Probability distributions EBP038A05
Lecture slides

Nicky van Foreest, Ruben van Beesten, Arpan Rijal

February 9, 2022

1 WEEK 1

1.1 LECTURE 1

Ex 1.1.1. Consider 12 football players on a football field. Eleven of them are players of F.C. Barcelona, the other one is an arbiter. We select a random player, uniform. This player must take a penalty. The probability that a player of Barcelona 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 1.1.2. Why is it not interesting to consider the case $\{X = 0, Y = 0\}$?

Ex 1.1.3. 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 1.1.4. What is c?

Ex 1.1.5. What is $F_X(i)$? What is $F_Y(j)$?

Ex 1.1.6. Are X and Y dependent? If so, why, because $1 = F_{X,Y}(4, 4) = F_X(4)F_Y(4)$?

Ex 1.1.7. What is $\mathsf{P}\{Z = k\}$?

Ex 1.1.8. 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 1.1.9. What is $\mathsf{P}\{Z = 4\}$?

Remark 1.1.10. 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 $X \sim \text{Pois}(\lambda)$ and $Y \sim \text{Pois}(\mu)$. (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.

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

4. Why do people often/sometimes (?) model the claim sizes as iid $\sim \text{Norm}(\mu, \sigma^2)$? 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., $\mathsf{P}\{X+Y>5-2\}$?

Ex 1.1.11. We have a continuous r.v. $X \geq 0$ with finite expectation. Use 2D integration and indicators to prove that

$$\mathsf{E}[X] = \int_0^\infty x f(x)\,\mathrm{d}x = \int_0 G(x)\,\mathrm{d}x, \tag{1.1.1}$$

where $G(x)$ is the survival function.

Ex 1.1.12. A variation on BH.7.1. Alice is prepared to wait 20 minutes for Bob, while Bob doesn't want to wait longer than 10 minutes. What is the probability that they meet?

Use the fundamental bridge and indicator functions to write this probability as a 2D integral. Then use repeated integration to solve the 2D integral.

1.2 LECTURE 2

Ex 1.2.1. Let $L=\min\{X,Y\}$, where $X,Y \sim \text{Geo}(p)$ and independent. What is the domain of L? Then, use the fundamental bridge and 2D LOTUS to show that

$$\mathsf{P}\{L \geq i\} = q^{2i} \implies L \sim \text{Geo}(1-q^2).$$

Ex 1.2.2. Let $M=\max\{X,Y\}$, where $X,Y \sim \text{Geo}(p)$ and independent. Show that

$$\mathsf{P}\{M=k\} = 2pq^k(1-q^k) + p^2q^{2k}.$$

Ex 1.2.3. Explain that

$$\mathsf{P}\{L=i, M=k\} = 2p^2q^{i+k}\,I_{k>i} + p^2q^{2i}\,I_{i=k}).$$

Ex 1.2.4. With the previous exercise, use marginalization to compute the marginal PMF $\mathsf{P}\{M=k\}$.x

Ex 1.2.5. Now take X,Y iid and $\sim \text{Exp}(\lambda)$. Use the fundamental bridge to show that for $u \leq v$, the joint CDF has the form

$$F_{L,M}(u,v) = \mathsf{P}\{L\leq u, M\leq v\} = 2\int_0^u (F_Y(v)-F_Y(x))f_X(x)\,\mathrm{d}x.$$

Ex 1.2.6. Take partial derivatives to show that for the joint PDF,

$$f_{L,M}(u,v) = 2f_X(u)f_Y(v)\,I_{u\leq v}.$$

2

WEEK 2

WEEK 3

3

4

WEEK 4

5

WEEK 5

WEEK 6

6

WEEK 7

7

7.1 HINTS

h.1.1.11. Check the proof of BH.4.4.8

h.1.2.1. The fundamental bridge and 2D LOTUS have the general form

$$\mathsf{P}\left\{g(X,Y\right\} \in A\} = \mathsf{E}\left[I_{g(X,Y)\in A}\right] = \sum_i \sum_j I_{g(i,j)\in A}\,\mathsf{P}\left\{X = i, Y = j\right\}.$$

Take $g(i,j) = \min\{i,j\}$.

h.1.2.2. Use 2D LOTUS on $g(x,y) = I_{\max\{x,y\}=k}$.

7.2 SOLUTIONS

s.1.1.1. Here is the joint PMF:

$$\mathsf{P}\{P = A, S = 1\} = \frac{1}{12}0.5 \qquad \mathsf{P}\{P = A, S = 0\} = \frac{1}{12}0.5 \tag{7.2.1}$$

$$\mathsf{P}\{P = B, S = 1\} = \frac{11}{12}0.7 \qquad \mathsf{P}\{P = B, S = 0\} = \frac{11}{12}0.3. \tag{7.2.2}$$

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 $\mathsf{P}\left\{X = x, Y = y\right\} = \mathsf{P}\{X = x\}\,\mathsf{P}\left\{Y = y\right\}$. 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.1.1.2. When the claim sizes are 0, then the insurance company does not receive a claim.

