Probability distributions EBP038A05: 2020-2021 Assignments

Nicky D. van Foreest and student assistents

January 25, 2021

CONTENTS

GENERAL INFORMATION

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

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

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

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

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

The rules:

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

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

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

1 ASSIGNMENT 1

1.1 *Have you read well?*

Ex 1.1. In your own words, explain what is

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

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

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

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

Ex 1.4. Express Bayes' formula for two rvs X and Y in terms of the joint CDF, i.e., provide a formula.

Ex 1.5. Let X be uniformly distributed on the set $\{0,1,2\}$ and let $Y \sim \text{Bern}(1/4)$; X and Y are independent. Present a contingency table for the X and Y. What is the interpretation of the column sums the table? What is the interpretation of the row sums of the table? Suppose you change some of the entries in the table, are X and Y still independent?

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

Ex 1.7. Apply the chicken-egg story. Families enter a zoo in a given hour. Some families have one child, other two, and so on. What are the 'eggs' in this context, and what is the meaning of 'hatching'?

Ex 1.8. Claim: We have two rvs X and Y on $\mathbb{R}^+$. It is given that $F_{X,Y}(x,y) = F_X(x)F_Y(y)$ for $x,y \leq 1/3$. Then X and Y are necessarily independent.

Ex 1.9. I select a random guy from the street, his height $X \sim \text{Norm}(1.8, 0.1)$, and I select a random woman from the street, her height is $Y \sim \text{Norm}(1.7, 0.08)$. I claim that since I selected the man and the woman independently, their heights are independent. Briefly comment on this claim. (Hint, from this exercise you should memorize this: **independence is a property of the joint CDF, not of the rvs**.)

Ex 1.10. Correct? For any two rvs X and Y on $\mathbb{R}^+$ with marginals F_X and F_Y. Then $\mathsf{P}\{X \leq x, Y \leq y\} = F_X(x)F_Y(x)$.

Ex 1.11. Theorem 7.1.11. What is the meaning of the notation $X|N=n$?

Ex 1.12. Correct? X,Y two discrete rvs with CDF $F_{X,Y}$. We can compute the PDF as $\partial_x\partial_y F_{X,Y}(x,y)$.

1.2 *Exercise at about exam level*

Ex 1.13. This is about the simplest model for an insurance company that I can think of. We start with an initial capital $I_0 = 2$. The company receives claims and contributions every period, a week say. In the ith period, we receive a contribution X_i uniform on the set $\{1, 2, \ldots, 10\}$ and a claim C_i uniform on $\{0, 1, \ldots 8\}$.

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

1.3 *Coding skills*

Ex 1.14. Use simulation to estimate the answer of BH.7.1. Run the code below and explain line 9 of python code or line 7 of the R code.

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

Compare the value of the simulation to the exact value.

```
1  import numpy as np
2
3  np.random.seed(3)
4
5  num = 10
6
7  a = np.random.uniform(size=num)
8  b = np.random.uniform(size=num)
9  success = np.abs(a - b) < 0.25
10 print(a)
11 print(b)
12 print(success)
13 print(success.mean(), success.var())
```

```
1  set.seed(3)
2
3  num <- 10
4
5  a <- runif(num)
6  b <- runif(num)
7  success <- abs(a-b) < 0.25
8  a
9  b
10 success
11 paste(mean(success), var(success))
```

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

Ex 1.15. Let $X \sim \text{Exp}(3)$. Find a simple expression for $\mathsf{P}\{1 < X \leq 4\}$ and compute the value. Then use simulation to check this value. Finally, use numerical integration to compute this value. What are the numbers? Explain lines 11, 21 and 26 of the python code or lines 7, 17 and 23 of the R code.

