EK381 Probability, Statistics, and Data Science for Engineers — Boston University
Prof Carruthers — Summer 2025

Homework 4: finish by 6/3.

Reading: Notes: Chapter 2, 3. **Videos:** 2.5, 2.6, 3.1 - 3.4

Problem 4.1 (**Video 2.5, 2.6**, **Quick Calculations**)

Calculate each of the requested quantities.

(a) Let X be Binomial$(5, \frac{1}{2})$. Calculate $\mathbb{P}[1 \leq X \leq 4]$ and $\mathbb{P}\big[X = 2 \,\big|\, 1 \leq X \leq 4\big]$.

Solution:

$$\mathbb{P}[1 \leq X \leq 4] = 1 - \big(P_X(0) + P_X(5)\big) = 1 - \left(\binom{5}{0}\left(\frac{1}{2}\right)^5 + \binom{5}{5}\left(\frac{1}{2}\right)^5\right) = 1 - \frac{2}{2^5} = \frac{15}{16}$$

$$\mathbb{P}\big[X = 2 \,\big|\, 1 \leq X \leq 4\big] = \frac{\mathbb{P}\big[\{X = 2\} \cap \{1 \leq X \leq 4\}\big]}{\mathbb{P}[1 \leq X \leq 4]} = \frac{P_X(2)}{\mathbb{P}[1 \leq X \leq 4]} = \frac{\binom{5}{2}\left(\frac{1}{2}\right)^5}{\frac{15}{16}} = \frac{\frac{10}{32}}{\frac{15}{16}} = \frac{1}{3}$$

(b) Let Y be Geometric$(1/2)$. Calculate $\mathsf{Var}[Y + 4]$ and $\mathbb{E}[3Y + 0.2]$.

Solution:

The variance of a linear function is $\mathsf{Var}[aY + b] = a^2\mathsf{Var}[Y]$.
Therefore, $\mathsf{Var}[Y + 4] = \mathsf{Var}[Y] = \frac{1-p}{p^2} = \frac{1/2}{(1/2)^2} = 2$.
By linearity of expectation, $\mathbb{E}[3Y + 0.2] = 3\mathbb{E}[Y] + 0.2 = 3 \cdot \frac{1}{p} + 0.2 = 3 \cdot 2 + 0.2 = 6.2$.

(c) Let X be Bernoulli(p) with $\mathbb{E}[2^X] = \frac{5}{4}$. Determine the value of p and of $\mathbb{E}[X^2]$.

Solution:

First, note that $\mathbb{E}[2^X] = \sum_{x \in R_X} 2^x\, P_X(x) = 2^0 \cdot P_X(0) + 2^1 \cdot P_X(1) = (1-p) + 2p = 1 + p$. Therefore, $1 + p = \frac{5}{4}$ and $p = \frac{1}{4}$. It follows that $\mathbb{E}[X^2] = \sum_{x \in R_X} x^2\, P_X(x) = 0^2 \cdot P_X(0) + 1^2 \cdot P_X(1) = p = \frac{1}{4}$.

(d) Let X be Discrete Uniform$(1, 12)$. Let B the event that X is strictly greater than 4 *and* strictly less than 8. Calculate $\mathbb{E}[X|B]$ and $\mathbb{E}[X^2|B]$.

Solution:

Note that the new conditional PMF $P_{X|B}(x)$ is Discrete Uniform$(5, 7)$ by restricting the PMF to the values in $B = \{5, 6, 7\}$ and rescaling, dividing by $\mathbb{P}[B] = \frac{3}{12}$. Therefore,

$$\mathbb{E}[X|B] = \frac{5 + 7}{2} = 6; \quad \mathsf{Var}[X|B] = \frac{(7 - 5 + 1)^2 - 1}{12} = \frac{8}{12} = \frac{2}{3}$$

$$\mathbb{E}[X^2|B] = \mathsf{Var}[X|B] + \big(\mathbb{E}[X|B]\big)^2 = \frac{2}{3} + 6^2 = \frac{110}{3}$$

(e) Let X be Poisson(2). Calculate $\mathbb{E}[3X-1]$ and $\mathbb{E}[3-X^2]$.

Solution:

$$\begin{aligned}\mathbb{E}[3X-1] &= 3\mathbb{E}[X] - 1 = 3\cdot 2 - 1 = 5\\ \mathbb{E}[3-X^2] &= 3 - \mathbb{E}[X^2] = 3 - ((\mathbb{E}[X])^2 + \mathsf{Var}(X)) = 3 - (4+2) = -3\ .\end{aligned}$$

Problem 4.2 (**Video 2.5, 2.6**) You are examining blood cells under a microscope. Your microscope slide is divided up into a grid consisting of tiny squares. The number of blood cells you count in a single square, denoted by X, can be modeled as a Poisson(2) random variable.[1]

(a) What is the probability that X is less than 2?

