Probability Theory for EOR 2021/2022

—

Solutions to Exercises

—

Week 3

December 15, 2021

PMFs and CDFs

1. Let X be the number of people needed to obtain a birthday match. Of course $P(X = 1) = 0$. The event $X = 2$ occurs when the second person has the same birthday as the first, so the second person has 365 birthday options, one of which is successful. Therefore $P(X = 2) = \frac{1}{365}$.

   The event $X = 3$ occurs when there are two persons in the room with a different birthday (D_2), and the third person entering the room has the same birthday as one of the persons already in the room (S_3). For the first two persons to have a different birthday $X = 2$ should not happen, so $P(D_2) = 1 - P(X = 2)$. The third person has 365 birthday options, two of which are successful. Therefore,

$$\begin{aligned} P(X = 3) &= P(D_2 \cap S_3) \\ &= P(S_3|D_2)P(D_2) \\ &= \frac{2}{365}(1 - P(X = 2)) \\ &= \frac{2}{365}\left(1 - \frac{1}{365}\right). \end{aligned}$$

   The event $X = 4$ occurs when there are three persons in the room with a different birthday (D_3), and the fourth person entering the room has the same birthday as one of the persons already in the room (S_4). For the first three persons to have a different birthday $X = 2$ *and* $X = 3$ should not happen. Notice that $X = 2$ and $X = 3$ are disjoint events. So, $P(D_3) = P((X = 2 \cup X = 3)^C) = 1 - P(X = 2) - P(X = 3)$. The fourth person has 365 birthday options, three of which are successful. Therefore,

$$\begin{aligned}
P(X=4) &= P(D_3 \cap S_4) \\
&= P(S_4|D_3)P(D_3) \\
&= \frac{3}{365}(1 - P(X=2) - P(X=3)) \\
&= \frac{3}{365}\left(1 - \frac{2}{365}\right)\left(1 - \frac{1}{365}\right).
\end{aligned}$$

We can continue to find that

$$p_X(k) = P(X=k) = \frac{k-1}{365}\prod_{m=1}^{k-2}\left(1 - \frac{m}{365}\right), \qquad \text{for } k = 1, \ldots, 366,$$

and $p_X(k) = 0$ otherwise. Notice that we assume that $\prod_{m=1}^{j}(1 - m/365) = 1$ for $j = -1, 0$.

Solutions are not unique, alternatively, it is also to consider first choose $k-1$ different birthdays, and then take into the orderings into account $(k-1)!$ and only for the kth arrival to have a match there are $(k-1)$ choices (same result):

$$P(X=k) = \frac{\binom{365}{k-1}(k-1)!(k-1)}{365^k}.$$

2a. Denote by X the random variable that is equal to i if the first success takes place on the ith trial. We need to find the PMF of X. Furthermore, denote by F_{k-1} the number of failures up to the kth trial and by S_k the event that there is a success on the kth trial. Notice that the trials are independent, so F_{k-1} is independent of S_k. Since the probability of success is $1/2$, a success on the first trial happens with probability

$$P(X=1) = \frac{1}{2}.$$

For the first success to happen on the second trial, it cannot happen on the first trial, so

$$P(X=2) = P(F_1 \cap S_2) = P(F_1)P(S_2) = \frac{1}{2^2},$$

where the second equality uses independence between F_1 and S_2. Continuing in this fashion, we find

$$p_X(k) = P(X=k) = \frac{1}{2^k} \quad k = 1, \ldots, \infty,$$

and $P(X=k) = 0$ otherwise. We can quickly check whether this is a valid PMF. It is clear that $p_X(k) \geq 0$ for all k. Also,

$$\begin{aligned}
\sum_{k=-\infty}^{\infty} p_X(k) &= \sum_{k=1}^{\infty} \frac{1}{2^k} \\
&= -1 + \sum_{k=0}^{\infty} \frac{1}{2^k} \\
&= -1 + \frac{1}{1-\frac{1}{2}} \\
&= -1 + 2 = 1.
\end{aligned}$$

We conclude that the provided PMF is a valid PMF.

