ENGRD 2700: Basic Engineering Probability and Statistics
Fall 2019

Homework 4

Due Friday, Oct 11th 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. Let X have the following PMF:

$$f(x) = P(X = x) = \frac{1}{n}, \qquad x = 0, 1, 2, \ldots, n-1$$

(a) Find the cumulative distribution function, $F(x)$. Recall that $F(x)$ is defined for all $x \in (-\infty, \infty)$. When $x < 0$, $F(x) = 0$. Also, when $x \geq n-1$, $F(x) = 1$. The function has a jump of size $1/n$ whenever x is in the support of X. Putting everything together yields

$$F(x) = \begin{cases} 0 & x < 0 \\ 1/n & 0 \leq x < 1 \\ 2/n & 1 \leq x < 2 \\ \cdots & \\ (n-1)/n & n-2 \leq x < n-1 \\ 1 & x \geq n-1 \end{cases}$$

(b) Compute $E[X]$ and $\mathrm{Var}(X)$. For this problem, you may find the following facts useful:

$$\sum_{i=0}^{n-1} i = \frac{n(n-1)}{2}, \qquad \sum_{i=0}^{n-1} i^2 = \frac{n(n-1)(2n-1)}{6}.$$

First, computing $E[X]$:

$$\begin{aligned} E[X] &= \sum_{x=0}^{n-1} xP(X = x) \\ &= \sum_{x=0}^{n-1} x \cdot \frac{1}{n} \\ &= \frac{1}{n} \sum_{x=0}^{n-1} x \\ &= \frac{1}{n} \cdot \frac{n(n-1)}{2} \\ &= \frac{n-1}{2} \end{aligned}$$

where in the fourth line, we used a fact specified in the problem description.

We next find $E[X^2]$:

$$\begin{aligned}
E[X^2] &= \sum_{x=0}^{n-1} x^2 P(X = x) \\
&= \frac{1}{n} \sum_{x=0}^{n-1} x^2 \\
&= \frac{1}{n} \cdot \frac{n(n-1)(2n-1)}{6} \\
&= \frac{(n-1)(2n-1)}{6}
\end{aligned}$$

where in the last line, we again used a fact in the problem description. Putting everything together:

$$\begin{aligned}
\mathrm{Var}(X) &= E[X^2] - (E[X])^2 \\
&= \frac{(n-1)(2n-1)}{6} - \left[\frac{n-1}{2}\right]^2 \\
&= \frac{2n^2 - 3n + 1}{6} - \frac{n^2 - 2n + 1}{4} \\
&= \frac{4n^2 - 6n + 2}{12} - \frac{3n^2 - 6n + 3}{12} \\
&= \frac{n^2 - 1}{12}.
\end{aligned}$$

2. Suppose that a baseball game between teams A and B is tied at the end of the 9th inning. To determine a winner, extra innings will be played until one team scores more runs in an inning than the other. Suppose that the probability that A scores more than B in an inning is 0.2, the probability that B scores more than A in an inning is 0.3, and the probability that the score remains tied is 0.5. The innings are independent of each other.

   (a) What is the distribution of the number X of extra innings that need to be played until a winner is determined (including the last one where one team scores more runs)? To answer this question, state the possible values of X and give a formula for $P(X = k)$ for all relevant k.
   The variable X has a geometric distribution, with $P(X = k) = 0.5^{k-1} 0.5 = 0.5^k$, valid for $k = 1, 2, \ldots$

   (b) What is the probability that at least 5 extra innings are required to determine a winner?
   At least 5 extra innings are required if and only if the first 4 innings are drawn. This happens with probability $0.5^4 = 0.0625$. This can also be computed by using $P(X \geq 5) = 1 - P(X = 1) - P(X = 2) - P(X = 3) - P(X = 4) = 0.0625$.

   (c) Eventually, one of the two teams wins. Compute the probability that B eventually wins. (Optional: You may try to use the same reasoning to compute the probability of A eventually winning. If the two probabilities do not sum to 1, then something is wrong.)
   P(B wins) $= \sum_{k=1}^{\infty} P(B \text{ wins and } X = k\ ) = \sum_{k=1}^{\infty} 0.5^{k-1} 0.3 = 0.3/(1 - 0.5) = 0.6$.
   Alternatively, $P(B \text{ wins }) = P(B \text{ wins in an extra inning} \mid \text{game finishes in that extra inning}) = P(B \text{ wins in an extra inning})\ / P(A \text{ wins or } B \text{ wins an the extra inning}) = .3/(.3 + .2) = 0.6$.

