Probability distributions
Lectures, questions and solutions

Nicky van Foreest and Ruben van Beesten

February 8, 2021

CONTENTS

1 LECTURE 1

Ex 1.1. Consider 12 football players on a football field. Eleven of them are players of FC Barcelone, the other one is an arbiter. We select a random player, uniform. This player must take a penalty. The probability that a player of Barcelone scores is 70%, for the arbiter it is 50%. Let $P \in \{A, B\}$ be r.v that corresponds to the selected player, and $S \in \{0,1\}$ be the score.

1. What is the PMF? In other words, determine $\mathsf{P}\{P = B, S = 1\}$ and so on for all possibilities.
2. What is $\mathsf{P}\{S = 1\}$? What is $\mathsf{P}\{P = B\}$?
3. Show that S and P are dependent.

An insurance company receives on a certain day two claims $X, Y \geq 0$. We will find the PMF of the loss $Z = X + Y$ under different assumptions.

The joint CDF $F_{X,Y}$ and joint PMF $p_{X,Y}$ are assumed known.

Ex 1.2. Why is it not interesting to consider the case $\{X = 0, Y = 0\}$?

Ex 1.3. Find an expression for the PMF of $Z = X + Y$.

Suppose $p_{X,Y}(i, j) = c\, I_{i=j} I_{1\leq i\leq 4}$.

Ex 1.4. What is c?

Ex 1.5. What is $F_X(i)$? What is $F_Y(j)$?

Ex 1.6. Are X and Y dependent? If so, why, because $1 = F_{X,Y}(4,4) = F_X(4)F_Y(4)$?

Ex 1.7. What is $\mathsf{P}\{Z = k\}$?

Ex 1.8. What is $\mathsf{V}[Z]$?

Now take X, Y iid $\sim$ Unif$(\{1,2,3,4\})$ (so now no longer $p_{X,Y}(i, j) = I_{i=j} I_{1\leq i\leq 4}$).

Ex 1.9. What is $\mathsf{P}\{Z = 4\}$?

Remark 1.1. We can make lots of variations on this theme.

1. Let $X \in \{1,2,3\}$ and $Y \in \{1,2,3,4\}$.
2. Take $X \sim$ Pois(λ) and $Y \sim$ Pois(μ). (Use the chicken-egg story)
3. We can make X and Y such that they are (both) continuous, i.e., have densities. The conceptual ideas[1] don't change much, except that the summations become integrals.
4. Why do people often/sometimes (?) model the claim sizes as iid $\sim$ Norm(μ, σ^2)? There is a slight problem with this model (can real claim sizes be negative?), but what is the way out?
5. The example is more versatile than you might think. Here is another interpretation.

   A supermarket has 5 packets of rice on the shelf. Two customers buy rice, with amounts X and Y. What is the probability of a lost sale, i.e., $\mathsf{P}\{X + Y > 5\}$? What is the expected amount lost, i.e., $\mathsf{E}[\max X + Y - 5, 0]$?

   Here is yet another. Two patients arrive in to the first aid of a hospital. They need X and Y amounts of service, and there is one doctor. When both patients arrive at 2 pm, what is the probability that the doctor has work in overtime (after 5 pm), i.e., $\mathsf{P}\{X + Y > 5 - 2\}$?

[1] Unless you start digging deeper. Then things change drastically, but we skip this technical stuff.

2 LECTURE 2

Read the problems of `memoryless\_excursions.pdf`. All the problems in that document relate to topics discussed in Sections BH.7.1 and BH.7.2, and quite a lot of topics you have seen in the previous course on probability theory.

3 LECTURE 3

Ex 3.1. We ask a married woman on the street her height X. What does this tell us about the height Y of her spouse? We suspect that taller/smaller people choose taller/smaller partners, so, given X, a simple estimator $\hat{Y}$ of Y is given by

$$\hat{Y} = aX + b.$$

(What is the sign of a if taller people tend to choose taller people as spouse?) But how to determine a and b? A common method is to find a and b such that the function

$$f(a,b) = \mathsf{E}\left[(Y - \hat{Y})^2\right]$$

is minimized. Show that the optimal values are such that

$$\hat{Y} = \mathsf{E}[Y] + \rho\frac{\sigma_Y}{\sigma_X}(X - \mathsf{E}[X]),$$

where ρ is the correlation between X and Y and where σ_X and σ_Y are the standard deviations of X and Y respectively.

