

Probability distributions: Questions on the
discussion board

EBPo38Ao5

TAs, Nicky, Ruben

February 8, 2021

CONTENTS

1	Chapter 7	3
2	Hints	4
3	Solutions	4

1 CHAPTER 7

Ex 1.1. Can you elaborate on the solutions of Ex 1.4 and 1.5 of Lecture 1? Specifically, suppose $p_{X,Y}(i,j) = c I_{i=j} I_{1 \leq i \leq 4}$.

What is c ? What is $F_X(i)$? What is $F_Y(j)$?

Ex 1.2. How can the joint PDF of three random variables X, Y, Z be expressed in terms of conditional distributions? Specifically, assume that Y and Z are conditionally independent given X . How can we then simplify the result and calculate the joint PDF of Y, Z ?

2 HINTS

3 SOLUTIONS

s.1.1. Since the joint PMF should sum to 1 when the sum is taken over all possible values of X and Y , we find the following equation:

$$c \sum_i \sum_j I_{i=j} I_{1 \leq i \leq 4} = \sum_i \sum_j c I_{i=j} I_{1 \leq i \leq 4} = \sum_i \sum_j p_{X,Y}(i,j) = 1. \quad (3.1)$$

The question is now what the possible values of X and Y are.

The indicator $I_{1 \leq i \leq 4}$ is equal to 0 if i does not satisfy the inequality $1 \leq i \leq 4$, so $p_{X,Y}(i,j) = 0$ if i does not satisfy the inequality $1 \leq i \leq 4$. Hence, the only possible values for i are 1, 2, 3 and 4.

The indicator $I_{i=j}$ is equal to 0 if i and j are not equal. Since $1 \leq i \leq 4$, we must also have $1 \leq j \leq 4$. So we can write (3.1) as

$$c \sum_{i=1}^4 \left[\sum_{j=1}^4 I_{i=j} I_{1 \leq i \leq 4} \right] = 1. \quad (3.2)$$

Let us now look how this inner summation looks for a specific value of i , say $i = 1$. So we take this summation and replace all occurrences of i by 1:

$$\begin{aligned} \sum_{j=1}^4 I_{1=j} I_{1 \leq 1 \leq 4} &= \sum_{j=1}^4 I_{1=j} \\ &= I_{1=1} + I_{1=2} + I_{1=3} + I_{1=4} \\ &= 1 + 0 + 0 + 0 = 1. \end{aligned}$$

Here I first use that $I_{1 \leq 1 \leq 4} = 1$ because $1 \leq 1 \leq 4$ is true, and then I write out the sum by considering the four possible values for j .

We can do this in the same way for $i = 2$. So we take this summation and replace all occurrences of i by 1:

$$\begin{aligned} \sum_{j=1}^4 I_{2=j} I_{1 \leq 2 \leq 4} &= \sum_{j=1}^4 I_{2=j} \\ &= I_{2=1} + I_{2=2} + I_{2=3} + I_{2=4} \\ &= 0 + 1 + 0 + 0 = 1. \end{aligned}$$

In the same way, the sum will also equal 1 for $i = 3$ and $i = 4$. The reason is that, for a given i , there is exactly one value of j for which the indicator $I_{i=j}$ equals 1.

This means that (3.2) reduces to

$$c \sum_{i=1}^4 1 = 1. \quad (3.3)$$

Hence, $4c = 1$, so $c = \frac{1}{4}$. This answers Exercise 1.4.

For Exercise 1.5, we have to compute $F_X(i)$, i.e. the marginal CDF of X . The following solution was given in the solution file of the lectures:

$$F_X(k) = F_{X,Y}(k, \infty) = \sum_{i \leq k} \sum_j p_{X,Y}(i, j) \quad (3.4)$$

$$= \frac{1}{4} \sum_{i \leq k} \sum_j I_{i=j} I_{1 \leq i \leq 4} \quad (3.5)$$

$$= \frac{1}{4} \sum_{i \leq k} I_{1 \leq i \leq 4} \quad (3.6)$$

$$= k/4. \quad (3.7)$$

I will now explain each of these steps in more detail. The first equality is a way of writing that we sum over all possible values of the second index j . We could also write it as $P(X \leq k) = P(X \leq k, Y \leq \infty)$, which is true since $Y \leq \infty$ always holds.

For the second equality, note that $F_{X,Y}(k, \infty) = P(X \leq k, Y \leq \infty)$, so we have to sum over all possible values of X that are at most k and all possible values of Y .

In the third equality, we simply fill in the given joint PMF.

In the fourth equality, we again use that for a given i , there is exactly one value of j for which the indicator $I_{i=j}$ equals 1. Hence,

$$\sum_j I_{i=j} I_{1 \leq i \leq 4} = I_{1 \leq i \leq 4} \sum_j I_{i=j} = I_{1 \leq i \leq 4}.$$

Note that we can take $I_{1 \leq i \leq 4}$ out of this summation because it does not depend on j . If this is still confusing, take a specific value of i and work it out, like I did above.

In the fifth and final step, I count the number of $i \leq k$ that satisfy $1 \leq i \leq 4$. Assuming that $k \leq 4$, we see that $i \leq k$ already implies that $i \leq 4$, so we count the number of integers i that satisfy $1 \leq i \leq k$. There are exactly k such integers, so the sum equals k and hence the final answer is $k/4$ (for $1 \leq k \leq 4$).

This answer was quite long, but I hope it helps. Working with indicators can be tricky at first, but it is a very useful tool once you get used to them.

s.1.2. In general we can write

$$f_{X,Y,Z}(x, y, z) = f_{Y,Z|X}(y, z|x) f_X(x) \quad (3.8)$$

$$= f_{Y|Z,X}(y|z, x) f_{Z|X}(z|x) f_X(x). \quad (3.9)$$

The first equality is similar to the case with two random variables. In the second equality, we use that $f_{Y,Z|X}(y, z|x)$ lives in a world where we know $X = x$; therefore the conditioning on X should remain in both densities in the next line. Essentially, we use that conditional

densities *are* densities, so this is just the rule $f_{Y|Z}(y|z)f_Z(z)$, but then with conditioning on X added.

Note that here I chose to condition on X first and then on Z , but this is not the only possibility. In general, it depends on the problem what will be the most convenient variable to condition on. Here is an example with first conditioning on Z and then on Y :

$$f_{X,Y,Z}(x,y,z) = f_{X|Y,Z}(x|y,z)f_{Y|Z}(y|z)f_Z(z).$$

If Y and Z are conditionally independent given X , then knowing Z does not give any information about Y if we already know X . In formulas, this means that $f_{Y|Z,X}(y|z,x) = f_{Y|X}(y|x)$. Hence, (3.9) reduces to

$$f_{X,Y,Z}(x,y,z) = f_{Y|X}(y|x)f_{Z|X}(z|x)f_X(x). \quad (3.10)$$

We can then fill in the given PDF of X and the PDFs of Y given X and Z given X . To determine the joint PDF of Y, Z only, we have to marginalize out X and compute the integral

$$f_{Y,Z}(y,z) = \int_{-\infty}^{\infty} f_{Y|X}(y|x)f_{Z|X}(z|x)f_X(x) dx. \quad (3.11)$$