Solution:

$$\mathbb{P}[X<2] = P_X(0) + P_X(1) = \frac{2^0}{0!}e^{-2} + \frac{2^1}{1!}e^{-2} = 3e^{-2}$$

(b) Given that X is less than 4, what is the probability that X is less than 2?

Solution:

$$\begin{aligned}\mathbb{P}[X<2|X<4] &= \frac{\mathbb{P}[\{X<2\}\cap\{X<4\}]}{\mathbb{P}[X<4]} = \frac{\mathbb{P}[X<2]}{\mathbb{P}[X<4]}\\ \mathbb{P}[X<4] &= P_X(0)+P_X(1)+P_X(2)+P_X(3)\\ &= e^{-2}\left(\frac{2^0}{0!}+\frac{2^1}{1!}+\frac{2^2}{2!}+\frac{2^3}{3!}\right) = e^{-2}\left(1+2+2+\frac{4}{3}\right) = \frac{19}{3}e^{-2}\\ \mathbb{P}[X<2|X<4] &= \frac{3e^{-2}}{\frac{19}{3}e^{-2}} = \frac{9}{19}\end{aligned}$$

(c) Given that X is less than 4, what is the conditional expected value of X?

Solution:

Let $B = \{0,1,2,3\}$.

$$P_{X|B}(x) = \begin{cases}\dfrac{P_X(x)}{\mathbb{P}[X\in B]} & x\in B\\ 0 & x\notin B\end{cases} = \begin{cases}\frac{e^{-2}}{\frac{19}{3}e^{-2}} & x=0\\ \frac{2e^{-2}}{\frac{19}{3}e^{-2}} & x=1\\ \frac{2e^{-2}}{\frac{19}{3}e^{-2}} & x=2\\ \frac{\frac{4}{3}e^{-2}}{\frac{19}{3}e^{-2}} & x=2\\ 0 & \text{otherwise}\end{cases} = \begin{cases}\frac{3}{19} & x=0\\ \frac{6}{19} & x=1\\ \frac{6}{19} & x=2\\ \frac{4}{19} & x=3\\ 0 & \text{otherwise}\end{cases}$$

[1]Interestingly, this is the model used by William Gosset in his first publication: *Student. "On the error of counting with a haemacytometer." Biometrika (1907): 351-360.* Gosset's day job was Head Experimental Brewer at Guinness, and he was only allowed to publish under the pseudonym "Student." He is also responsible for Student's t-distribution and Student's t-test, which we will encounter a bit later in the course.

$$\mathbb{E}[X|B] = \sum_{x \in B} x\, P_{X|B}(x) = 0 \cdot \frac{3}{19} + 1 \cdot \frac{6}{19} + 2 \cdot \frac{6}{19} + 3 \cdot \frac{4}{19} = \frac{30}{19}$$

(d) Assume that the number of blood cells is independent across squares in the grid. Let Y be the number of squares you examine until the first square you find with 3 or more blood cells. What kind of random variable is Y? (Don't forget the parameters.)

Solution:

By the complement property, $\mathbb{P}[X \geq 3] = 1 - \mathbb{P}[X < 3] = 1 - 5e^{-2}$.
Y is a Geometric$(1 - 5e^{-2})$ random variable.

Problem 4.3 (**Video 2.5, 2.6**) After Team Avatar (or the BoomerAang Gang) arrives in a small Fire Nation town, Sokka is talking to local residents to determine the number of days until Sozin's Comet arrives, which we denote by the random variable X. Initially, Sokka thinks the comet is equally likely to arrive in the next 2 to 30 days. Sokka is particularly worried about the mounting costs at the local inn, which charges a one-time check-in fee of 3 gold coins (or ban) and a daily rate of 2 gold coins (or ban). Let Y denote the total inn cost for X days. (To be very clear, if the comet arrives in X days, the total inn cost is $Y = 3 + 2 \cdot X$ gold coins.)

(a) What kind of a random variable is X? (Don't forget the parameters.)

Solution:

X is a Discrete Uniform$(2, 30)$ random variable.

(b) What is the expected value for the total cost Sokka will have to pay the inn?

Solution:

Since we can look up $\mathbb{E}[X] = \dfrac{2+30}{2} = 16$, it is convenient to use the linearity of expectation.

$$\mathbb{E}[Y] = \mathbb{E}[2X + 3] = 2\,\mathbb{E}[X] + 3 = 2 \cdot 16 + 3 = 35$$

(c) Sokka only has 49 gold coins. What is the probability that the money will run out before the comet arrives?