3. Suppose X is a continuous random variable with probability density function (pdf)

$$f(x) = \begin{cases} cx^{-6} & 1 \leq x < \infty \\ 0 & x < 1 \end{cases}$$

   (a) Find c.
   Since the pdf must integrate to one, we have

$$\int_{-\infty}^{\infty} f(x)\, dx = \int_{1}^{\infty} \frac{c}{x^6}\, dx = -\frac{c}{5x^5}\bigg|_1^{\infty} = \frac{c}{5}.$$

It follows that $c = 5$.

(b) Compute $F(x)$, the cumulative distribution function (cdf) of X.
When $x < 1$, $F(x) = 0$. If $x \geq 1$, then

$$P(X \leq x) = \int_1^x \frac{5}{t^6}\,dt = -\frac{1}{t^5}\bigg|_1^x = 1 - \frac{1}{x^5}.$$

Thus, the cdf of X is

$$F(x) = \begin{cases} 0 & x < 1 \\ 1 - \frac{1}{x^5} & x \geq 1. \end{cases}$$

(c) Find the 40^{th} percentile of X, i.e., the number γ such that $P(X \leq \gamma) = 0.4$.
We seek γ such that

$$1 - \frac{1}{\gamma^5} = 0.4.$$

Algebra yields that $\gamma = 0.6^{-\frac{1}{5}} \approx 1.1076$.

(d) Compute $E[X]$ and $\mathrm{Var}(X)$.
First, finding $E[X]$ and $E[X^2]$:

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x f(x)\,dx = \int_1^{\infty} x\frac{5}{x^6}\,dx = -\frac{5}{4x^4}\bigg|_1^{\infty} = \frac{5}{4} \\ E[X^2] &= \int_1^{\infty} x^2\frac{5}{x^6}\,dx = -\frac{5}{3x^3}\bigg|_1^{\infty} = \frac{5}{3}, \end{aligned}$$

and so $\mathrm{Var}(X) = 5/3 - (5/4)^2 = 5/48 = 0.1042$.

(e) Compute $E[X^5]$.
We have that

$$E[X^5] = \int_1^{\infty} x^5\frac{5}{x^6}\,dx = 5\ln(x)\bigg|_1^{\infty} = \infty.$$

4. Suppose you and your friend just finished shopping at Wegmans and are checking out separately. You use one of those self-checkout kiosks and know that you will finish checking out in exactly 5 minutes. Your friend is waiting in a traditional checkout lane, and the amount of time until your friend finishes checking out is uniformly distributed between 0 and 15 minutes. The two of you will leave Wegmans together when you both finish checking out. Find $E[T]$, where T is the number of minutes between now and the time you leave.
Let X be a $\mathrm{Uniform}(0, 15)$ random variable, which has density

$$f(x) = \begin{cases} \frac{1}{15} & 0 \leq x \leq 15 \\ 0 & \text{otherwise} \end{cases}.$$

We have that $T = \max\{X, 5\}$, which has expectation

$$\begin{aligned} E[T] &= \int_0^{15} \max\{x, 5\}\frac{1}{15}\,dx \\ &= \int_0^5 5\,\frac{1}{15}\,dx + \int_5^{15} x\,\frac{1}{15}\,dx \\ &= \frac{1}{3}\int_0^5 1\,dx + \frac{1}{15}\int_5^{15} x\,dx \\ &= \frac{25}{3} \approx 8.3333. \end{aligned}$$

5. The gain from one share of stock in company i over the coming year is X_i, where $i = 1, \ldots, 10$. (Negative values of X_i represent losses.) Suppose that the X_i are independent Normal$(100, 196)$ random variables (i.e., with mean $\mu = 100$ and variance $\sigma^2 = 196$).