Ex 3.2. Using scaling laws often can help to find errors. For instance, the prediction $\hat{Y}$ should not change whether we measure the height in meters or centimetres. In view of this, explain that

$$\hat{Y} = \mathsf{E}[Y] + \rho\frac{\mathsf{V}[Y]}{\sigma_X}(X - \mathsf{E}[X])$$

must be wrong.

Ex 3.3. N people throw their hat in a box. After shuffling, each of them takes out a hat at random. How many people do you expect to take out their own hat (i.e., the hat they put in the box); what is the variance? In BH.7.46 you have to solve this analytically. In the exercise here you have to write a simulator for compute the expectation and variance.

INDICATORS ARE GREAT functions, and I suspect you underestimated the importance of these functions. They help to keep your formulas clean. You can use them in computer code as logical conditions, or to help counting relevant events, something you need when numerically estimating multi-D integrals for machine learning for instance. And, even though I(=NvF) often prefer to use figures over algebra to understand something, when it comes to integration (and reversing the sequence of integration in multiple integrals) I find indicators easier to use.

Moreover, you should know that in fact, *expectation* is the fundamental concept in probability theory, and the *probability of an event is defined* as

$$\mathsf{P}\{A\} := \mathsf{E}[I_A]. \tag{3.1}$$

Thus, the fundamental bridge is actually an application of LOTUS to indicator functions. REREAD BH.4.4!

Here are some more examples

Ex 3.4. What is $\int_{-\infty}^{\infty} I_{0\leq x\leq 3}\,\mathrm{d}x$?

Ex 3.5. What is

$$\int x\, I_{0\le x\le 4}\, \mathrm{d}x? \tag{3.2}$$

When we do an integral over a 2D surface we can first integrate over the x and then over the y, or the other way around, whatever is the most convenient. (There are conditions about how to handle multi-D integral, but for this course these are irrelevant.)

Ex 3.6. What is

$$\iint xy\, I_{0\le x\le 3}\, I_{0\le y\le 4}\, \mathrm{d}x\, \mathrm{d}y? \tag{3.3}$$

Ex 3.7. What is

$$\iint I_{0\le x\le 3}\, I_{0\le y\le 4}\, I_{x\le y}\, \mathrm{d}x\, \mathrm{d}y? \tag{3.4}$$

4 SOLUTIONS

s.1.1. Here is the joint PMF:

$$\mathsf{P}\{P = A, S = 1\} = \frac{1}{12}0.5 \qquad \mathsf{P}\{P = A, S = 0\} = \frac{1}{12}0.5 \tag{4.1}$$

$$\mathsf{P}\{P = B, S = 1\} = \frac{11}{12}0.7 \qquad \mathsf{P}\{P = B, S = 0\} = \frac{11}{12}0.3. \tag{4.2}$$

Now the marginal PMFs

$$\mathsf{P}\{S = 1\} = \mathsf{P}\{P = A, S = 1\} + \mathsf{P}\{P = B, S = 1\} = 0.042 + 0.64 = 0.683 = 1 - \mathsf{P}\{S = 0\}$$
$$\mathsf{P}\{P = B\} = \frac{11}{12} = 1 - \mathsf{P}\{P = A\}.$$

For independence we take the definition. In general, for all outcomes x, y we must have that $\mathsf{P}\{X = x, Y = y\} = \mathsf{P}\{X = x\}\,\mathsf{P}\{Y = y\}$. For our present example, let's check for a particular outcome:

$$\mathsf{P}\{P = B, S = 1\} = \frac{11}{12} \cdot 0.7 \neq \mathsf{P}\{P = B\}\,\mathsf{P}\{S = 1\} = \frac{11}{12} \cdot 0.683$$

The joint PMF is obviously not the same as the product of the marginals, which implies that P and S are not independent.

