("EV: ", var * labda, E(VXN)))

VE = V(EXN)
print(paste("VE: ", mu * mu * labda, VE))
```

5.4 *BH.9.7*

We use simulation to solve BH.9.7. Read it now, i.e., before reading the text below, then read the code below. Note how short this code is; amazing, isn't it?

Python Code

```
import numpy as np
from scipy.stats import uniform
import matplotlib.pyplot as plt

np.random.seed(3)


N = 1000
a, b = 0, 1000000
V = uniform(a, b).rvs(N)

x_range = np.linspace(b / 5, b / 2, num=50)
y = np.zeros_like(x_range)

for i, b in enumerate(x_range):
    payoff = (V - b) * (b >= V / 4)
    y[i] = payoff.mean()


plt.plot(x_range, y)
plt.show()
```

R Code

```r
set.seed(3)

N = 1000
a = 0
b = 1000000
V = runif(N, min = a, max = b)

x_range = seq(b / 5, b / 2, length.out = 50)
y = rep(0, length(x_range))


i = 1
for (b in x_range) {
  payoff = (V - b) * (b >= V / 4)
  y[i] = mean(payoff)
  i = i + 1
}

plot(x_range, y, type = "l", col = "blue")
```

Ex 5.13. For the Python code, use the scipy documentation to explain why $V \sim \text{Unif}([0, 10^6])$. For R, explain the function `runif()`.

Ex 5.14. What are the smallest and the largest value of `x_range`?

Ex 5.15. Run the code above and make a graph. Include the graph in your report, and explain what you see in the graph. For instance, is there a maximum? If so, can you explain where the maximum occurs? Can you explain how the maximum should be?

Ex 5.16. Suppose after seeing the graph of the payoffs, and this graph would only increase, or decrease, how would you change `x_range`? Do you expect to see a maximum?

Ex 5.17. For small N, e.g. N=10, you can get quite strange values. Why is that?

Ex 5.18. Change the acceptance threshold from to $V/4$ to $V/5$ (or $V/6$, or some other value you like), and make a graph of the payoffs. Include the graph in your report.

Ex 5.19. Change the payoff function to e.g $\sqrt{V-b}$, or some weird function that you like particularly such as $\sin|V-b|$ (any non-trivial function goes). Make a graph of the mean and the standard deviation of the payoff. Can you explain your graph?

5.5 *Challenges*

Try this challenge if you would like to get a 10.

5.5.1 *Challenge: Benford's law*

In this exercise, we discuss Benford's law. Recall that the first step towards this law was taken in Lecture 5, Exercise 3. In this exercise, we showed that if X, Y are iid uniform on $[1,10)$, that then the density of $Z = XY$ is given by

$$f_Z(z) = \frac{\log(\min\{10,z\}) - \log(\max\{1,z/10\})}{81} I_{1\le z\le 100}.$$

Note that log denotes the natural logarithm.
Benford's law is a statement about the distribution of the first digit of the product of sufficiently many variables that are iid uniform on $[1,10)$. We first consider the first digit of the product of two such variables, i.e. the first digit of Z.

Ex 5.20. Let K be the first digit of Z. Show that the PMF of K is given by

$$P(K=k) = \frac{9k\log(k) - 9(k+1)\log(k+1) + 9 + 10\log(10)}{81}$$

for $k \in \{1,2,\dots,9\}$.

Ex 5.21. Check that $\sum_{k=1}^{9} P(K=k) = 1$. This can be done nicely by recognizing a *telescoping sum*: many terms cancel because they appear once with a minus and once with a plus.

Another way to derive the first digit of Z is to first divide Z by 10 if $Z \ge 10$. This yields a random variable W with support $[1,10)$. Clearly, the division doesn't affect the first digit. The next exercise asks to derive the resulting density. This can be a bit tricky; you should check your answer by verifying that the distribution of the first digit of W matches the distribution of the first digit of Z.

Ex 5.22. Let $W = Z$ if $1 \le Z < 10$ and $W = \frac{Z}{10}$ if $10 \le Z < 100$. Derive the density of W.

We now turn to the product of more than two (independent) random variables. It would be very tedious to do this analytically, so we will instead use some code. However, to do this we have to approximate the continuous uniform variable by a discrete random variable. We use the discrete uniform distribution on $\{1+0.5\cdot\frac{9}{s}, 1+1.5\cdot\frac{9}{s}, 1+2.5\cdot\frac{9}{s}, \dots, 1+(s-0.5)\cdot\frac{9}{s}\}$; in total this set has s elements. However, a product of two elements from this set may not again be an element of this set. To solve this, we identify all elements of the interval $\left(1+k\cdot\frac{9}{s}, 1+(k+1)\cdot\frac{9}{s}\right)$ with $1+(k+0.5)\cdot\frac{9}{s}$. We now use a loop to approximate the distribution of the product of $p+1$ random variables by looking at all possible values of the product of p random variables and

one additional uniformly distributed random variable. Note that in the code, s is called `steps` and p is called `p_idx`.

Executing the code may take a while. If it takes more than 1 minute, you may decrease `steps`, but please do note that you did so.

Python Code