2b. Denote by Y the random variable that is equal to i if the missing event (success of failure) occurs on the ith trial. Notice that after the first trial, we are in the situation of a. Suppose the outcome of the first trial is success (S_1). Then $P(Y = 2|S_1) = P(X = 1), P(Y = 3|S_1) = P(X = 2)$, etcetera. Using the law of total probability,

$$\begin{aligned}
P(Y = 2) &= P(Y = 2|S_1)P(S_1) + P(Y = 2|S_1^C)P(S_1^C) \\
&= \frac{1}{2}(P(X = 1) + P(X = 1)) = P(X = 1) = \frac{1}{2}.
\end{aligned}$$

This reasoning actually holds for all subsequent trials as well, so we find that

$$\begin{aligned}
p_Y(k) = P(Y = k) &= P(Y = k|S_1)P(S_1) + P(Y = k|S_1^C)P(S_1^C) \\
&= \frac{1}{2}(P(X = k-1) + P(X = k-1)) \\
&= P(X = k-1) \\
&= \frac{1}{2^{k-1}} \quad k = 2, 3, \ldots
\end{aligned}$$

and $p_Y(k) = 0$ otherwise.

3. The CDF of X is defined as

$$F_X(x) = P(X \leq x)$$

Similarly, for Y, we have

$$F_Y(y) = P(Y \leq y)$$

Now $Y = \mu + \sigma X$, then

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(\mu + \sigma X \leq y) \\ &= P\left(X \leq \frac{y-\mu}{\sigma}\right) \\ &= F_X\left(\frac{y-\mu}{\sigma}\right). \end{aligned}$$

8. Let X be equal to k if k is your most valuable prize ($k = 5, \ldots, 100$). Suppose your most valuable price is \$27. Then you draw 4 out of the first 26 boxes and the 27th box. Using the naive definition of probability

$$P(X = 27) = \frac{\binom{26}{4}}{\binom{100}{5}}.$$

In general,

$$p_X(k) = P(X = k) = \frac{\binom{k-1}{4}}{\binom{100}{5}} \quad \text{for } k = 5, \ldots, 100,$$

and $p_X(k) = 0$ otherwise.

9a. The properties of a valid CDF are

1. Increasing: If $x_1 \leq x_2$, then $F(x_1) \leq F(x_2)$.
2. Right-continuous: $F(a) = \lim_{x \to a^+} F(x)$.
3. Convergence to 0 and 1 in the limits: $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 0$.

To show (i), notice that $F(x_1) = pF_1(x_1) + (1-p)F_2(x_1) \leq pF_1(x_2) + (1-p)F_2(x_2) = F(x_2)$, where the inequality holds because F_1 and F_2 are CDFs *and* $0 < p < 1$. To show (ii)

$$\begin{aligned} \lim_{x \to a^+} F(x) &= \lim_{x \to a^+} (pF_1(x) + (1-p)F_2(x)) \\ &= p \cdot \lim_{x \to a^+} F_1(x) + (1-p) \cdot \lim_{x \to a^+} F_2(x) \qquad \text{(Algebraic Limit Theorem)} \\ &= pF_1(a) + (1-p)F_2(a) \\ &= F(a). \end{aligned}$$

To show (iii)

$$\begin{aligned}
\lim_{x\to-\infty} F(x) &= \lim_{x\to-\infty} pF_1(x) + (1-p)F_2(x) \\
&= p \cdot \lim_{x\to-\infty} F_1(x) + (1-p) \cdot \lim_{x\to-\infty} F_2(x) \\
&= 0, \\
\lim_{x\to\infty} F(x) &= \lim_{x\to\infty} pF_1(x) + (1-p)F_2(x) \\
&= p \cdot \lim_{x\to-\infty} F_1(x) + (1-p) \cdot \lim_{x\to-\infty} F_2(x) \\
&= p + (1-p) \\
&= 1.
\end{aligned}$$

9b. This uses the definition of the CDF and the LOTP. Define X_1 as the the random variable with distribution F_1 and X_2 as the random variable with distribution F_2 and X as the random variable with distribution F. Define Y as the random variable equal to 1 if the coin lands heads and equal to 0 if the coin lands tails. Then,

$$\begin{aligned}
F(x) = P(X \le x) &= P(X \le x|Y=1)P(Y=1) + P(X \le x|Y=0)P(Y=0) \quad \text{(LOTP)} \\
&= P(X_1 \le x)P(Y=1) + P(X_2 \le x)P(Y=0) \\
&= p \cdot F_1(x) + (1-p) \cdot F_2(x).
\end{aligned}$$

10a. For a PMF to be valid, we require that $\sum_n p_X(n) = 1$. In this case, the support of the PMF are the positive integers, so we get $\sum_{n=1}^{\infty} p_X(n) = 1$. It is asked whether it is possible that $p_X(n) = \frac{c}{n}$. In this case, $\sum_{n=1}^{\infty} \frac{c}{n}$ diverges. We conclude there does not exist a valid PMF as suggested.

