Probability distributions EBP038A05: 2020-2021 Assignments

Nicky D. van Foreest and student assistents

January 14, 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 $f_{X,Y}(x,y) = 2\,I_{x \leq y}$?

1. Are X and Y independent?

2. Compute $F_{X,Y}(x,y)$.

Ex 1.3. Correct (that is, is the following 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 $X \sim \text{N}(1.8, 0.1)$, and I select a random woman from the street, her height is $Y \sim \text{N}(1.7, 0.08)$. 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 $\text{P}\{X \leq x, Y \leq y\} = F_X(x)F_Y(x)$.

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 Exercises at about exam level

Ex 1.13. We throw an unbiased die with six sides; the result of the ith throw is X_i.

1. What is the sample space of the two throws (X_1, X_2)?

2. What is the joint CDF?

3. What is the joint PMF?

4. Marginalize out X_2 to show that $P\{X_1 = 5\} = 1/6$.

5. Use the fundamental bridge and indicators to compute $P\{X_1 \; X_2\}$.

6. Use the fundamental bridge and indicators to compute $P\{|X_1 - X_2| \; 1\} = 1/6$.

7. Use the fundamental bridge and indicators to compute $P\{|X_1 - X_2| \leq 1\}$.

8. How would you use simulation to estimate $P\{|X_1 - X_2| \leq 1\}$?

Ex 1.14. We select a random married couple (a man and a woman). His height is $X \sim N(1.8, 0.1)$, her height is $Y \sim N(1.7, 0.08)$ in meters.

1. What is the sample space of (X, Y)?

2. If your answer to question 1 is correct, you must have noticed that potentially the height of the man and the woman can be negative. Is this a problem for this model?

3. What is the joint CDF?

4. What is the joint PDF?

5. Marginalize out Y to show that $X \sim N(1.8, 0.1)$.

6. Use the fundamental bridge and indicators to write $P\{X \; Y\}$ as an integral. You don't have to solve the integral.

7. Use the fundamental bridge and indicators to write $P\{|X - Y| \; 0.1\}$ as an integral. You don't have to solve the integral.

Ex 1.15. 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 $P\{I_1 \; 0\}$?

7. What is $P\{I_1' \; 0\}$?

8. What is $P\{I_2 \; 0\}$?

9. What is $P\{I_2' \; 0\}$?

10. Provide an interpretation in terms of the inventory of rice, say, at a supermarket for I_1 and I_1'.

Ex 1.16. 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.

1.3 *Coding skills*

Ex 1.17. Use simulation to estimate the answer of BH.7.1. Run the code below and explain line 9 of python code or line 7 of the R code.

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

Compare the value of the simulation to the exact value.

```python
import numpy as np

np.random.seed(3)

num = 10

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

```r
set.seed(3)

num <- 10

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

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

Ex 1.18. Let $X \sim \text{Exp}(3)$. Find a simple expression for $P\{1 \ X \le 4\}$ and compute the value. Then use simulation to check this value. Finally, use numerical integration to compute this value. What are the numbers? Explain lines 11, 21 and 26 of the python code or lines 7, 17 and 23 of the R code.

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

labda = 3

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

```python
10
11    success = (X > 1) * (X < 4)
12    # print(success)
13    print(success.mean(), success.std())
14
15
16    def F(x):  # CDF
17        return 1 - np.exp(-labda * x)
18
19
20    def f(x):  # density
21        return labda * np.exp(-labda * x)
22
23
24    print(F(4) - F(1))
25
26    I = quad(f, 1, 4)
27    print(I)
```

```r
1     labda <- 3
2
3     X <- rexp(1000, rate = labda)
4     # X
5     mean(X)
6
7     success <- (X > 1) * (X < 4)
8     # print(success)
9     paste(mean(success), sd(success))
10
11
12    CDF <- function(x) {  # CDF
13      return(1 - exp(-labda * x))
14    }
15
16    f <- function(x) {    # density
17      return(labda * exp(-labda * x))
18    }
19
20
21    CDF(4) - CDF(1)
22
23    I = integrate(f, 1, 4)
24    I
```

1.4 *Challenges*

Ex 1.19. 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 $N \sim \mathrm{P}(\lambda)$ it is shown in BH 7.1.9 that X and Y are independent and $X \sim \mathrm{P}(\lambda p)$ and $Y \sim \mathrm{P}(\lambda(1 - p))$. This exercise asks for the converse.