```python
import math

steps = 900
products = 15
p_unif = [1.0/steps] * steps
p_mat = [p_unif]

for p_idx in range(1, products):
    p_vec = [0] * steps
        for s1 in range(steps):
            for s2 in range(steps):
                product = (1 + (s1 + 0.5)*9/steps) * (1 + (s2 + 0.5)*9/steps)
                prod_probability = p_mat[p_idx - 1][s1] * 1/steps

                if product > 10:
                    product = product/10

                prod_idx = math.floor((product-1)/9 * steps)
                p_vec[prod_idx] += prod_probability

    p_mat.append(p_vec)


p_digits = []
for p_idx in range(products):
    vec = []
    for digit in range(1, 10):
        pd = sum(p_mat[p_idx][((digit-1)*steps//9):(digit*steps//9)])
        vec.append(round(pd, 6))
    p_digits.append(vec)

print(p_digits)
```

R Code

```r
steps <- 900
products <- 15
p_unif <- rep(1/steps, steps)
p_mat <- matrix(0, nrow = steps, ncol = products)
p_mat[, 1] <- p_unif

for (p_idx in 2:products) {
  p_vec <- rep(0, steps)
  for (s1 in 1:steps) {
    for (s2 in 1:steps) {
      product <- (1 + (s1 - 0.5)*9/steps) * (1 + (s2 - 0.5)*9/steps)
      prod_probability <- p_mat[s1, p_idx - 1] *  1/steps

      if (product > 10) {
        product <- product/10
      }

      prod_idx <- ceiling((product-1)/9 * steps)
      p_vec[prod_idx] <- p_vec[prod_idx] + prod_probability
    }
  }
  p_mat[, p_idx] = p_vec
}

p_digits <- matrix(0, nrow = 9, ncol = products)
for (p_idx in 1:products) {
  for (digit in 1:9) {
    pd = sum(p_mat[((digit-1)*(steps/9)+1):(digit*(steps/9)), p_idx])
    p_digits[digit, p_idx] = round(pd, 6)
  }
}
p_digits
```

Ex 5.23. Explain line P.13 (R.12) of the code.

Ex 5.24. Briefly comment on the results for $p = 2$ compared the exact result derived in the first exercise. Why is it important to make this comparison?

When looking at the results for larger p, it seems that the probabilities converge. The limit random variable B then satisfies the property that the first digit of B and the first digit of BU (where $U \sim \text{Unif}(1,10)$) are identically distributed. Proving this is quite challenging (even for the challenge). In addition, we first need to know what the distribution of B is.

To guess the distribution of the first digit of B, we look at the results of our code and try some transformations to see if this yields familiar numbers. It turns out that the first digit M of B has the following distribution:

$$P(M=k)=\log_{10}\left(\frac{k+1}{k}\right),$$

for $k\in\{1,2,\ldots,9\}$.

Ex 5.25. Briefly comment on these exact values of $P(M=k)$ compared to the values for $p=15$ that result from the code. Give two reasons why the code results are not exact. Which reason do you think is the most important?

Ex 5.26. Check that $\sum_{k=1}^{9} P(M=k)=1$. You can again use a telescoping sum.

Besides the theoretical aspects covered in this challenge, Benford's law states that the first digit of numbers of naturally occurring sets that span several orders of magnitude, such as vote counts by county (or municipality), transaction sizes, etc., approximately follow this distribution. Initially this was just seen as an interesting curiosity of no practical value, but recently it has been used in fraud detection. If you're interested, you might check out this YouTube video by Numberphile: https://www.youtube.com/watch?v=XXjlR2OK1kM&ab_channel=Numberphile

6 ASSIGNMENT 6

6.1 *BH.9.25*

This simulation exercise is based on BH.9.25. Please read the exercise first, and then the code below.

Ex 6.1. How does this code work? Use the print statements to find out what S is, and explain why we multiply by 2 and subract 1 in one of the lines marked as 'this'.

Python Code

```
import numpy as np
from scipy.stats import bernoulli

np.random.seed(3)

n = 5
num = 10

p = 0.5
S = bernoulli(p).rvs([num, n]) * 2 - 1 # this
# print(S)

x = np.zeros([num, n])
x[:, 0] = 100  # this
# print(x)
f = 0.25

for i in range(1, n):
    x[:, i] = x[:, i - 1] * (1 + f * S[:, i]) # this

print(x.mean(axis=0), x.std(axis=0))
```

R Code

```r
set.seed(3)

n = 5
num = 10

p = 0.5
S = matrix(rbinom(n * num, size = 1, prob = p), num, n) * 2 - 1 # this
# print(S)

x = matrix(0, num, n)
x[, 1] = 100 # this
#print(x)
f = 0.25

for (i in 2:n) {
  x[, i] = x[, i - 1] * (1 + f * S[, i]) # this
}

print(colMeans(x))
print(apply(x, MARGIN = 2, sd))
```

Ex 6.2. 1. Run the code below.

2. Check the values of `x`. What do you notice?

3. Then compare both figures. You should see that in `Kelly-no-jitter.pdf` many of the markes (the `x` below the y-axis) fall on top of each other. To repair for this, we add some noise (jitter) to the x and y values of the markes. Explain how randomness is used here.

4. Jitter plots, like this one, are meant to give some more insight into the data to check whether we use the right number of bins. Change the number of bins to 2, include the plot with the jitters, and explain how the marks help to see that two bins is much too small.

Python Code