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

```python
from scipy.integrate import quad

labda = 3

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

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


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


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


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

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

```r
labda <- 3

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

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


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

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


```

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

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

1.4 *Challenges*

Ex 1.16. Consider the again the chicken-egg story (BH 7.1.9): A chicken lays a random number of eggs N an each egg independently hatches with probability p and fails to hatch with probability $q = 1-p$. For $N \sim \text{Pois}(\lambda)$ it is shown in BH 7.1.9 that X and Y are independent and $X \sim \text{Pois}(\lambda p)$ and $Y \sim \text{Pois}(\lambda(1-p))$. This exercise asks for the converse.

Let $X \sim \text{Pois}(\lambda)$ be the number of eggs that hatch and let $Y \sim \text{Pois}(\mu)$ be the number of eggs that do not hatch. Assume that X and Y are independent. Prove $X + Y =: N \sim \text{Pois}(\lambda + \mu)$.

Ex 1.17. Assume that X has the Cauchy distribution.

1. Does $\mathsf{E}\left[\frac{X}{X^2+1}\right]$ exist? If so, find its value.
2. Does $\mathsf{E}\left[\frac{|X|}{X^2+1}\right]$ exist? If so, find its value.

2 ASSIGNMENT 2

2.1 *Have you read well?*

Ex 2.1. What is the difference between 1D LOTUS and 2D LOTUS?

Ex 2.2. Example 7.2.2. Write down the integral to compute $\mathsf{E}\left[(X-Y)^2\right]$. You don't have to solve the integral.

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

2.2 *Exercises at about exam level*

Ex 2.4. Check first BH 7.2.3. When X, Y iid $\sim$ Norm(0,1), then $X - Y \sim$ Norm(0,2). However, when X, Y iid $\sim$ Pois(λ), then prove first that $X + Y \sim$ Pois(2λ), but note that $X - Y$ is not $\sim$ Pois(0). Explain this difference between the Poisson and normal distribution.

Ex 2.5. Derive the results of BH 7.3.6 without smart tricks. Thus, you have to use the fundamental bridge to show that

$$\begin{aligned}
&\mathsf{E}[ML] = \mathsf{E}[X]\,\mathsf{E}[Y] = 1, && \mathsf{E}[M] = 3/2, && \mathsf{E}[L] = 1/2,\\
&\mathsf{E}\left[L^2\right] = 1/4, && \mathsf{E}\left[M^2\right] = 2\,\mathsf{E}\left[X^2\right] - \mathsf{E}\left[L^2\right] = 7/2 &&\\
&\mathsf{V}[M] = \mathsf{E}\left[M^2\right] - (\mathsf{E}[M])^2, && \mathsf{V}[L] = \mathsf{E}\left[L^2\right] - (\mathsf{E}[L])^2. &&
\end{aligned}$$

You can use the document 'Memoryless exercursions' to see how to solve these problems.

Ex 2.6. Take $X \sim$ Unif($\{-2,-1,1,2\}$) and $\eta = X^2$. What is the correlation coefficient of X and η? If we would consider another distribution for X, would that change the correlation?

Ex 2.7. Let X be the result of the throw of a coin. It is given that $\mathsf{P}\{X = H\} = p = 1 - \mathsf{P}\{X = T\}$. When $X = H$, we choose a fair die with 4 sides with values $1,2,3,4$, when $X = T$ we choose a fair die with 6 sides with values $1,\dots,6$. Let Y_i be the value of the ith throw with the die.

1. What is the PMF of X and Y_1?
2. Marginalize the answer of part a to show that $\mathsf{P}\{X = H\} = p$.
3. What is $\mathsf{P}\{Y_1 = 1\}$?
4. What is $\mathsf{P}\{X = H \mid Y_1 = 1\}$?
5. What is $\mathsf{P}\{X = H \mid Y_1 = 1, Y_2 = 2\}$?
6. What is $\mathsf{P}\{X = H \mid Y_1 = Y_2 = \cdots = Y_n = 1\}$?

Ex 2.8. We have a machine that consists of two components. The machine works as long as not both components have failed. Let X_i be the lifetime of component i.

1. What is the interpretation of $\min X_1, X_2$?
2. What is the interpretation of $\max X_1, X_2$?
3. If X_1, X_2 iid $\sim$ Exp(10) (in hours), what is the expected time until the machine fails?
4. If X_1, X_2 iid $\sim$ Exp(10) (in hours), what is the probability that the machine is still 'up' (i.e., not failed) at time $T = 50$?

2.3 *Coding skills*

Ex 2.9. In this exercise we verify the answers of BH.5.6.5. Read this example first.

For the python code below, run it for a small number of sample; here I choose `samples=2` to enable to see the pattern.

1. In line 11 we print the value of `X` in line 10. What is the meaning of `X`?
2. What is the meaning of `T` in line 12?
3. What do we print in line 14?
4. What is `expected`?
5. What is the `cumsum` of `expected`?
6. Now that you understand what is going on, rerun the simulation for a larger number of samples, e.g., 1000.