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

2 ASSIGNMENT 2, TBD

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 $\mathrm{Corr}(X,X)$? 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 $\sim N(0,1)$, then $X - Y \sim N(0,2)$. However, when X, Y iid $\sim P(\lambda)$, then prove first that $X + Y \sim P(2\lambda)$, but note that $X - Y$ is not $\sim P(0)$. 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

$$E[ML] = E[X]E[Y] = 1, \qquad E[M] = 3/2, \qquad E[L] = 1/2,$$
$$E[L^2] = 1/4, \qquad E[M^2] = 2E[X^2] - E[L^2] = 7/2$$
$$V[M] = E[M^2] - (E[M])^2, \qquad V[L] = E[L^2] - (E[L])^2.$$

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

Ex 2.6. Take $X \sim U(\{-2, -1, 1, 2\})$ 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 $P\{X = H\} = p = 1 - P\{X = T\}$. 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 $1, \ldots, 6$. 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 $P\{X = H\} = p$.

3. What is $P\{Y_1 = 1\}$?

4. What is $P\{X = H \mid Y_1 = 1\}$?

5. What is $P\{X = H \mid Y_1 = 1, Y_2 = 2\}$?

6. What is $P\{X = H \mid Y_1 = Y_2 = \cdots = Y_n = 1\}$?

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 $\sim \text{Exp}(10)$ (in hours), what is the expected time until the machine fails?

4. If X_1, X_2 iid $\sim \text{Exp}(10)$ (in hours), what is the probability that the machine is still 'up' (i.e., not failed) at time $T = 50$?

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.

```python
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.

```python
1   import numpy as np
2
3   np.random.seed(3)
4
5
6   def find_number_of_maxima(X):
7       num_max = 0
8       M = -np.infty
9       for x in X:
10          if x > M:
11              num_max += 1
12              M = x
```

```
13          return num_max
14
15
16   num = 10
17   X = np.random.uniform(size=num)
18   print(X)
19
20   print(find_number_of_maxima(X))
21
22   samples = 100
23   Y = np.zeros(samples)
24   for i in range(samples):
25       X = np.random.uniform(size=num)
26       Y[i] = find_number_of_maxima(X)
27
28   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 $X \sim \mathrm{Exp}(\lambda)$, 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.

```
1    import numpy as np
2
3    np.random.seed(3)
4
5
6    num = 5 # small sample at first, for checking.
7    start, end = 0, 3
8    labda = num / (end - start)  # per minute
9    print(1 / labda)
10
11   A = np.sort(np.random.uniform(start, end, size=num))
12   print(A)
13   print(A[1:])
14   print(A[:-1]
```

```
15  X = A[1:] - A[:-1]
16  print(X)
17
18  print(X.mean(), X.std())
```

2.4 *Challenges*

3 ASSIGNMENT 3, TBD

Topics of chapter 8.

3.1   *Have you read well?*

Ex 3.1. Explain in your own words:

1. What is a prior?

2. What is a conjugate prior?

Ex 3.2. Look up on the web: what is the conjugate prior of the multinomial distribution?

3.2 *About exam level*

Ex 3.3. Let X, Y iid $\sim U([0,1])$.

1. What is the joint CDF of $X + Y, XY$?

2. What is the joint PDF of $X +^Y XY$?

3. Compute $P\{X + Y \leq 1, XY \leq 2/9\}$.

3.3 *Coding skills*

Ex 3.4. How many ping pong balls fit into an Airbus Beluga? One way to answer this is as follows. According to this wikipage the cargo volume V of this airplane is $1500\,\text{m}^3$. But this number is based on the physical dimensions that is available to store containers, tanks, and so on. So, I estimate the volume as about twice that amount, i.e., $V = 2500\,\text{m}^3$. The volume of a ping pong ball is $v = 4\pi r^3/3 = 33.49333333333333\,\text{cm}^3$ with $r = 2$ cm. A plain division gives 74.6268656716418 ping pong balls. Note, I left out the 10^6 conversion from meters to cm, and I do not take into the sphere packing factor. Besides that, I hope you agree with me that providing an result with the precision as given here is plain ridiculous. (But from reason incomprehensible to me, even professional econometricians like to report results with 10 digits or more, without questioning the precision.)

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

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

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

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

2. Explain lines 8 and 11 of the python code or lines 5 and 8 of the R code. (If you are interested in both languages, also comment on the difference)

3. Use the code below to provide these estimates.

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

The numerical results suggest the interesting guess $\mathsf{V}[N] \approx \mathsf{V}[V] * \mathsf{V}[v]$, but is this true more generally? In [3.6] we research this problem in more detail.

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

num = 500

np.random.seed(3)

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

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

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

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

```
1   num <- 500
2
3   set.seed(3)
4
5   V = rnorm(num, 2500, 500)
6   v = rnorm(num, 33.5, 0.5)
7
8   N = V / v
9   paste(mean(N), sd(N))
10
11  2500/33.5
12  sqrt(500*0.5)
```

Ex 3.5. The baker of Poincare. TBD

Henri Poincaré was a French mathematician who taught at the Sorbonne around 1900. The following anecdote about him is probably fabricated, but it makes an interesting probability problem. Supposedly Poincaré suspected that his local bakery was selling loaves of bread that were lighter than the advertised weight of 1 kg, so every day for a year he bought a loaf of bread, brought it home and weighed it. At the end of the year, he plotted the distribution of his measurements and showed that it fit a normal distribution with mean 950 g and standard deviation 50 g. He brought this evidence to the bread police, who gave the baker a warning. For the next year, Poincaré continued the practice of weighing his bread every day. At the end of the year, he found that the average weight was 1000 g, just as it should be, but again he complained to the bread police, and this time they fined the baker.