```python
import numpy as np
from scipy.stats import bernoulli, uniform
import matplotlib.pyplot as plt


np.random.seed(3)

n = 10
n_experiments = 50

p = 0.4
S = bernoulli(p).rvs([n, n_experiments]) * 2 - 1  # this
# print(S)

x = np.zeros([n, n_experiments])
x[0, :] = 100  # this
f = 0.25

for i in range(1, n):
    x[i, :] = x[i - 1, :] * (1 + f * S[i, :])  # this

print(x[-1, :].mean(), x[-1, :].std())
print(x[-1, :])

plt.hist(x[-1, :], bins=20, density=True)
plt.plot(x[-1, :], np.full_like(x[-1, :], -0.01), "xk", markeredgewidth=1)
plt.savefig("figures/Kelly-no-jitter.pdf")


x_jitter = x[-1, :] + uniform(loc=-3, scale=6).rvs(size=n_experiments)
y_jitter = np.full_like(x[-1, :], -0.01) + uniform(loc=-0.01, scale=0.02).rvs(
    size=n_experiments
)
plt.plot(x_jitter, y_jitter, "xk", markeredgewidth=1)
plt.savefig("figures/Kelly-jitter.pdf")

```

R Code

```r
library(ggplot2)
set.seed(3)
n = 10
n_experiments = 50

p = 0.4
S = matrix(0, n, n_experiments)
for(i in 1:n){
  S[i,1:n_experiments] = rbinom(n_experiments, 1, p) * 2 -1 # this
}
#print(S)

x = matrix(0, n, n_experiments)
x[1,] = 100
f = 0.25

for(i in 2:n){
  x[i,] = x[i-1,] * (1 + f * S[i,])  # this
}

pdf(file="figures/Kelly-no-jitter.pdf",
    width = 8, height = 7, bg = "white",
    colormodel = "cmyk", paper = "A4")

df1 = data.frame(x[nrow(x),])
colnames(df1) = c("X")
df1$Y = -0.01

p = ggplot()+
  geom_histogram(df1, mapping = aes(x = X, y = ..density..), bins = 20)+
  geom_point(df1, mapping =aes(x=X, y = Y))+
  theme(axis.title.x=element_blank(),axis.title.y = element_blank())
p
dev.off()

pdf(file="figures/Kelly-jitter.pdf",
    width = 8, height = 7, bg = "white",
    colormodel = "cmyk", paper = "A4")

x_jitter = x[nrow(x),] + runif(n_experiments,-3, 3)
df2 = data.frame(x_jitter)
colnames(df2) = c("x_jitter")
df2$y_jitter = -0.01 + runif(n_experiments,-0.01, 0.01)

p + geom_histogram(df2, mapping = aes(x = x_jitter, y = ..density..), bins = 20)+
  geom_point(df2, mapping =aes(x=x_jitter, y = y_jitter))+
  theme(axis.title.x=element_blank(),axis.title.y = element_blank())
p
dev.off()
```

6.2 *BH.9.37*

With this exercise we illustrate why (and how) to use bootstrapping. At first it appears a bit strange: we have a set of observations (samples) $X = \{X_1, X_2, \ldots, X_n\}$. For this we can compute the mean, standard deviation, and the empirical distribution with standard procedures, so why sample (bootstrap) from X?

One type of question that seems hard to answer by classical means is to characterize the median, in particular the standard deviation and distribution of the median, of the population (that is, the population from which we took the sample X).

Bootstrapping is nowadays used a lot in data science.

Ex 6.3. Run the code below, and then explain what is `sample`, `bootstrap` and `medians`. Why do we use `axis=1`?

Ex 6.4. Change some numbers (such as the range of the random numbers from which we obtained the initial sample). Include your code, make the plot, and include it in your assignment.

Python Code

```python
import numpy as np
import matplotlib.pyplot as plt

np.random.seed(3)

sample = np.random.randint(200, size=100)
n_boot = 5000
bootstrap = np.random.choice(sample, replace=True, size=(n_boot, len(sample)))

medians = np.median(bootstrap, axis=1)
# print(medians.mean(), medians.std())
# print(np.percentile(medians, [2.5, 97.5]))

plt.hist(medians, bins=100, density=True)
plt.axvline(np.percentile(medians, 2.5), c='r', lw=2)
plt.axvline(medians.mean(), c='r', lw=2)
plt.axvline(np.percentile(medians, 97.5), c='r', lw=2)
plt.savefig("figures/bootstrap.pdf")
```

R Code

```r
library(ggplot2)
library(matrixStats)

np.random.seed(3)

sample = sample(1:200,100, replace=TRUE)
n_boot = 5000
bootstrap = matrix(0,n_boot,length(sample))
for(i in 1:nrow(bootstrap)){
  bootstrap[i,] = sample(sample, length(sample), replace = TRUE)
}

medians = rowMedians(bootstrap)
#print(paste(mean(medians), sd(medians)))
#print(quantile(medians, c(.025, 0.975)))

pdf(file="figures/bootstrap.pdf",
    width = 8, height = 7, bg = "white",
    colormodel = "cmyk", paper = "A4")

df1 = data.frame(medians)
ggplot()+
geom_histogram(df1, mapping = aes(x = medians, y = ..density..), bins = 100)+
geom_vline(xintercept = quantile(medians, 0.025), color = "red", size=2) +
geom_vline(xintercept = quantile(medians, 0.9725), color = "red", size=2)+
geom_vline(xintercept = mean(medians), color = "red", size=2)

dev.off()
```

Ex 6.5. Finally, let us compare the information we obtained from bootstrap to the medians we would obtain when we would sample from the real population.

1. What is here the real population?
2. In a real life situation, what is less costly: real sampling or bootstrapping?
3. Explain the ideas of the code below.
4. Compare this figure to the one obtained by bootstrapping. Comment on major differences.

Python Code

