ENGRD 2700: Basic Engineering Probability and Statistics
Fall 2019

Homework 3 Solutions

Due Friday, October 4 by 11:59pm. Submit to Gradescope by clicking the name of the assignment. See `https://people.orie.cornell.edu/yudong.chen/engrd2700_2019fa.html#homework` for detailed submission instructions.

The same stipulations from Homework 1 (e.g., independent work, computer code, etc.) still apply.

1. The lifetime T, in years, of the light bulb you just purchased satisfies

$$P(T > t) = e^{-t/3} \qquad \text{for all } \; t \geq 0.$$

Suppose the bulb has lasted more than x years, where $x \geq 0$. Given this information, what's the conditional probability that it will last at most $x + 2$ years? Does your answer depend on the value of x?

The desired probability is given by

$$P(T \leq x + 2 | T > x) = \frac{P(T \leq x + 2 \text{ and } T > x)}{P(T > x)} = \frac{P(x < T \leq x + 2)}{P(T > x)}$$

Since

$$\begin{aligned} P(x < T \leq x + 2) = P(T \leq x + 2) - P(T \leq x) &= (1 - P(T > x + 2)) - (1 - P(T > x)) \\ &= P(T > x) - P(T > x + 2) \\ &= e^{-x/3} - e^{-(x+2)/3}, \end{aligned}$$

we conclude that

$$P(T \leq x + 2 | T > x) = \frac{e^{-x/3} - e^{-(x+2)/3}}{e^{-x/3}} = 1 - e^{-2/3} \approx 0.49.$$

Note that this is the same as $P(T \leq 2) = 1 - P(T > 2) = 1 - e^{-2/3}$, the probability that the bulb lasts at most 2 years. In this case we say that the distribution of T is *memoryless*; knowing that the bulb has lasted more than x years, the probability that it will last two additional years is the same as if we had a brand-new bulb. Therefore, the answer does not depend on the value of x.

2. Suppose a family with two children is selected at random, and the genders of the children are noted in the order of their birth (f for female, m for male). Assume that the possible outcomes $\{ff, fm, mf, mm\}$ are all equally likely.

   (a) A family tells us that the first child is female. Given this information, what's the probability that both children are female?

   Solution 1: Let F be the event that the first child is female. Then

$$\begin{aligned} P(ff|F) &= \frac{P(ff \cap F)}{P(F)} \\ &= \frac{P(ff)}{P(ff) + P(fm)} \\ &= \frac{1/4}{(1/4) + (1/4)} \\ &= \frac{1}{2} \end{aligned}$$

Solution 2: we can also use the independence of the genders between the first child and the second child to solve this question. Note that since the possible outcomes of ff, fm, mf, mm are equally likely, then each of the event has a probability of $\frac{1}{4}$. That is $P(ff) = P(fm) = P(mf) = P(mm) = \frac{1}{4}$. Then for the event of ff, we have:

$$\begin{aligned}
&P(\text{the first child is female and the second child is female}) \\
&= P(ff) \\
&= \frac{1}{4} \\
&= \frac{1}{2} \cdot \frac{1}{2} \\
&= P(f) \cdot P(f) \\
&= P(\text{first child is female}) \times P(\text{second child is female})
\end{aligned}$$

We can apply the same argument to the rest of the three cases. Then we get P(the first child is female/male and the second child is female/male) = P(the first child is female/male)× P(the second child is female/male) . According to the definition of independence, if $P(A \cap B) = P(A) \times P(B)$, event A is independent of event B . Therefore, by the previous results, the gender of the first child is independent of the gender of second child. Therefore,

$$\begin{aligned}
&P(\text{the first child is female and the second child is female} \,|\, \text{the first child is female}) \\
&= P(\text{the first child is female} \,|\, \text{the first child is female}) \times P(\text{the second child is female} \,|\, \text{the first child is female}) \\
&= 1 \cdot P(\text{the second child is female} \,|\, \text{the first child is female}) \\
&= P(\text{the second child is female}) = \frac{1}{2}
\end{aligned}$$

(b) Another family tells us that at least one of the children is female. Given this information, what's the probability that both children are female?

Let A be the event that there's at least one female child. Then

$$\begin{aligned}
P(ff|A) &= \frac{P(ff \cap A)}{P(A)} \\
&= \frac{P(ff)}{P(ff) + P(fm) + P(mf)} \\
&= \frac{1/4}{(1/4) + (1/4) + (1/4)} \\
&= \frac{1}{3}.
\end{aligned}$$

(c) Assume that, if a child is female, then her name is Katie with some small probability p independently of the gender and naming of the other child, and that no boys are named Katie.