s.1.2. When the claim sizes are 0, then the insurance company does not receive a claim.

s.1.3. By the fundamental bridge,

$$\begin{aligned}
\mathsf{P}\{Z = k\} &= \sum_{i,j} I_{i+j=k}\, p_{X,Y}(i,j) &(4.3)\\
&= \sum_{i,j} I_{i,j\geq 0}\, I_{j=k-i}\, p_{X,Y}(i,j) &(4.4)\\
&= \sum_{i=0}^{k} p_{X,Y}(i, k-i). &(4.5)
\end{aligned}$$

s.1.4. $c = 1/4$ because there are just four possible values for i and j.

s.1.5. Use marginalization:

$$\begin{aligned}
F_X(k) &= F_{X,Y}(k,\infty) = \sum_{i\leq k}\sum_j p_{X,Y}(i,j) &(4.6)\\
&= \frac{1}{4}\sum_{i\leq k}\sum_j I_{i=j}\, I_{1\leq i\leq 4} &(4.7)\\
&= \frac{1}{4}\sum_{i\leq k} I_{1\leq i\leq 4} &(4.8)\\
&= k/4, &(4.9)\\
F_Y(j) &= j/4. &(4.10)
\end{aligned}$$

s.1.6. The equality in the question must hold for all i, j, not only for $i = j = 4$. If you take $i = j = 1$, you'll see immediately that $F_{X,Y}(1,1) \neq F_X(1)F_Y(1)$:

$$\frac{1}{4} = F_{X,Y}(1,1) \neq F_X(1)F_Y(1) = \frac{1}{4}\frac{1}{4}. \tag{4.11}$$

s.1.7. $\mathsf{P}\{Z=2\} = \mathsf{P}\{X=1, Y=1\} = 1/4 = \mathsf{P}\{Z=4\}$, etc. $\mathsf{P}\{Z=k\} = 0$ for $k \not\in \{2,4,6,8\}$.

s.1.8. Here is one approach

$$
\begin{aligned}
\mathsf{V}[Z] &= \mathsf{E}\left[Z^2\right] - (\mathsf{E}[Z])^2 && (4.12)\\
\mathsf{E}\left[Z^2\right] &= \mathsf{E}\left[(X+Y)^2\right] = \mathsf{E}\left[X^2\right] + 2\,\mathsf{E}[XY] + \mathsf{E}\left[Y^2\right] && (4.13)\\
(EZ)^2 &= (\mathsf{E}[X] + \mathsf{E}[Y])^2 && (4.14)\\
&= (\mathsf{E}[X])^2 + 2\,\mathsf{E}[X]\,\mathsf{E}[Y] + (\mathsf{E}[Y])^2 && (4.15)\\
&\Longrightarrow && (4.16)\\
\mathsf{V}[Z] &= \mathsf{E}\left[Z^2\right] - (\mathsf{E}[Z])^2 && (4.17)\\
&= \mathsf{V}[X] + \mathsf{V}[Y] + 2(\mathsf{E}[XY] - (\mathsf{E}[X]\,\mathsf{E}[Y])) && (4.18)\\
\mathsf{E}[XY] &= \sum_{ij} ij\, p_{X,Y}(i,j) = \frac{1}{4}(1+4+9+16) = \ldots && (4.19)\\
\mathsf{E}\left[X^2\right] &= \ldots && (4.20)
\end{aligned}
$$

The numbers are for you to compute.

s.1.9.

$$
\begin{aligned}
\mathsf{P}\{Z=4\} &= \sum_{i,j} I_{i+j=4}\, p_{X,Y}(i,j) && (4.21)\\
&= \sum_{i=1}^{4}\sum_{j=1}^{4} I_{j=4-i}\frac{1}{16} && (4.22)\\
&= \sum_{i=1}^{3}\frac{1}{16} && (4.23)\\
&= \frac{3}{16}. && (4.24)
\end{aligned}
$$

s.3.1. We take the partial derivatives of f with respect to a and b, and solve for a and b. In the derivation, we use that

$$
\rho = \frac{\mathsf{Cov}[X,Y]}{\sqrt{\mathsf{V}[X]\,\mathsf{V}[Y]}} = \frac{\mathsf{Cov}[X,Y]}{\sigma_X\sigma_Y} \Longrightarrow \rho\frac{\sigma_Y}{\sigma_X} = \frac{\mathsf{Cov}[X,Y]}{\mathsf{V}[X]}. \qquad (4.25)
$$