```
1  import numpy as np
2  from scipy.stats import expon
3
4  np.random.seed(10)
5
6  labda = 4
7  num = 3
8  samples = 2
9
10 X = expon(scale=labda).rvs((samples, num))
11 print(X)
12 T = np.sort(X, axis=1)
13 print(T)
14 print(T.mean(axis=0))
15
16 expected = np.array([labda / ((num - j)) for j in range(num)])
17 print(expected)
18 print(expected.cumsum())
```

Ex 2.10. Let's check BH.7.48. Read and solve it first.

For the python code below, explain how the small function in lines 6 to 13 works. (You should know that `x += 1` is an extremely useful abbreviation of the code `x = x +1`). Then explain the code in lines 25 and 26.

```
1  import numpy as np
2
3  np.random.seed(3)
4
5
```

```python
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())
```

Ex 2.11. Why is the Exponential Distribution so important? This exercise provides some motivation.

At the Paris metro, a train arrives every 3 minutes on a platform. Suppose that 50 people arrive between the departure of a train and an arrival. It seems entirely reasonable to me to model the arrival times of the individual people as distributed on the interval [0,3]. What is the distribution of the inter-arrival times of these people? It turns out to be exponential! Story BH.13.4.2 provides an explanation (It is not forbidden to read the book beyond what you have to do for this course!); here we use simulation to check this fact.

For the python code:

1. Explain the result of line 12.
2. Compare the result of line 12 and 13; explain what is `A[1:]`.
3. Compare the result of line 12 and 14; explain what is `A[:-1`.
4. Explain what is `X`.
5. Why do I compare $1/\lambda$ and `X.mean()`?
6. Recall that when $X \sim \text{Exp}(\lambda)$, that $\mathsf{E}[X] = \sigma(X)$. Hence, what do you expect to see for `X.std()`?
7. Run the code for a larger sample, e.g. 50, and discuss (very briefly) your results.

```python
import numpy as np

np.random.seed(3)


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

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

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

2.4 *Challenges*

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

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

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

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

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

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

(a) Determine the optimal value x^* for x and the corresponding objective value in case d is deterministic.

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

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

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

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

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

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

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

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

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

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

Ex 2.12. To get some more insight into the model, suppose (for now) that $d \sim U\{10, 20\}$. Solve the model, i.e., find the optimal amount x^*. *Hint: First, compute the value of* $\mathsf{E}[v(d,x)]$ *as a function of x. Then find the optimal value of x.*

Ex 2.13. 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, $\Sigma_{12} = \Sigma_{21} = Cov(d_m, d_f) = \rho\sigma_m\sigma_f = 25\rho$. Also, we know that either $\rho \approx 1$ or $\rho \approx -1$.

Ex 2.14. Calculate x^* and the corresponding objective value for the case $\rho = -1$. (Do not read $\rho = 1$, this case is is not simple.)

Ex 2.15. Consider the two extreme cases $\rho = 1$ and $\rho = -1$. In which case will the shop owner have lower expected total costs? Provide a short, intuitive explanation. *Hint: you don't have to compute* x^* *for the case where* $\rho = 1$ *(this is not easy!).*