Now, a third family tells us that they have at least one child named Katie. Given this information, what's the probability that both children are female?

Let L be the event that there's at least one female child named Katie. Then

$$P(ff|L) = \frac{P(ff \cap L)}{P(L)}.$$

Note that

$$\begin{aligned} P(ff \cap L) &= P(L|ff)P(ff) \\ &= [p(1-p) + (1-p)p + p^2] \cdot (1/4) = \frac{2p - p^2}{4} \end{aligned}$$

and, according to the law of total probability,

$$\begin{aligned} P(L) &= P(L|ff)P(ff) + P(L|fm)P(fm) + P(L|mf)P(mf) + P(L|mm)P(mm) \\ &= (2p - p^2) \cdot (1/4) + p \cdot (1/4) + p \cdot (1/4) + 0 \cdot (1/4) = \frac{4p - p^2}{4}. \end{aligned}$$

So,

$$P(ff|L) = \frac{2p - p^2}{4p - p^2}.$$

P.S. When p is small, the above probability is approximately 1/2. Comparing this with part (b), note that knowing the name of one of the daughters increases the probability that both are female from 1/3 to 1/2! Can you explain this phenomenon intuitively?

3. A train station has installed a system for determining whether bags contain explosives. It has a 90% chance of correctly identifying a bag containing explosives as dangerous, and a 99% chance of correctly classifying a bag without explosives as safe. Suppose that the train station screens 4 million bags per year, and that 10 of these bags are expected to contain explosives.

   (a) A bag is identified by the system as dangerous. What's the probability that it actually contains explosives?

   Let IE be the event that the bag is identified as dangerous, and let CE be the event that it actually contains explosives. Then

   $$P(CE|IE) = \frac{P(CE \cap IE)}{P(IE)}.$$

   Since

   $$P(CE \cap IE) = P(IE|CE)P(CE) = (0.90) \cdot \left(\frac{10}{4 \times 10^6}\right) = 2.25 \times 10^{-6}$$

   and

   $$\begin{aligned} P(IE) &= P(IE|CE)P(CE) + P(IE|CE^c)P(CE^c) \\ &= (0.90) \cdot \left(\frac{10}{4 \times 10^6}\right) + (1 - 0.99) \cdot \left(1 - \frac{10}{4 \times 10^6}\right) \approx 0.01, \end{aligned}$$

   we conclude

   $$P(CE|IE) \approx \frac{2.25 \times 10^{-6}}{0.01} = 0.000225$$

   (b) If we want the probability in part (a) to be at least 0.5, what should the probability of correctly identifying a bag without explosives be?

   We want

   $$P(CE|IE) = \frac{P(IE|CE)P(CE)}{P(IE|CE)P(CE) + P(IE|CE^c)P(CE^c)} \geq \frac{1}{2}.$$

Rearranging, we get

$$P(IE|CE^c) \leq \frac{P(IE|CE)P(CE)}{P(CE^c)}$$
$$= \frac{2.25 \times 10^{-6}}{\left(1 - \frac{10}{4\times 10^6}\right)} = 2.25 \times 10^{-6}.$$

So, we need

$$P(IE^c|CE^c) > 1 - 2.25 \times 10^{-6} \approx 0.9999978.$$

(c) Would it be possible to make the probability in part (a) at least 0.5 by increasing the chance of correctly identifying bags containing explosives? Justify your answer.

No , because $P(CE|IE) \geq 1/2$ would require

$$P(IE|CE) \geq \frac{P(IE|CE^c)P(CE^c)}{P(CE)}$$
$$= \frac{(1-0.99)(1-\frac{10}{4\times 10^6})}{\frac{10}{4\times 10^6}}$$
$$\approx 4000 > 1.$$

Alternative solution: Let us consider the best scenario, which is that all bags that contain explosives have been correctly identified. Note that even in this situation there would still be $0.01 \times (4000000 - 10) \approx 40000$ false positives . That is we still have 40000 bags which do not contain explosives but have been identified as containing explosives. Then, $P(CE|IE) = \frac{10}{10+40000} < 0.5$. Therefore it is not possible.