10b. To show that such a PMF exists, we simply assume that the PMF at n is equal to $\frac{c}{n^2}$ for some $c > 0$. Then,

$$\sum_{n=1}^{\infty} \frac{c}{n^2} = c\frac{\pi^2}{6}.$$

So if we choose $c = \frac{6}{\pi^2}$, we have a PMF that is valid.

12a. Denote by X describe the number of eyes of a fair die throw. Define $Y = X$ if $X \le 2$ and $Y = 2$ otherwise.

12b. If the inequality is strict for some x, then we know that

$$\sum_x P(X = x) < \sum_x P(Y = y) \tag{1}$$

However, summing over all outcomes should give 1 on the l.h.s. (left-hand side) and r.h.s. (right-hand side), so we have arrived at a contradiction. We conclude it is not possible to find discrete r.v.'s such that the stated inequality holds.

13. We have

$$\begin{aligned} P(X=a) &= \sum_z P(X=a|Z=z)P(Z=z) && \text{(LOTP)} \\ &= \sum_z P(Y=a|Z=z)P(Z=z) && \text{(Given in the question)} \\ &= P(Y=a). && \text{(Reversed LOTP)} \end{aligned}$$

Named distributions

17. There are enough seats if 100 or less people show up. Denote by N the number of people who show up. We are looking for $P(N \leq 100)$. N follows a binomial distribution (since each person shows up independently of the other), with parameters $n = 110$ and $p = 0.9$. Then,

$$P(N \leq 100) = \sum_{x=0}^{100} \binom{110}{x} \cdot 0.9^x \cdot 0.1^{110-x} = 0.67.$$

Probably a good idea not to calculate this by hand. In R, type `pbinom(100,110,0.9)`.

18a. Suppose that n games are played with $n = 4, 5, 6, 7$. If A wins, it of course wins the final match and has won three of the preceeding $n-1$ matches. Denote by W_i the number of matches won out of i mathces. Since the outcomes of the matches are independent, the probability that $W_{n-1} = 3$ is

$$P(W_{n-1} = 3) = \binom{n-1}{3} p^3 (1-p)^{n-4}.$$

The probability of winning is then

$$\begin{aligned} P(A) &= \sum_{n=4}^{7} P(A|W_{n-1}=3)P(W_{n-1}=3) \\ &= p \sum_{n=4}^{7} P(W_{n-1}=3) \\ &= p^4 \cdot \sum_{n=4}^{7} \binom{n-1}{3} (1-p)^{n-4}. \end{aligned}$$

18b. If the teams keep playing, then you can rephrase the question to: what is the probability that team A wins at least 4 out of seven matches? Notice that they can never first lose 4 (and therefore, lose the series) and then win 4, so we don't include scenarios in which team A actually loses the series. Denote by X the number of matches won by team A, then the probability of winning the series is,

$$P(A) = \sum_{x=4}^{7} P(X = x) = \sum_{x=4}^{7} \binom{7}{x} p^x (1-p)^{7-x}.$$

24a. Denote by X_8 the number of heads in 8 tosses, so $X_8 \sim \text{Bin}(8, 0.5)$, and by HH the event that the first two tosses are heads. Notice that we have $X = 2$ if in the remaining 8 tosses we don't see any heads. Similarly, we have $X = 3$ if in the remaining 8 tosses we have 1 heads. In general,

$$P(X = x|HH) = P(X_8 = x - 2) \text{ for } x = 2, \dots, 10.$$

24b. Denote by A the event that at least two heads are tossed. Then,

$$P(A) = 1 - \binom{10}{0}\frac{1}{2^{10}} - \binom{10}{1}\frac{1}{2^{10}}.$$

For $x = 2, \dots, 10$, using the definition of conditional probability,

$$\begin{aligned} P(X = x|A) &= \frac{P(X = x, A)}{P(A)} = \frac{P(X = x)}{P(A)} = \frac{\binom{10}{x}\frac{1}{2^{10}}}{1 - \binom{10}{0}\frac{1}{2^{10}} - \binom{10}{1}\frac{1}{2^{10}}} \\ &= \binom{10}{x} \cdot \frac{1}{1024 - 1 - 10}. \end{aligned}$$