```python
import numpy as np
import matplotlib.pyplot as plt

np.random.seed(3)

samples = np.random.randint(200, size=(500, 100))
medians = np.median(samples, axis=1)
plt.hist(medians, bins=100, density=True)
plt.savefig("figures/real_medians.pdf")
```

R Code

```r
library(ggplot2)
library(matrixStats)

set.seed(3)

sample = matrix(sample(200, 500 * 100, replace = T), nrow = 500, ncol = 100)
medians = rowMedians(sample)

pdf(file="figures/real_medians.pdf",
    width = 8, height = 7, bg = "white",
    colormodel = "cmyk", paper = "A4")

df1 = data.frame(medians)
ggplot()+
geom_histogram(df1, mapping = aes(x = medians, y = ..density..), bins = 100)

dev.off()
```

6.3 *BH.9.50*

This simulation exercise is based on BH.9.50. Please read the exercise first. Do the exercise while thinking about the code.

Ex 6.6. Here is some code to simulate N. Explain the non-trivial steps. (By now you should have an idea about boiler-plate code and what code is specific for the problem.)

Python Code

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

np.random.seed(3)

num = 100
N = np.zeros(num)
Labda = expon(scale=1)
labdas = Labda.rvs(num)
for i in range(num):
    labda = labdas[i]
    N[i] = poisson(labda).rvs()

print(N.mean(), N.var())
adam = Labda.mean()
eve = Labda.mean() + Labda.var()
print(adam, eve)
```

R Code

```r
library(ggplot2)
library(matrixStats)

set.seed(3)

num = 100
N = rep(0,num)
labdas = rexp(num, 1)

for(i in 1:num){
  labda = labdas[i]
  N[i] = rpois(1,labda)
}

adam = 1/1
eve = 1/1 + 1/(1 * 1)
print(paste(adam, eve))
```

Ex 6.7. Run the above code and compare the result to the theoretical value. What do you observe? Rerun the code with `num = 10000`. Does this result in a better estimate? What do you learn from this?

Ex 6.8. The next step is to estimate the mean and variance of the total claim size S. Here is the code, run it and explain the relevant parts.

Python Code

```python
import numpy as np
from scipy.stats import poisson, expon, lognorm

np.random.seed(3)

num = 100
Labda = expon(scale=1)
labdas = Labda.rvs(num)
mu, sigma = 3, 1
X = lognorm(loc=mu, s=sigma)
S = np.zeros(num)
for i in range(num):
    N = poisson(labdas[i]).rvs()
    S[i] = X.rvs(N).sum()

print(f"{S.mean()=}")
adam = X.mean() * Labda.mean()
print("mean_theoretical: ", adam)
print(f"{S.var()=}")
eve = Labda.mean() * X.var() + X.mean() ** 2 * (Labda.mean() + Labda.var())
print("var_theoretical: ", eve)
```

R Code

```r
set.seed(3)

num = 100
labdas = rexp(num, 1)
mu = 3
sigma = 1
S = rep(0,num)

for(i in 1:num){
  N = rpois(1,labdas[i])
  S[i] = sum(rlnorm(N, mean = 0, sd = sigma)+ mu)
}

theta = exp(sigma * sigma / 2)
var = (theta * theta) * (exp(sigma * sigma) - 1)

print(paste("S.mean()=", mean(S)))
adam = (theta + mu) * (1/1)
print(paste("mean_theoretical: ", adam))
print(paste("S.var()=", var(S)))
eve = (1/1) * var + (theta + mu) ** 2 * (1/1 + 1/1/1)
print(paste("var_theoretical: ", eve))
```

Ex 6.9. Here is the code to plot the distribution. Change λ such that it is $\sim$ Exp(1/2) and include the figure.

Python Code

```python
import numpy as np
from scipy.stats import poisson, expon
import matplotlib.pyplot as plt

np.random.seed(3)

num = 100
N = np.zeros(num)
Labda = expon(scale=1)
labdas = Labda.rvs(num)
for i in range(num):
    labda = labdas[i]
    N[i] = poisson(labda).rvs()

plt.hist(N, density=True)
plt.savefig("figures/bh-9.50-fig.pdf")
```

R Code

```r
set.seed(3)

num = 100
N = rep(0,num)
labdas = rexp(num, 1)
for(i in 1:num){
  labda = labdas[i]
  N[i] = rpois(1,labda)
}

pdf(file="figures/bh-9.50-fig.pdf",
    width = 8, height = 7, bg = "white",
    colormodel = "cmyk", paper = "A4")

df1 = data.frame(N)
ggplot()+
  geom_histogram(df1, mapping = (aes(x = N, y = ..density..)))

dev.off()
```

Ex 6.10. Finally, let's do the integration numerically to compute $\mathsf{P}\{N = 3\}$. Explain the code. Why, in particular, do I compare the result against `geom(1/2).pmf(4)`? (Hint, read the documentation of `scipy.stats.geom` (`dgeom()` function in R) to understand why.)