4. A casino offers you the following game. There's a pot that initially contains 1 dollar. A fair coin is tossed. If it comes up tails, the amount of money in the pot is doubled, and the coin is tossed again. The game ends once the coin comes up heads, at which point you get whatever is in the pot. Let the random variable X denote the amount of money you win by playing this game.

(a) What is the set $\mathcal{X}$ of possible values for X? The possible values for X are $\mathcal{X} = \{2^i : i = 0, 1, 2, 3, \dots\}$

(b) Write down the PMF $p_X(x)$ of X for $x \in \mathcal{X}$.

If the first toss is heads (which happens with probability 1/2) you win $1 = 2^0$ dollar, and if the first toss is tails and the second one is heads (which happens with probability $(1/2)\cdot(1/2) = 1/2^2$), you win $2 = 2^1$ dollars. In general,

$$p_X(2^i) = \frac{1}{2^{i+1}} \qquad \text{for } i = 0, 1, 2, \dots$$

(c) What's the probability that you'll win more than 40 dollars?

Since your winnings can only be powers of 2,

$$P(\text{win more than 40 dollars}) = 1 - P(\text{win less than 40 dollars})$$
$$= 1 - P(X \leq 2^5)$$
$$= 1 - \sum_{i=0}^{5} p_X(2^i)$$
$$= 1 - \sum_{i=0}^{5} \frac{1}{2^{i+1}}$$
$$= 1 - \frac{1}{2} \cdot \left(\frac{1-(1/2)^6}{1-1/2}\right) = \frac{1}{64} \approx 0.0156.$$

(d) Compute $E[X]$.

Let $\mathcal{X}$ be the set of all earnings from the game for which the PMF $p_X(x)$ is non-zero. Then

$$\begin{aligned}
E[X] &= \sum_{x\in\mathcal{X}} x p_X(x) \\
&= \sum_{i=0}^{\infty} 2^i p_X(2^i) \\
&= \sum_{i=0}^{\infty} 2^i \cdot \frac{1}{2^{i+1}} = \sum_{i=0}^{\infty} \frac{1}{2} = \frac{1}{2} + \frac{1}{2} + \cdots = \infty.
\end{aligned}$$

(e) Define another random variable $Y = \min\{X, 2^{10}\}$. Find the set $\mathcal{Y}$ of possible values for Y, write down the PMF of Y, and compute $E[Y]$.

The possible values for set $\mathcal{Y}$ are $\{2^i : i = 0, 1, 2, \ldots, 10\}$

The PMF of $p_Y(y)$ is

$$p_Y(2^i) = \begin{cases} \frac{1}{2^{i+1}} & \text{for } i = 0, 1, 2, \ldots, 9, \\ P(X \geq 2^{10}) = \frac{1}{2^{10}} & \text{for } i = 10, \end{cases}$$

It follows that

$$\begin{aligned}
E[Y] &= \sum_{y\in\mathcal{Y}} y p_Y(y) \\
&= \sum_{i=0}^{9} 2^i \cdot \frac{1}{2^{i+1}} + 2^{10} P(X \geq 2^{10}) \\
&= \frac{1}{2} \cdot 10 + 2^{10} \cdot \frac{1}{2^{10}} = 5 + 1 = 6.
\end{aligned}$$

5. A random variable X has the following cumulative distribution function (CDF):

$$F(x) = \begin{cases} 0 & x < -2 \\ 0.2 & -2 \leq x < -1 \\ 0.3 & -1 \leq x < 0 \\ 0.7 & 0 \leq x < 1 \\ 0.8 & 1 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$

(a) Find $f(x)$, its probability mass function (PMF).

From the lecture notes we know that the PMF is obtained by analyzing the "jumps" of the CDF. Therefore, we have

$$f(x) = \begin{cases} 0.2 & x = -2 \\ 0.1 & x = -1 \\ 0.4 & x = 0 \\ 0.1 & x = 1 \\ 0.2 & x = 2 \end{cases}$$

(b) Compute $E[X]$ and $\mathrm{Var}(X)$.