Why? Because the shape of the distribution was asymmetric. Unlike the normal distribution, it was skewed to the right, which is consistent with the hypothesis that the baker was still making 950 g loaves, but deliberately giving Poincaré the heavier ones. Exercise 5.6 Write a program that simulates a baker who chooses n loaves from a distribution with mean 950 g and standard deviation 50 g, and gives the heaviest one to Poincaré. What value of n yields a distribution with mean 1000 g? What is the standard deviation?

3.4 *Challenges*

Ex 3.6. This is a continuation of [3.4]. TBD.

```
Als V[V] = 0 of V[v] = 0, is rechts 0, links niet perse.
Maar er is nog een betere reden waarom dit niet kan kloppen. Definieer V' = 2V, v'=2v en N' = V'/v'

Ik denk dat ik het heb opgelost. Merk op dat we aannemen dat V en v onafhankelijk zijn. Dus
V[V/v] = E[V^2/v^2] - E[V/v]^2 = E[V^2]E[1/v^2] - E[V]^2E[1/v]^2 = V[V]V[1/v] + E[V]^2 E[1/v^2]  +
Als we er van uit gaan dat we V en v voldoende precies weten, geldt er V[V]V[1/v] << E[V]^2 V[1/v] +
De eerste-orde Taylorbenadering van 1/v rond E[v] is 2/E[v] - v/E[v]^2. We kunnen dus benaderen V[1
Dezelfde benadering geeft ook E[1/v]^2 \approx 1/ E[v]^2. Dus het resultaat is al met al  V[V/v]\ap
Ofwel: V[V/v]/E[V/v]^2=  V[v]/E[v]^2 +  V[V]/E[V]^2, dus SCV(V/v) \approx SCV(V)  + SCV(v).
Overigens: de benadering zit hier in SCV(1/v) \approx SCV(v). Immers, SCV(V/v) = SCV(V)  + SCV(1/v)
```

4   ASSIGNMENT 4, TBD

4.1   *Have you read well?*

4.2   *Coding skills*

Ex 4.1. Bekijk het maximale verschil van de verdeling van de som van 3 uniformen, en de normale verdeling.

4.2.1   *Bayesian priors, Testing for rare deceases, Making the plot of Exercise 7.86*

In line with Exercise 8.33, we are now going to analyze the effect on $P\{D|T\}$ 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\Hence, even though it is claimed that PCR tests 'have analytical sensitivity and specificity of greater than 95\Simply put, the specificity and sensitivity are not precisely known, hence this must affect $P\{D|T\}$.

To help you, we show how to make one graph. Then we ask you to make a few on your own, and comment on them.

4.2.2   *Redoing the computation of the Example 2.3.9*

I write `p_D_g_T$` for $P\{D|T\}$. Here is how this can be implemented in python.

```
1  sensitivity = 0.95
2  specificity = 0.95
3  p_D = 0.01
4
5  p_T = sensitivity * p_D + (1-specificity)*(1-p_D)
6  p_D_g_T  = sensitivity * p_D/p_T
7  p_D_g_T
```

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

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

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

4.

4.3   *Compound Poisson distribution, hitting times, and overshoot distribution*

5   ASSIGNMENT 5, TBD

5.1   *Have you read well?*

6   BAYES' BILLIARDS

Take $n = 100$ samples.

6.1 *One coin*

1. Take success probability $p = 1/2$.

2. Make matrix with n rows, n columns. Each row is an experiment of n throws of the coin.

3. Plot the histogram of the number of heads.

6.2   *Three coins*

1. Take three coins with success probabilities $p = 1/4, 1/2, 3/4$.

2. Make a simulation for each coin.

3. If we select a coin with probability $1/3$, the total histogram is the $1/3$ times the sum of the histograms of each of the coins. That is, $P\{X = k\} = P\{X = k|C = i\} P\{C = i\}$, where C is one of chosen coins; here we take $P\{C = i\} = 1/3$.

6.3   *five coins*

1. Select with uniform probability one out of five coins with success probabilities $p = i/5$, $i = 1, \ldots, 5$.

2. Make a simulation for each coin.

3. Make the histogram $\mathsf{P}\{X = k\} = \mathsf{P}\{X = k | C = i\}\, \mathsf{P}\{C = i\}$, where $\mathsf{P}\{C = i\} = 1/5$.

7 ASSIGNMENT 6, TBD

7.1 *Have you read well?*