Python Code

```python
import numpy as np
from scipy.integrate import quad
from scipy.stats import expon, geom

Labda = expon(scale=1)

def integrand(labda):
    return np.exp(-labda) * labda ** 3 / 6 * Labda.pdf(labda)

res = quad(lambda labda: integrand(labda), 0, np.infty)
print(res)
print(geom(1 / 2).pmf(3))
print(geom(1 / 2).pmf(4))
```

R Code

```r
library(cubature)

integrand = function(labda){
  return (exp(-labda) * labda ** 3 / 6 * exp(-labda) )
}

res = adaptIntegrate(function(x){integrand(x[1])}, lowerLimit = 0, upperLimit = Inf)
print(res)
print(dgeom(3 - 1, 1 / 2))
print(dgeom(4 - 1, 1 / 2))
```

6.4 *BH.10.2.3*

Let us try to understand the weak law of large numbers by means of simulation. An easy example is to take $X_i \sim \text{Unif}(0,1)$, so that is what we do here.

Python Code

```python
import numpy as np
from numpy.random import uniform

np.random.seed(3)

n = 10
N = 50 # num samples

mu = 1 / 2
var = 1 / 12
eps = 0.1

X = uniform(size=[num_samples, n])

Y = X.mean(axis=1)

larger = np.abs(Y - mu) > eps
count = larger.sum()
P = count / N
RHS = var / (n * eps * eps)
print(P, RHS)
```

R Code

```r
set.seed(3)

n = 10
N = 50 # num samples

mu = 1 / 2
var = 1 / 12
eps = 0.1

X = matrix(runif(N * n), N, n)

Y = rowMeans(X)

larger = abs(Y - mu) > eps
count = sum(larger)
P = count / N
RHS = var / (n * eps * eps)
print(P)
print(RHS)
```

Ex 6.11. Explain `mu` and `var`.

,

Ex 6.12. What is `Y`? What is the symbol that BH use for this?

Ex 6.13. What are the meanings of `larger` and `count`?

Ex 6.14. What is `RHS` in the notation of BH?

Ex 6.15. What inequality of BH do we check by printing `RHS` and `P`?

Ex 6.16. Choose some different values for n and the sample size N. Is the inequality always true?

6.5 *Challenges*

Try this challenge if you would like to get a 10.

6.5.1 *Challenge: Betting*

Consider the setting of BH.9.25, which you also studied in the coding section. We use the notation from that exercise. In this exercise we will discuss how to set f, the betting fraction. In particular, we will discuss the *Kelly criterion*, which states that the betting fraction should be $f = 2p - 1$ if $p > \frac{1}{2}$ is the winning probability.

We discuss its relationship to expected utility theory, which you will also study in Introduction to Mathematical Economics. Expected utility theory states that bets should be chosen to maximize expected utility. So we solve $\max_{0 \le f \le 1} \mathsf{E}[U(X_{n+1}) \mid X_n]$.

Ex 6.17. Show that solving the maximization problem for the utility function $U(x) = \log(x)$ yields the betting fractions from the Kelly criterion, $f = 2p - 1$ if $p > \frac{1}{2}$ and $f = 0$ if $p \le \frac{1}{2}$.

Other people may have a different utility function, which yields a different betting fraction.

Ex 6.18. Calculate the utility maximizing betting fraction f if $U(x) = \sqrt{x}$.

Note that for both of these utility functions, the betting fraction f does not depend on the wealth X_n before the gamble, but in general f does depend on X_n.

Now that we have two different betting fractions, we compare them. For that, we first need the following result:

Ex 6.19. Assume that f does not depend on X_n. Let $x_0 = 1$. Show that there exist constants a, b such that $\log(X_n) = aW + b$ and $W \sim \text{Bin}(n, p)$, and determine a and b in terms of f.

Theorem 10.3.6. states that (for sufficiently large n) we can approximate a random variable with the binomial distribution $W \sim \text{Bin}(n, p)$ by a random variable with the normal distribution $\text{Norm}(np, np(1-p))$. While you will only learn about the proof of this next week, we are already going to use this approximation here.

Ex 6.20. Two people (Carl and Daria) participate in n rounds of this betting game. Their games are independent. Carl's initial wealth is $x_0 = 1$ and Daria's initial wealth is $y_0 = 1$. We denote Carl's wealth after n rounds by X_n and Daria's wealth after n rounds by Y_n. Carl chooses f according to the Kelly criterion, i.e. $f = 2p - 1$. Daria chooses f to be the utility maximizing betting fraction for $U(x) = \sqrt{x}$. Use the previously mentioned normal approximation to derive an approximation for the difference $\log(Y_n) - \log(X_n)$.

Kelly's criterion does not mention utility functions, it just recommends to set $f = 2p - 1$ regardless of one's utility function. The next exercise is meant to give some insight why.

Ex 6.21. Use `pnorm` in R, or `norm.cdf` in Python, to approximate $P(X_n > Y_n)$ for some chosen values for n and p. What do you think that happens if $n \to \infty$ for a fixed p? Explain why this is an argument to use the Kelly criterion regardless of one's utility function. Also, explain why maximizing utility suggests a different f in spite of this result.

7 ASSIGNMENT 7

7.1 *BH.10.9*

Ex 7.1. In the code below we compute or simulate the various expressions of BH.10.9. Check the results and compare it to the analytical results. Are both results in line? If not, do you know why that might be the case? What can you do about that?

I don't know how to build $\mathsf{E}[\mathsf{V}[Y|X]]$ in any sensible way. Since Y is independent of X, the conditioning does not have any influence. Thus, this numerical exercise is skipped.