The formula for $E[X]$ given the PMF is

$$E[X] = \sum_{x\in\mathcal{X}} x f(x),$$

where $\mathcal{X}$ is the set of points where the PMF is non-zero. Hence,

$$\begin{aligned} E[X] &= (-2)\cdot(0.2)+(-1)\cdot(0.1)+(0)\cdot(0.4)+(1)\cdot(0.1)+(2)\cdot(0.2) \\ &= 0 \end{aligned}$$

Alternatively, one can deduce that $E[X]=0$ by observing that the PMF is symmetric around 0 . To find $\text{Var}(X)$ we use the formula

$$\text{Var}(X) = E[X^2] - (E[X])^2.$$

First, we compute

$$\begin{aligned} E[X^2] &= (-2)^2\cdot(0.2)+(-1)^2\cdot(0.1)+(0)^2\cdot(0.4)+(1)^2\cdot(0.1)+(2)^2\cdot(0.2) \\ &= 1.8 \end{aligned}$$

Hence $\text{Var}(X) = 1.8 - (0)^2 = 1.8$

(c) Compute $E[\sin(X)]$.
From the lecture notes we know how to compute expectation of a function of a random variable,

$$\begin{aligned} E(\sin(X)) &= \sum_{x\in S}\sin(x)f(x) \\ &= \sin(-2)\cdot(0.2)+\sin(-1)\cdot(0.1)+\sin(0)\cdot(0.4)+\sin(1)\cdot(0.1)+\sin(2)\cdot(0.2) \\ &= 0 \end{aligned}$$

Alternatively, observe that $\sin(x)$ is an odd function. Since the PMF is symmetric around the origin, it follows that the expectation is zero because the positive and negative parts of $\sin(x)$ cancel.

6. Stephanie Kerry is a very good basketball player. Each time she attempts a free throw, she misses it with a probability of only 7% (independently of other free throw attempts). This month, she will attempt 1000 free throws.

   (a) What is the distribution of X, the total number of free throws that Stephanie misses this month? Give its name and compute its parameters.
   X follows a Binomial distribution $B(n,p)$, with parameters $n=1000$ and $p=0.07$.

   (b) What is the probability that Stephanie will miss at least 61 free throws this month?
   This happens when the number of students that show up is greater than 61. By means of R or otherwise, we obtain

   $$\begin{aligned} P(X\geq 61) &= 1-P(X\leq 60) \\ &\approx 1-0.118 = 0.882 \end{aligned}$$

   (c) Use a Poisson approximation to give an approximation for the probability that Stephanie will miss at least 61 free throws this month.
   We will use the Poisson distribution with $\lambda = 1000\times 0.07 = 70$ to approximate the distribution of X. By means of R or otherwise, we obtain

   $$\begin{aligned} P(X\geq 61) &= 1-P(X\leq 60) \\ &\approx 1-0.127 = 0.873 \end{aligned}$$

   (d) By the end of the 15th day of the month, Stephanie has already missed 55 free throws (though we don't know how many free throws she has attempted). Given this information, what is the chance that she will miss *at least* 60 free throws in total this month? Use the Poisson approximation in answering this question.

Given that Stephanie has already missed 55 free throws, the question is asking for the probability that there are greater than or equal to 60 misses. Therefore, we have

$$\begin{aligned} P(X \geq 60 | X \geq 55) &= \frac{P(X \geq 60)}{P(X \geq 55)} \\ &= \frac{1 - P(X \leq 59)}{1 - P(X \leq 54)} \\ &\approx \frac{1 - 0.102}{1 - 0.028} \\ &= 0.924\,. \end{aligned}$$

(e) At the end of the month, Stephanie will look at her total number of missed free throws, X. If $X \geq 40$, she will put 5 dollars in a jar ***for each*** free throw she misses. For example, she puts 200 dollars in the jar if $X = 40$. If $X < 40$, she will leave the jar empty. What is the expected number of dollars in the jar? Use the Poisson approximation in answering this question.

$$\begin{aligned} \text{Expected number of dollars} &= \sum_{k=0}^{39} 0 \times \frac{e^{-70}70^k}{k!} + \sum_{k=40}^{\infty} 5k \times \frac{e^{-70}70^k}{k!} \\ &= 5 \times 70 \times \sum_{k=40}^{\infty} \frac{e^{-70}70^{k-1}}{(k-1)!} \\ &= 5 \times 70 \times \sum_{k=39}^{\infty} \frac{e^{-70}70^k}{k!} \\ &= 350 \times P(X \geq 39) \\ &= 350(1 - P(X \leq 38)) \\ &\approx 350 \end{aligned}$$

Another (approximate) approach is to compute a table of costs and Poisson probabilities, and use it to compute the sum as described in the first line above until the terms get negligible.