   For the follow questions, **you should first derive an expression of the probability in terms of the CDF $\Phi(\cdot)$ of a standard normal r.v., and then give a numerical answer.**

   (a) Compute $P(X_1 \geq 90)$, the probability that company 1's gain is at least 90% of its expectation.
   Note that
   $$\frac{X_1 - 100}{\sqrt{196}} \sim Z,$$
   i.e., a standard normal random variable. Thus
   $$P(X_1 \geq 90) = P\left(\frac{X_1 - 100}{14} \geq \frac{90 - 100}{14}\right) = 1 - \Phi\left(-\frac{10}{14}\right) \approx 0.7612,$$
   where $\Phi(x)$ is the cdf of the standard normal random variable.

   (b) Compute $P(\sum_{i=1}^{10} X_i \geq 900)$, the probability that the combined gain of all 10 companies is at least 90% of the expectation.
   Because the X_i are independent, it follows that $\sum_{i=1}^{10} X_i$ is also a normal random variable, but with mean $10 \cdot 100 = 1000$ and variance $10 \cdot 196 = 1960$. Thus
   $$\begin{aligned} P\left(\sum_{i=1}^{10} X_i \geq 900\right) &= P\left(\frac{\sum_{i=1}^{10} X_i - 1000}{\sqrt{1960}} \geq \frac{900 - 1000}{\sqrt{1960}}\right) \\ &= 1 - \Phi\left(-\frac{100}{\sqrt{1960}}\right) \\ &\approx 0.9881. \end{aligned}$$
   Comparing parts (a) and (b), one interesting observation is that pooling reduces risk.

   (c) Compute $P(X_1 - 2X_2 \geq 10)$.
   Because X_1 and X_2 are independent, $X_1 - 2X_2$ is also normally distributed, but with mean $100 - 2 \cdot 100 = -100$ and variance $(1 + 2^2) \cdot 196 = 980$. (Here, we're using the fact that $-2X_2$ is also a normal random variable.) Thus
   $$P(X_1 - 2X_2 \geq 10) = P\left(\frac{X_1 - 2X_2 + 100}{\sqrt{980}} \geq \frac{10 + 100}{\sqrt{980}}\right) = 1 - \Phi\left(\frac{110}{\sqrt{980}}\right) \approx 0.0002.$$

6. Suppose that X is an exponential random variables with parameter $\lambda = 3$.

   (a) What is the cumulative distribution function of $Y = -2X + 2$?
   We have
   $$\begin{aligned} F_Y(y) = P(Y \leq y) = P(-2X + 2 \leq y) = P(X \geq 1 - \frac{y}{2}) = 1 - F_X(1 - \frac{y}{2}) \\ = \begin{cases} e^{\frac{3}{2}y - 3} & y \leq 2, \\ 1 & y > 2. \end{cases} \end{aligned}$$

   (b) What is the mean and variance of Y?
   Note that since $X \sim$ Exponential(3), we have $E[X] = \frac{1}{3}$ and $\text{Var}(X) = \frac{1}{9}$. Using linearity of expectation, we see that $E[Y] = E[-2X + 2] = -2E[X] + 2 = -2\left(\frac{1}{3}\right) + 2 = \frac{4}{3}$ and $\text{Var}(Y) = \text{Var}(-2X + 2) = 2^2\text{Var}(X) = \frac{4}{9}$.

   (c) What is the 0.9 quantile of Y?
   We want to find $\hat{y}$ satisfying $F_Y(\hat{y}) = 0.9$. We have
   $$\begin{aligned} F_Y(\hat{y}) &= e^{\frac{3}{2}\hat{y} - 3} = 0.9 \\ \iff \log(0.9) &= \frac{3}{2}\hat{y} - 3, \end{aligned}$$
   hence we arrive at $\hat{y} = 2 + \frac{2}{3}\log(0.9) \approx 1.93$.

(d) What is the probability density function of Y?

From part (a), we see that $F_Y(y) = \begin{cases} e^{\frac{3}{2}y-3} & y \leq 2 \\ 1 & y > 2. \end{cases}$

Hence we differentiate to obtain the density:

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \begin{cases} \frac{3}{2}e^{\frac{3}{2}y-3} & y \leq 2 \\ 0 & y > 2. \end{cases}$$