Solution:

First, we note that 49 gold coins will buy $\dfrac{49-3}{2} = 23$ days at the inn. Therefore, if the comet arrives in 24 or more days, Sokka will run out of money. The PMF for a Discrete Uniform$(2, 30)$ is

$$P_X(x) = \begin{cases} \dfrac{1}{29} & x = 2, 3, \ldots, 30 \\ 0 & \text{otherwise.} \end{cases}$$

Therefore,

$$\mathbb{P}[\{\text{money runs out}\}] = \mathbb{P}[X \geq 24] = \sum_{x=24}^{30} P_X(x) = 7 \cdot \frac{1}{29} = \frac{7}{29}$$

(d) After talking to local residents, Sokka is now sure that the earliest the comet might arrive is in 5 days *and* he is also sure that the comet will arrive within 15 days or less. Let us call this new information event B. Determine the resulting conditional PMF $P_{X|B}(x)$ for the number days until the comet arrives. What kind of random variable corresponds to this conditional PMF? (Don't forget the parameters.)

Solution:

We can express this information as $B = \{x \in R_X : 5 \leq x \leq 15\}$ and calculate that

$$\mathbb{P}[X \in B] = \mathbb{P}[5 \leq X \leq 15] = \sum_{x=5}^{15} P_X(x) = 11 \cdot \frac{1}{29} = \frac{11}{29} \ .$$

The conditional PMF is thus

$$P_{X|B}(x) = \begin{cases} \dfrac{P_X(x)}{\mathbb{P}[X \in B]} & x \in B \\ 0, & \text{otherwise} \end{cases} = \begin{cases} \frac{\frac{1}{29}}{\frac{11}{29}} & x = 5, 6, \ldots, 15 \\ 0, & \text{otherwise} \end{cases} = \begin{cases} \dfrac{1}{11} & x = 5, 6, \ldots, 15 \\ 0, & \text{otherwise,} \end{cases}$$

which corresponds to a Discrete Uniform$(5, 15)$ random variable.
(*Despite the pending calamity, Sokka is thrilled that they will not go over budget.*)

(e) Given Sokka's information, calculate the conditional expected value and conditional variance for the total inn cost.

Solution:

Using the formulas for the expected value and variance for a Discrete Uniform$(5, 15)$ random variable, we know that

$$\mathbb{E}[X|A] = \frac{5+15}{2} = 10 \qquad \mathsf{Var}[X|A] = \frac{(15-5+1)^2 - 1}{12} = \frac{121-1}{12} = 10 \ .$$

Using the linearity of expectation,

$$\mathbb{E}[Y|A] = \mathbb{E}[2X + 3|A] = 2\,\mathbb{E}[X|A] + 3 = 2 \cdot 10 + 3 = 23 \ ,$$

Using the variance of a linear function,

$$\mathsf{Var}[Y|A] = \mathsf{Var}[2X + 3|A] = 2^2\,\mathsf{Var}[X|A] = 4 \cdot 10 = 40 \ .$$

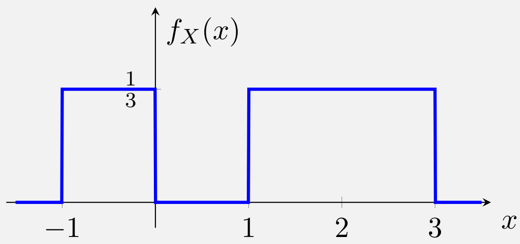
Problem 4.4 (**Video 3.1, 3.2**, **Lecture Problem**)

Consider a continuous random variable X with the following PDF:

$$f_X(x) = \begin{cases} c & -1 \leq x \leq 0 \text{ or } 1 \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

(a) Sketch the PDF of X. Be sure to label your axes.

Solution:



(b) Determine the value of c that satisfies the normalization property. Set c to this value for the remainder of the problem. This can be done without integration, so your answer should be a number.

Solution:

Since the PDF consists of two rectangles, we can directly calculate the total area to be $c(1+2) = 3c$. Setting $c = 1/3$ satisfies the normalization property.

(c) What is the expected value of X?

Solution:

$$\mathbb{E}[X] = \int_{-1}^{0} \frac{x}{3} dx + \int_{1}^{3} \frac{x}{3} dx = \frac{7}{6}.$$

(d) What is the variance of X?

Solution:

$$\mathsf{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

where $\mathbb{E}[X]$ is from part (c) and

$$\mathbb{E}[X^2] = \int_{-1}^{0} \frac{x^2}{3} dx + \int_{1}^{3} \frac{x^2}{3} dx = 3$$

Thus, $\mathsf{Var}[X] = 3 - \left(\frac{7}{6}\right)^2 = \frac{59}{36}$.