Hence,

$$
\begin{aligned}
f(a,b) &= \mathsf{E}\left[(Y-\hat{Y})^2\right] \\
&= \mathsf{E}\left[(Y-aX-b)^2\right] \\
&= \mathsf{E}\left[Y^2\right] - 2a\,\mathsf{E}[YX] - 2b\,\mathsf{E}[Y] + a^2\,\mathsf{E}\left[X^2\right] + 2ab\,\mathsf{E}[X] + b^2 \\
\partial_a f &= -2\,\mathsf{E}[YX] + 2a\,\mathsf{E}\left[X^2\right] + 2b\,\mathsf{E}[X] = 0 \\
&\implies a\,\mathsf{E}\left[X^2\right] = \mathsf{E}[YX] - b\,\mathsf{E}[X] \\
\partial_b f &= -2\,\mathsf{E}[Y] + 2a\,\mathsf{E}[X] + 2b = 0 \\
&\implies b = \mathsf{E}[Y] - a\,\mathsf{E}[X] \\
a\,\mathsf{E}\left[X^2\right] &= \mathsf{E}[YX] - \mathsf{E}[X](\mathsf{E}[Y] - a\,\mathsf{E}[X]) \\
&\implies a(\mathsf{E}\left[X^2\right] - \mathsf{E}[X]\,\mathsf{E}[X]) = \mathsf{E}[YX] - \mathsf{E}[X]\,\mathsf{E}[Y] \\
&\implies a = \frac{\mathsf{Cov}[X,Y]}{\mathsf{V}[X]} = \rho\frac{\sigma_Y}{\sigma_X} \\
b &= \mathsf{E}[Y] - \rho\frac{\sigma_Y}{\sigma_X}\,\mathsf{E}[X] \\
\hat{Y} &= aX + b \\
&= \rho\frac{\sigma_Y}{\sigma_X}X + \mathsf{E}[Y] - \rho\frac{\sigma_Y}{\sigma_X}\,\mathsf{E}[X] \\
&= \mathsf{E}[Y] + \rho\frac{\sigma_Y}{\sigma_X}(X - \mathsf{E}[X]).
\end{aligned}
$$

What a neat formula! Memorize the derivation, at least the structure. You'll come across many more optimization problems.

What if $\rho = 0$?

s.3.2. If we measure X in centimetres instead of metres, then X, $\mathsf{E}[X]$ and σ_X are all multiplied by 100, and the prediction $\hat{Y}$ should also be expressed in centimetres But $\mathsf{V}[Y]$ scales as length squared. This messes up the units.

s.3.3. Let us first do one run.

```python
import numpy as np

np.random.seed(3)

N = 4
X = np.arange(N)
np.random.shuffle(X)
print(X)
print(np.arange(N))
print((X == np.arange(N)))
print((X == np.arange(N)).sum())
```

Here are the results of the print statements: $X = [3102]$. The matches are [False True False False]; we see that $X[1] = 1$ (recall, python arrays start at index 0, not at 1, so $X[1]$ is the second element of X, not the first), so that the second person picks his own hat. The number of matches is therefore 1 for this simulation.

Now put the people to work, and let them pick hats for 50 times.

```python
import numpy as np

np.random.seed(3)

num_samples = 50
N = 5

res = np.zeros(num_samples)
for i in range(num_samples):
    X = np.arange(N)
    np.random.shuffle(X)
    res[i] = (X == np.arange(N)).sum()

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

Here is the number of matches for each round: [0. 1. 1. 0. 1. 0. 1. 0. 1. 1. 1. 2. 2. 1. 0. 1. 1. 1. 0. 2. 0. 1. 2. 2. 0. 0. 0. 1. 0. 1. 3. 1. 1. 2. 3. 0. 1. 0. 3. 1. 2. 0. 2. 0. 1. 0. 3. 0. 1. 0.] The mean and variance are as follows: $\mathsf{E}[X] = 0.96$ and $\mathsf{V}[X] = 0.8384$.

For your convenience, here's the R code