Python Code

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

np.random.seed(3)

X = expon(2)
Y = expon(2)

print(np.exp(-X.mean()), X.expect(lambda x: np.exp(-x)))

n_sample = 1000
X_sample = X.rvs(n_sample)
Y_sample = Y.rvs(n_sample)
p1 = sum(X_sample > Y_sample + 3) / n_sample
p2 = sum(Y_sample > X_sample + 3) / n_sample
print(p1, p2)

p3 = sum(X_sample > Y_sample - 3) / n_sample
print(p1, p3)

EXY = X_sample @ Y_sample / n_sample
EX4 = X.expect(lambda x: x ** 4)
print(EX4, EXY * EXY)

p = sum(np.abs(X_sample + Y_sample) > 3) / n_sample
print(p, X.mean())
```

R Code

```r
library(cubature)
set.seed(3)

expectation = adaptIntegrate(function(x){exp(-(x[1]+2)) * exp(-x[1]) },
              lowerLimit = c(0), upperLimit = c(Inf))
print(paste(exp(-mean(rexp(10000,1)+2)), expectation$integral))

n_sample = 1000
X_sample = rexp(n_sample, 1) + 2
Y_sample = rexp(n_sample, 1) + 2
p1 = sum(X_sample > Y_sample +  3) / n_sample
p2 = sum(Y_sample > X_sample + 3) / n_sample
print(paste(p1, p2))

p3 = sum(X_sample > Y_sample - 3) / n_sample
print(paste(p1, p3))

EXY = X_sample %*% Y_sample / n_sample
expectation2 = adaptIntegrate(function(x){((x[1]+2) ** 4) * exp(-x[1]) },
              lowerLimit = c(0), upperLimit = c(Inf))

print(paste(expectation2$integral, EXY * EXY))

p = sum(abs(X_sample + Y_sample) > 3) / n_sample
print(paste(p, (2) + 1))
```

Ex 7.2. Why do I use simulation to estimate $\mathsf{E}[XY]$? Recall an earlier assignment in which we used numerical integration. How successful was that?

Ex 7.3. What would you do to increase the credibility of the claims? (Note these claims are based on case checking and simulation? There is no actual proof, however, they can act as counter-examples.)

7.2 *BH.10.28*

Ex 7.4. Let us first plot the PMF of the standardized version of X_n for $n = 10$. Explain the code below BEFORE making the plot.

Python Code

```python
import numpy as np
from scipy.stats import poisson, norm
import matplotlib.pyplot as plt

n = 10
X = poisson(n)
N = norm()

pmf = np.zeros(2 * n)
for i in range(len(pmf)):
    pmf[i] = X.pmf(i)
supp = (np.linspace(0, len(pmf), num=len(pmf)) - n) / np.sqrt(n)


plt.plot(supp, pmf)  # this
plt.plot(supp, norm.pdf(supp))
plt.savefig("figures/bh-10-28.pdf")
```

R Code

```r
library(ggplot2)

n = 10

pmf = dpois(1:(2 * n), n)
supp = (seq(0, length(pmf), length.out = length(pmf)) - n) / sqrt(n)

pdf(file="figures/bh-10-28.pdf",
    width = 8, height = 7, bg = "white",
    colormodel = "cmyk", paper = "A4")

df1 = data.frame(supp)
df1$pmf = pmf # this
df1$norm_pdf = dnorm(supp,0,1)

p = ggplot(df1) +
    geom_line(mapping =(aes(x=supp, y = pmf)), color = "blue")+
    geom_line(mapping =(aes(x=supp, y = norm_pdf)), color = "orange")+
    theme(axis.title.x=element_blank(),axis.title.y = element_blank())
p

dev.off()
```

Ex 7.5. Now make the plot and explain what is wrong.

Ex 7.6. The line marked as 'this' should be replaced with

Python Code

```
plt.plot(supp, pmf * np.sqrt(n))
```

R Code

```
df1$pmf = pmf * sqrt(n)
```

Explain why. (You should memorize that comparing PMFs and PDFs is not straightforward.)

Ex 7.7. Make the plot of $n = 2$ and $n = 20$. Include both plots and explain the results.

Ex 7.8. The code below computes $M_Y(s)$ where Y is the standardized version of X_n. Then it compares M_Y to the MGF of the standard normal distribution.

1. Explain how it works.

Python Code

```
import numpy as np
from scipy.stats import poisson, norm
import matplotlib.pyplot as plt

n = 10
X = poisson(n)
num = 50
S = np.linspace(-1, 1, num)
M = np.zeros(num)
pmf = X.pmf(range(100)) # this

for i in range(num):
    for j in range(len(pmf)):
        M[i] += np.exp(S[i] * (j - n) / np.sqrt(n)) * pmf[j]


plt.plot(S, M, label="M X")
plt.plot(S, np.exp(S * S / 2), label="M N")
plt.xlabel("s")
plt.legend()
plt.savefig("figures/bh-10-28-mgf1.pdf")
```

R Code