25a. Assume that X is the number of heads for Alice, then $V = n - X$ is the number of tails and $V = n - X \sim \text{Bin}(n, 1/2)$. Similarly, assume that Y is the number of heads for Bob, then $W = n + 1 - Y$ is the number of tails and $W = n + 1 - Y \sim \text{Bin}(n + 1, 1/2)$. Of course, $P(X < Y) = P(V < W)$ since V has the same distribution as X and W has the same distribution as Y, X and Y are independent and V and W are independent.

25b. Now

$$P(X < Y) = P(n - X < n + 1 - Y) = P(X > Y - 1) = P(X \geq Y).$$

Also, since $X < Y$ and $X \geq Y$ are disjoint and their union has probability equal to one, we have $1 = P(X < Y) + P(X \geq Y) = 2P(X < Y)$. We conclude that $P(X < Y) = 1/2$.

26. If $X \sim \text{HGeom}(w, b, n)$, then

$$P(X = k) = \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}.$$

For $n - X$, we have

$$\begin{aligned} P(n - X = k) &= P(X = n - k) \\ &= \frac{\binom{w}{n-k}\binom{b}{k}}{\binom{w+b}{n}}. \end{aligned}$$

We see that $X \sim \text{Hgeom}(b, w, n)$.

28. The probability of the number of chicks that hatch H is binomial with parameters n and p ($H \sim \text{Bin}(n, p)$). Each chicken has probability pr of hatching a chick that survives, so the number of surviving chicks is also binomial, now with parameters n and pr.

31a. Denote by X the number of correct guesses. It is helpful to think for a moment about a particular ordering, for example:

$$M, M, T, M, T, T$$

Refer to these as the milk and tea locations. To get k of the milk teas correct, the lady needs to select k out of 3 milk locations and $3 - k$ out of 3 tea locations. In total, she can select 3 out of 6 locations for her guess. So, we find that

$$P(X = k) = \frac{\binom{3}{k}\binom{3}{3-k}}{\binom{6}{3}}.$$

Notice that the particular ordering MMTMTT above is not material to the argument.

31b. Denote by L the claim of the lady and by M the event that the cup is milk first. We are looking for the posterior odds, so

$$odds(M|L) = \frac{P(M|L)}{P(M^C|L)}.$$

We first calculate the numerator

$$P(M|L) = \frac{P(L|M)P(M)}{P(L|M)P(M) + P(L|M^C)P(M^C)} = \frac{p_1 \frac{1}{2}}{p_1 \frac{1}{2} + (1 - p_2)\frac{1}{2}}.$$

where we have $P(L|M^C) = 1 - p_2$ because the cup is tea first, which the lady would correctly identify with probability p_2. The probability that she mistakes this for milk first is then $1 - p_2$. Some algebra now shows that

$$odds(M|L) = \frac{p_1}{1 - p_2}.$$

Independence of r.v.s.

39. Let Y be discrete uniform over the support and let $X = Y + 1$ when $Y < 10$ and $X = 1$ if $Y = 10$. Then $P(X = i) = P(Y = i - 1) = \frac{1}{10} = P(Y = i)$. If X and Y are independent, then

$$\begin{aligned} P(X = Y) &= \sum_y P(X = Y | Y = y) P(Y = y) && \text{(LOTP)} \\ &= \sum_y P(X = y) P(Y = y) > 0, \end{aligned}$$

where the final inequality holds since X and Y have the same support.

42. Yes, they have the same distribution (this is the example given in question 39.).

$$\begin{aligned} P(X < Y) &= \sum_{x=1}^{7} P(X < Y | X = x) P(X = x) \\ &= 6 \cdot \frac{1}{7}, \end{aligned}$$

Since for X equal to Monday through Saturday, Y will be larger, but for X equal to Sunday, Y will be smaller than X.