(e) What is $\mathbb{E}[2X^2 - 3X + 1]$?

Solution:

Using linearity of expectation, $\mathbb{E}[2X^2 - 3X + 1] = 2\mathbb{E}[X^2] - 3\mathbb{E}[X] + 1$ and we have both of these expectations from parts (c) and (d), respectively. Thus,

$$\mathbb{E}[2X^2 - 3X + 1] = 2 \cdot 3 - 3 \cdot \frac{7}{6} + 1 = \frac{7}{2}.$$

(f) Calculate the probability that $|X|$ is less than 2. (Your answer can be an integral, but you can also take advantage of the simple structure of the PDF.)

Solution:

$$\mathbb{P}[|X| < 2] = \int_{-2}^{2} f_X(x)\,dx = \int_{-1}^{0} \frac{1}{3}\,dx + \int_{1}^{2} \frac{1}{3}\,dx = \frac{2}{3}$$

(g) Let $B = (-2, 2)$. Determine the conditional PDF $f_{X|B}(x)$ of X given the event $\{X \in B\}$.

Solution:

$$f_{X|B}(x) = \begin{cases} \dfrac{f_X(x)}{\mathbb{P}[X \in B]} & x \in B \\ 0 & x \notin B \end{cases} = \begin{cases} \dfrac{1}{2} & -1 \leq x \leq 0 \text{ or } 1 \leq x \leq 2 \\ 0 & \text{otherwise.} \end{cases}$$

(h) Calculate the probability that $X < 0$ given that $|X|$ is less than 2.

Solution:

$$\mathbb{P}\big[X < 0 \,\big|\, |X| < 2\big] = \int_{-\infty}^{0} f_{X|B}(x)\,dx = \int_{-1}^{0} \frac{1}{2}\,dx = \frac{1}{2}$$

(i) Calculate the conditional expectation $\mathbb{E}[X|B]$ of X given that $|X|$ is less than 2.

Solution:

$$\mathbb{E}[X|B] = \int_{-\infty}^{\infty} x\, f_{X|B}(x)\,dx = \int_{-1}^{0} x\,\frac{1}{2}\,dx + \int_{1}^{2} x\,\frac{1}{2}\,dx = \frac{1}{2}$$

Problem 4.5 (**Video 3.3, 3.4**, **Lecture Problem**) Let X be a Gaussian$(-1, 4)$ random variable.

(a) Calculate $\mathbb{P}[X < 2]$. You may leave your answer in terms of the standard normal CDF $\Phi(z)$.

Solution:

First, note that $F_X(x) = \Phi\big(\frac{x-\mu}{\sigma}\big) = \Phi\big(\frac{x-(-1)}{2}\big)$. Therefore, here we have that

$$\mathbb{P}[X < 2] = F_X(2) = \Phi\left(\frac{2-(-1)}{2}\right) = \Phi\left(\frac{3}{2}\right).$$

(b) Calculate $\mathbb{P}[X < 0|X < 2]$. You may leave your answer in terms of the standard normal CDF $\Phi(z)$.

Solution:

Using the definition of conditional probability,

$$\mathbb{P}[X < 0|X < 2] = \frac{\mathbb{P}[\{X < 0\} \cap \{X < 2\}]}{\mathbb{P}[X < 2]} = \frac{\mathbb{P}[X < 0]}{\mathbb{P}[X < 2]} \ .$$

Using the CDF from the previous part,

$$F_X(0) = \Phi\left(\frac{0 - (-1)}{2}\right) = \Phi\left(\frac{1}{2}\right) \ ,$$

which yields

$$\mathbb{P}[X < 0|X < 2] = \frac{\Phi(\frac{1}{2})}{\Phi(\frac{3}{2})} \ .$$

(c) Calculate $\mathbb{E}[2X + 3]$ and $\mathsf{Var}[2X + 3]$.

Solution:

Using the linearity of expectation,

$$\mathbb{E}[2X + 3] = 2\mathbb{E}[X] + 3 = 2(-1) + 3 = 1 \ .$$

Similarly, using the formula for the variance of a linear function,

$$\mathsf{Var}[2X + 3] = 2^2\mathsf{Var}[X] = 16 \ .$$

(d) Let $Y = 2X + 3$. What kind of a random variable is Y?

Solution:

A linear function of a Gaussian random variable is itself a Gaussian random variable. Since a Gaussian random variable is fully determined by its mean and variance, all we need to do is determine the mean and variance of $Y = 2X + 3$. However, we have already done this in the previous part, so we have that Y is a Gaussian$(1, 16)$ random variable.