```r
library(ggplot2)

n = 10
num = 50
S = seq(-1, 1, length.out = num)
M = rep(0, num)
pmf = dpois(1:100,n) #this

for(i in 1:num){
  for(j in 1:length(pmf)){
    M[i] = M[i] + exp(S[i] * (j - n)/sqrt(n)) * pmf[j]
  }
}

pdf(file="figures/bh-10-28-mgf1.pdf",
    width = 8, height = 7, bg = "white",
    colormodel = "cmyk", paper = "A4")

df1 = data.frame(S)
df1$M_X = M
df1$M_N = exp(S * S / 2)

p = ggplot(df1) +
    geom_line(mapping =(aes(x=S, y = M_X, color = "M X")))+
    geom_line(mapping =(aes(x=S, y = M_N, color = "M N")))+
    scale_color_manual(name = " ", values = c("M X" = "blue", "M N" = "orange"))+
    theme(axis.title.y = element_blank())
p

dev.off()
```

Ex 7.9. The line marked as 'this' computes the PMF upfront. Why is that a good idea? Why do we stop the computation of the PMF at 100?

Ex 7.10. In the code below, one of the for loops is removed.

1. What is the change?
2. Will this change have an effect on the speed of the computation?
3. Which of the two alternatives do you find more readable?

Python Code

```python
import numpy as np
from scipy.stats import poisson, norm
import matplotlib.pyplot as plt

n = 10
X = poisson(n)
num = 50
S = np.linspace(-1, 1, num)
M = np.ones(num)

for i in range(num):
    M[i] = X.expect(lambda j: np.exp(S[i] * (j - n) / np.sqrt(n)))

plt.plot(S, M, label="M X")
plt.plot(S, np.exp(S * S / 2), label="M N")
plt.xlabel("s")
plt.legend()
plt.savefig("figures/bh-10-28-mgf2.pdf")
```

R Code

```r
library(ggplot2)

n = 10
num = 50
S = seq(-1, 1, length.out = num)
M = rep(1, num)

for(i in 1:num){
  j = 1:10000
  X = exp(S[i] * (j-n) / sqrt(n))
  M[i] = weighted.mean(X,dpois(j,n))
}

pdf(file="figures/bh-10-28-mgf2.pdf",
    width = 8, height = 7, bg = "white",
    colormodel = "cmyk", paper = "A4")

df1 = data.frame(S)
df1$M_X = M
df1$M_N = exp(S * S / 2)

p = ggplot(df1) +
    geom_line(mapping =(aes(x=S, y = M_X, color = "M X")))+
    geom_line(mapping =(aes(x=S, y = M_N, color = "M N")))+
    scale_color_manual(name = " ", values = c("M X" = "blue", "M N" = "orange"))+
    theme(axis.title.y = element_blank())
p

dev.off()
```

Ex 7.11. Change n to 20 and to 50. Does the quality improve? Choose some value for n that you like, and include the graph.

7.3 *BH.10.30*

Ex 7.12. Explain how the code below implements the solution of parts a and b BH.10.30. (Of course we skip the proof of the convergence to $g(\alpha)$ here.)

Python Code

```python
import numpy as np
from scipy.stats import bernoulli
import matplotlib.pyplot as plt

np.random.seed(3)

num = 100
p, alpha = 0.5, 0.5
revenue = 0.5 + 1.2 * bernoulli(p).rvs(num)

Y = np.ones(num)
Y[0] = 1
for i in range(1, num):
    Y[i] = (1 - alpha + alpha * revenue[i]) * Y[i - 1]

print(np.log(Y[-1]) / num)

# plt.plot(Y) # this

def g(alpha):
    return 0.5 * np.log((1 + 0.7 * alpha) * (1 - 0.5 * alpha))


plt.ylim(-10 * g(alpha), 10 * g(alpha)) # this
plt.axhline(g(alpha), c='r', lw=2, label="g")
plt.plot(np.log(Y)[1:] / range(1, num), label="(log Yn)/n")
plt.legend()
plt.savefig("figures/bh-10-30.pdf")
```

R Code

```r
library(cubature)
set.seed(3)

num = 100
p = 0.5
alpha = 0.5
revenue = 0.5 + 1.2 * rbinom(num, 1, p)

Y = rep(1, num)
Y[1] = 1
for(i in 2:num){
  Y[i] = (1 - alpha + (alpha * revenue[i])) * Y[i - 1]
}

print(log(Y[length(Y)]) / num)

# plot(Y, type = "l")

g = function(alpha){
  return(0.5*log((1 + 0.7 * alpha)*(1 - 0.5 *alpha)))
}

pdf(file="figures/bh-10-30.pdf",
    width = 8, height = 7, bg = "white",
    colormodel = "cmyk", paper = "A4")

df = data.frame(log(Y)/(1:num))
colnames(df) = c("lnY")
p = ggplot()+
    geom_hline(mapping = aes(yintercept = g(alpha), color = "g"), size = 2)+
    geom_line(df,mapping=aes(x = 1:num, y =lnY, color = '(log Yn)/n')) +
    scale_color_manual( name = " ", values = c("g" = "red", "(log Yn)/n" = "blue"))+
    ylim(-10* g(alpha), 10 * g(alpha)) + #this
    theme(axis.title.x = element_blank(), axis.title.y = element_blank())
p

dev.off()
```

Ex 7.13. Uncomment the relevant line to obtain a plot of Y_n as a function n. Make the plot and include it in your assigment.

Ex 7.14. Take `num` much larger, e.g., 100 000. What is $(\log Y_n)/n$ now?

Ex 7.15. Print $(\log Y_n)/n$ for various values of α. For instance, take $\alpha = 1/10, 2/10, 3/10, 4/10$.

Ex 7.16. If you like a simple challenge, include a plot of $(\log Y_n)/n$ as a function of α. However, skip this if you don't have time, or interest in this extension of the problem. (Perhaps, if you find programming hard, you should do it to improve your skills :-))

7.4 *BH.10.39*

BH.10.39.a asks about the first time such that some exponential rv exceeds a certain threshold. Part b is about when the sum of a number of r.v.s exceed a threshold. Such problems are called *hitting times*.

Python Code

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

np.random.seed(3)

X = expon(scale=1)

# part a

N = 10
res = np.zeros(N)
for i in range(N):
    n = 0
    while X.rvs() < 1:
        n += 1
    res[i] = n

print(res.mean(), res.std())

# part b

N = 100
res = np.zeros(N)
for i in range(N):
    M = X.rvs()
    n = 1
    while M < 10:
        M += X.rvs()
        n += 1
    res[i] = n

print(res.mean(), res.std())
```