```r
# set seed such that results can be recreated
set.seed(42)

# number simulations and people
numSamples <- 50
N <- 5

# initialize empty result vector
res <- c()

# for loop to simulate repeatedly
for (i in 1:numSamples) {

  # shuffle the N hats
  x <- sample(1:N)

  # number of people picking own hat (element by element the vectors x and
  # 1:N are compared, which yields a vector of TRUE and FALSE, TRUE = 1 and
  # FALSE = 0)
  correctPicks <- sum(x == 1:N)

  # append the result vector by the result of the current simulation
  res <- append(res, correctPicks)
}

# printing of observed mean and variance
```

```python
print(mean(res))
print(var(res))
```

s.3.4.

$$\int_{-\infty}^{\infty} I_{0\le x\le 3}\,\mathrm{d}x = \int_0^3 \mathrm{d}x = 3.$$

s.3.5.

$$\int x\, I_{0\le x\le 4}\,\mathrm{d}x = \int_0^4 x\,\mathrm{d}x = \dots$$

s.3.6.

$$\begin{aligned}\iint xy\, I_{0\le x\le 3} I_{0\le y\le 4}\,\mathrm{d}x\,\mathrm{d}y &= \int_0^3 x \int_0^4 y\,\mathrm{d}y\,\mathrm{d}x \\ &= \int_0^3 x \frac{y^2}{2}\Big|_0^4 \mathrm{d}x \\ &= \int_0^3 x\,\mathrm{d}x\, 8 = \dots\end{aligned}$$

s.3.7. Two solutions. First we integrate over y.

$$\begin{aligned}\iint I_{0\le x\le 3} I_{0\le y\le 4} I_{x\le y}\,\mathrm{d}x\,\mathrm{d}y &= \int I_{0\le x\le 3} \int I_{0\le y\le 4} I_{x\le y}\,\mathrm{d}y\,\mathrm{d}x && (4.26)\\ &= \int I_{0\le x\le 3} \int I_{\max\{x,0\}\le y\le 4}\,\mathrm{d}y\,\mathrm{d}x && (4.27)\\ &= \int_0^3 \int_{\max\{x,0\}}^4 \mathrm{d}y\,\mathrm{d}x && (4.28)\\ &= \int_0^3 y\Big|_{\max\{x,0\}}^4 \mathrm{d}x && (4.29)\\ &= \int_0^3 (4-\max\{x,0\})\,\mathrm{d}x && (4.30)\\ &= 12 - \int_0^3 \max\{x,0\}\,\mathrm{d}x && (4.31)\\ &= 12 - \int_0^3 x\,\mathrm{d}x && (4.32)\\ &= 12 - 9/2. && (4.33)\end{aligned}$$

Let's now instead first integrate over x.

$$\begin{aligned}
\iint I_{0\le x\le 3} I_{0\le y\le 4} I_{x\le y}\,\mathrm{d}x\,\mathrm{d}y &= \int I_{0\le y\le 4} \int I_{0\le x\le 3} I_{x\le y}\,\mathrm{d}x\,\mathrm{d}y && (4.34)\\
&= \int_0^4 \int I_{0\le x\le \min\{3,y\}}\,\mathrm{d}x\,\mathrm{d}y && (4.35)\\
&= \int_0^4 \int_0^{\min\{3,y\}} \mathrm{d}x\,\mathrm{d}y && (4.36)\\
&= \int_0^4 \min\{3,y\}\,\mathrm{d}y && (4.37)\\
&= \int_0^3 \min\{3,y\}\,\mathrm{d}y + \int_3^4 \min\{3,y\}\,\mathrm{d}y && (4.38)\\
&= \int_0^3 y\,\mathrm{d}y + \int_3^4 3\,\mathrm{d}y && (4.39)\\
&= 9/2 + 3. && (4.40)
\end{aligned}$$