s.1.1.3. By the fundamental bridge,

$$\mathsf{P}\{Z = k\} = \sum_{i,j} I_{i+j=k}\, p_{X,Y}(i,j) \tag{7.2.3}$$

$$= \sum_{i,j} I_{i,j\geq 0}\, I_{j=k-i}\, p_{X,Y}(i,j) \tag{7.2.4}$$

$$= \sum_{i=0}^{k} p_{X,Y}(i, k-i). \tag{7.2.5}$$

s.1.1.4. $c = 1/4$ because there are just four possible values for i and j.

s.1.1.5. Use marginalization:

$$F_X(k) = F_{X,Y}(k,\infty) = \sum_{i\leq k}\sum_j p_{X,Y}(i,j) \tag{7.2.6}$$
$$= \frac{1}{4}\sum_{i\leq k}\sum_j I_{i=j}\, I_{1\leq i\leq 4} \tag{7.2.7}$$
$$= \frac{1}{4}\sum_{i\leq k} I_{1\leq i\leq 4} \tag{7.2.8}$$
$$= k/4, \tag{7.2.9}$$
$$F_Y(j) = j/4. \tag{7.2.10}$$

s.1.1.6. 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 $F_{X,Y}(1,1) \neq F_X(1)F_Y(1)$:

$$\frac{1}{4} = F_{X,Y}(1,1) \neq F_X(1)F_Y(1) = \frac{1}{4}\frac{1}{4}. \tag{7.2.11}$$

s.1.1.7. $\mathsf{P}\{Z=2\} = \mathsf{P}\{X=1, Y=1\} = 1/4 = \mathsf{P}\{Z=4\}$, etc. $\mathsf{P}\{Z=k\} = 0$ for $k \not\in \{2,4,6,8\}$.

s.1.1.8. Here is one approach

$$\mathsf{V}[Z] = \mathsf{E}\left[Z^2\right] - (\mathsf{E}[Z])^2 \tag{7.2.12}$$
$$\mathsf{E}\left[Z^2\right] = \mathsf{E}\left[(X+Y)^2\right] = \mathsf{E}\left[X^2\right] + 2\,\mathsf{E}[XY] + \mathsf{E}\left[Y^2\right] \tag{7.2.13}$$
$$(EZ)^2 = (\mathsf{E}[X] + \mathsf{E}[Y])^2 \tag{7.2.14}$$
$$= (\mathsf{E}[X])^2 + 2\,\mathsf{E}[X]\,\mathsf{E}[Y] + (\mathsf{E}[Y])^2 \tag{7.2.15}$$
$$\Longrightarrow \tag{7.2.16}$$
$$\mathsf{V}[Z] = \mathsf{E}\left[Z^2\right] - (\mathsf{E}[Z])^2 \tag{7.2.17}$$
$$= \mathsf{V}[X] + \mathsf{V}[Y] + 2(\mathsf{E}[XY] - (\mathsf{E}[X]\,\mathsf{E}[Y])) \tag{7.2.18}$$
$$\mathsf{E}[XY] = \sum_{ij} ij\,p_{X,Y}(i,j) = \frac{1}{4}(1+4+9+16) = \ldots \tag{7.2.19}$$
$$\mathsf{E}\left[X^2\right] = \ldots \tag{7.2.20}$$

The numbers are for you to compute.

s.1.1.9.

$$\mathsf{P}\{Z=4\} = \sum_{i,j} I_{i+j=4}\, p_{X,Y}(i,j) \tag{7.2.21}$$
$$= \sum_{i=1}^{4}\sum_{j=1}^{4} I_{j=4-i}\frac{1}{16} \tag{7.2.22}$$
$$= \sum_{i=1}^{3}\frac{1}{16} \tag{7.2.23}$$
$$= \frac{3}{16}. \tag{7.2.24}$$

s.1.1.11. The trick is to realize that $x = \int_0^\infty I_{y\leq x}\,\mathrm{d}y$. Using this,

$$\begin{aligned}
\mathsf{E}[X] &= \int_0^\infty x f(x)\,\mathrm{d}x && (7.2.25)\\
&= \int_0^\infty \int_0^\infty I_{y\leq x} f(x)\,\mathrm{d}y\,\mathrm{d}x && (7.2.26)\\
&= \int_0^\infty \int_0^\infty I_{y\leq x} f(x)\,\mathrm{d}x\,\mathrm{d}y && (7.2.27)\\
&= \int_0^\infty \int_0^\infty I_{x\geq y} f(x)\,\mathrm{d}x\,\mathrm{d}y && (7.2.28)\\
&= \int_0^\infty \int_y^\infty f(x)\,\mathrm{d}x\,\mathrm{d}y && (7.2.29)\\
&= \int_0^\infty G(y)\,\mathrm{d}y. && (7.2.30)
\end{aligned}$$

s.1.1.12. Let A, B be the arrival times of Alice and Bob. They meet if $I_{A<B+1/3}\, I_{B<A+1/6}$ is true, i.e., is equal to 1. Therefore, by letting M be the event that they meet:

$$\mathsf{P}\{M\} = \mathsf{E}[I_{A<B+1/3}\, I_{B<A+1/6}] = \int_0^1\int_0^1 I_{x<y+1/3}\, I_{y<x+1/6}\,\mathrm{d}y\,\mathrm{d}x.$$