R Code

```r
set.seed(3)

# part a

N = 10
res = rep(0, N)
for (i in 1:N) {
  n = 0
  while (rexp(1, rate = 1) < 1) {
    n = n + 1
    res[i] = n
  }
}

print(mean(res))
print(sd(res))

# part b

N = 100
res = rep(0, N)
for (i in 1:N) {
  M = rexp(1, rate = 1)
  n = 1
  while(M < 10) {
    M = M + rexp(1, rate = 1)
    n = n + 1
    res[i] = n
  }
}

print(mean(res))
print(sd(res))
```

By now you have so much experience with reading code that you must be able to explain all of the code without intermediate steps to guide you.

Ex 7.17. Explain how the first part of the code simulates BH.10.39.a.

Ex 7.18. Explain how the second part of the code simulates BH.10.39.b.

7.5 *Challenges*

Try this challenge if you would like to get a 10.

7.5.1 *Challenge: Records*

In BH.7.48 you looked at the number of records in high jumping. Let X_j be how high the jth jumper jumped. As in that exercise, we assume that $X_1, X_2, \ldots$ are iid. with a continuous distribution and say that the jth jumper sets a record if X_j is larger than X_i for all $1 \leq i \leq j-1$. Let X_i^* denote the ith record, i.e., the height of highest jump for the first i jumps. We write f_X and F_X for the PDF and CDF of the iid jumping heights, and X for a random variable with density f_X. Finally, we write G_X for the survivor function of X, i.e. $G_X(x) = 1 - F_X(x)$.

It is not necessary to know the solution of BH.7.48 to do the challenge. In the challenge, we instead look at the distribution of the ith record, the expectation of the ith record and the expected improvement of the record: $\mathsf{E}\left[X_{n+1}^* - X_n^*\right]$.

Ex 7.19. Let $f_{X_{i+1}^*, X_i^*}$ be the joint PDF of the $(i+1)$th and the ith record. Prove that

$$f_{X_{i+1}^*, X_i^*}(u, v) = \frac{f_X(u)}{G_X(v)} f_{X_i^*}(v)\, I_{u>v}.$$

Note that $X_1^* = X_1$. We now derive the density of X_2^*, and then proceed with the general case. These are challenging problems, be sure to check out the hints if you are stuck.

Ex 7.20. Prove that

$$f_{X_2^*}(u) = -f_X(u) \log(G_X(u)).$$

Use a substitution. What is the derivative of the survivor function?

Ex 7.21. Prove that

$$f_{X_n^*}(u) = f_X(u) \cdot \frac{(-\log(G_X(u)))^{n-1}}{\Gamma(n)}.$$

If you need to prove something for all natural numbers n, it is always good to try using mathematical induction, especially since we already know something relating X_{n+1}^* and X_n^*. Note that you can again use the same substitution as in the previous exercise. After that, you need another substitution.

In general case, it is hard to compute the integral of $u f_{X_n^*}(u)$ which is required to compute $\mathsf{E}\left[X_n^*\right]$. For the exponential distribution however, it is still possible to find the result analytically.

Ex 7.22. Assume that $X \sim \text{Exp}(\lambda)$. Determine $\mathsf{E}\left[X_n^*\right]$, and hence the expected improvement of the record: $\mathsf{E}\left[X_{n+1}^* - X_n^*\right]$, using the PDF from the previous exercise. Do you recognize the PDF of X_n^* for $X \sim \text{Exp}(\lambda)$?

When X is exponentially distributed, we can also use directly use the properties of the exponential distribution to determine $\mathsf{E}\left[X_{n+1}^* - X_n^*\right]$.

Ex 7.23. In Exercise 6 of Assignment 5 you found an expression for $\mathsf{E}[X \mid X \geq a]$ if $X \sim \text{Exp}(\lambda)$. Use this expression and Adam's law to determine $\mathsf{E}\left[X_{n+1}^* - X_n^*\right]$ if $X \sim \text{Exp}(\lambda)$.

Ex 7.24. Is the exponential distribution a realistic model for record improvements? Why (not)? If not, why is it still good to look at this case? Explain briefly.

We now consider this model for other distributions as well. Although for many distributions finding analytical results is very difficult or impossible, it can still be interesting to make a plot for the record improvements for other distributions.

Ex 7.25. Assume that X has PDF $f_X(x) = 2xe^{-x^2}$ for $x > 0$, and $f_X(x) = 0$ otherwise. Plot $\mathsf{E}\left[X_{n+1}^* - X_n^*\right]$ as a function of n. You may use code for computing the survival function and the expectation, although it is possible (but not recommended) to do it analytically.