We can solve this integral by first integrating along y, and then along x. Let's focus on the integral over y first.

$$\begin{aligned}
\int_0^1 I_{x<y+1/3}\, I_{y<x+1/6}\,\mathrm{d}y &= \int_0^1 I_{x-1/3<y<x+1/6}\,\mathrm{d}y\\
&= \int_0^1 I_{\max\{0,x-1/3\}<y\min\{1,x+1/6\}}\,\mathrm{d}y\\
&= \min\{1, x+1/6\} - \max\{0, x-1/3\}
\end{aligned}$$

Now the integral over x:

$$\begin{aligned}
\int_0^1 (\min\{1,x+1/6\} - \max\{0,x-1/3\})\,\mathrm{d}x &= \int_0^1 \min\{1,x+1/6\}\,\mathrm{d}x - \int_0^1 \max\{0,x-1/3\}\,\mathrm{d}x\\
&= \int_0^{5/6} (x+1/6)\,\mathrm{d}x + \int_{5/6}^1 1\,\mathrm{d}x - \int_{1/3}^1 (x-1/3)\,\mathrm{d}x\\
&= 0.5x^2\Big|_0^{5/6} + 1/6\cdot 5/6 - 0.5x^2\Big|_{1/3}^1 + 1/3\cdot 2/3
\end{aligned}$$

Of course, we can find the probability with some simple geometric arguments (compute the area of two triangles). However, this does not work any longer if the density is not uniform. Then we have to do the integration, and that is the reason why I show above how to handle the general case.

s.1.2.1. With the hint,

$$\begin{aligned}
\mathsf{P}\{L\geq k\} &= \sum_i\sum_j I_{\min\{i,j\}\geq k}\,\mathsf{P}\{X=i, Y=j\}\\
&= \sum_{i\geq k}\sum_{j\geq k} \mathsf{P}\{X=i\}\,\mathsf{P}\{Y=j\}\\
&= \mathsf{P}\{X\geq k\}\,\mathsf{P}\{Y\geq k\} = q^k q^k = q^{2k}.
\end{aligned}$$

$\mathsf{P}\{L>i\}$ has the same form as $\mathsf{P}\{X>i\}$, but now with q^{2i} rather than q^i.

s.1.2.2.

$$\begin{aligned}
\mathsf{P}\{M=k\} &= \mathsf{P}\{\max\{X,Y\}=k\} \\
&= p^2 \sum_{ij} I_{\max\{i,j\}=k} q^i q^j \\
&= 2p^2 \sum_{ij} I_{i=k} I_{j<k} q^i q^j + p^2 \sum_{ij} I_{i=j=k} q^i q^j \\
&= 2p^2 q^k \sum_{j<k} q^j + p^2 q^{2k} \\
&= 2p^2 q^k \frac{1-q^k}{1-q} + p^2 q^{2k}
\end{aligned}$$

s.1.2.3.

$$\begin{aligned}
\mathsf{P}\{L=i, M=k\} &= 2\,\mathsf{P}\{X=i, Y=k\}\, I_{k>i} + \mathsf{P}\{X=Y=i\}\, I_{i=k} \\
&= 2p^2 q^{i+k} I_{k>i} + p^2 q^{2i} I_{i=k}.
\end{aligned}$$

s.1.2.4.

$$\begin{aligned}
\mathsf{P}\{M=k\} &= \sum_i \mathsf{P}\{L=i, M=k\} \\
&= \sum_i (2p^2 q^{i+k} I_{k>i} + p^2 q^{2i} I_{i=k}) \\
&= 2p^2 q^k \sum_{i=0}^{k-1} q^i + p^2 q^{2k} \\
&= 2pq^k(1-q^k) + p^2 q^{2k} \\
&= 2pq^k + (p^2 - 2p) q^{2k},
\end{aligned}$$

s.1.2.5. First the joint distribution. With $u \le v$,

$$\begin{aligned}
F_{L,M}(u,v) &= \mathsf{P}\{L \le u, M \le v\} \\
&= 2 \iint I_{x\le u, y\le v, x\le y} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= 2 \int_0^u \int_x^v f_Y(y)\,\mathrm{d}y\, f_X(x)\,\mathrm{d}x && \text{independence} \\
&= 2 \int_0^u (F_Y(v) - F_Y(x)) f_X(x)\,\mathrm{d}x.
\end{aligned}$$

s.1.2.6. Taking partial derivatives,

$$\begin{aligned}
f_{L,M}(u,v) &= \partial_v \partial_u F_{L,M}(u,v) \\
&= 2\partial_v \partial_u \int_0^u (F_Y(v) - F_Y(x)) f_X(x)\,\mathrm{d}x \\
&= 2\partial_v \left\{ (F_Y(v) - F_Y(u)) f_X(u) \right\} \\
&= 2 f_X(u) \partial_v F_Y(v) \\
&= 2 f_X(u) f_Y(v).
\end{aligned}$$