STAT 511 su 2020 hw 4

pmfs, pdfs, expected values, variance

At some point in this assignment you will need to use the identities

$$\sum_{i=1}^{n} i = n(n+1)/2, \quad \sum_{i=1}^{n} i^2 = n(n+1)(2n+1)/6, \quad \sum_{i=1}^{n} i^3 = n^2(n+1)^2/4.$$

1. Consider being dealt a hand of 5 cards from a 24-card Euchre deck, which has 9s, 10s, face cards, and aces in each of the four suits. Let the random variable X be the number of hearts in your hand.

   (a) Tabulate the probability distribution of X in a table like

x	$\dots$
$P_X(X = x)$	$\dots$

The probability of drawing x hearts is

$$P_X(X = x) = \frac{\binom{6}{x}\binom{18}{5-x}}{\binom{24}{5}} \quad \text{for } x = 0, 1, \dots, 5.$$

So we have

x	0	1	2	3	4	5
$P_X(X = x)$	0.2016	0.4320	0.2880	0.0720	0.0064	0.0001

We can get these from R very easily with

```
choose(6,0:5)*choose(18,5:0)/choose(24,5).
```

   (b) Find $P(X \leq 2)$.

$0.2016 + 0.4320 + 0.2880 = 0.9216.$

   (c) Find $\mathbb{E}X$.

We have

$$\mathbb{E}X = 0.2016(0) + 0.4320(1) + 0.2880(2) + 0.0720(3) + 0.0064(4) + 0.0001(5) = 1.2501.$$

Without rounding error the answer is 1.25, which we can get from R with

```
sum(0:5*choose(6,0:5)*choose(18,5:0)/choose(24,5)).
```

Note that this is equal to $(6/24)5$.

(d) Find $\mathrm{Var}\,X$.

We should first get $\mathbb{E}X^2$, which is

$$\mathbb{E}X^2 = 0.2016(0)^2 + 0.4320(1)^2 + 0.2880(2)^2 + 0.0720(3)^2 + 0.0064(4)^2 + 0.0001(5)^2 = 2.3369$$

More precisely, we have

$$\mathbb{E}X^2 = \texttt{sum((0:5)**2*choose(6,0:5)*choose(18,5:0)/choose(24,5))} = 2.336957.$$

So

$$\mathrm{Var}\,X = 2.336957 - (1.25)^2 = 0.774457.$$

2. The random variable Y represents the amount of flooding over a baseline equal to 1 in a certain city. It has the cdf

$$F_Y(y) = \begin{cases} 1 - \frac{1}{y^2}, & 1 \le y < \infty \\ 0, & -\infty < y < 1 \end{cases}$$

(a) Draw an accurate picture of F_Y.

(b) Verify that F_Y is a cdf (3 properties to verify).

(c) Find $P_Y(1 < Y < 2)$.

$$P_Y(1 < Y < 2) = F_Y(2) - F_Y(1) = (1 - 1/4) - (1 - 1) = 3/4.$$

(d) Find the pdf f_Y of Y. *Hint: consider separately the intervals* $-\infty < y < 1$ *and* $1 \le y < \infty$.

We have

$$f_Y(y) = \begin{cases} \dfrac{2}{y^3}, & y \ge 1 \\ 0 & y < 1. \end{cases}$$

(e) Draw an accurate picture of f_Y.

(f) Find the value M such that $P(Y \le M) = 1/2$. The value M is the median of Y.

We have

$$1/2 = \int_1^M \frac{2}{y^3} dy = \frac{1}{y^2}\Big|_1^M = 1 - \frac{1}{M} \qquad \Longleftrightarrow \qquad M = \sqrt{2}.$$

(g) Find $\mathbb{E}Y$.

$$\mathbb{E}Y = \int_1^\infty y \cdot \frac{2}{y^3} dy = -\frac{2}{y}\Big|_1^\infty = 2.$$

(h) Find $\operatorname{Var} Y$.

We have

$$\mathbb{E}Y^2 = \int_1^\infty y^2 \cdot \frac{2}{y^3} = 2\log y\Big|_1^\infty \to \infty.$$

Since the integral diverges Y does not have finite variance.

(i) Are we able to apply Chebychev's inequality to this random variable in order to find an interval within which it will lie with probability exceeding some lower bound? Explain why or why not.

This random variable does not have finite variance, so we cannot use Chebychev's inequality.

3. Let X be a random variable with pdf given by

$$f_X(x) = Cx^2(1-x)\mathbf{1}(0 < x < 1),$$

where $C > 0$ and $\mathbf{1}(\cdot)$ is the indicator function.

(a) Find the value of the constant C such that f_X is a valid pdf.

Since the pdf must integrate to 1, we have

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} Cx^2(1-x)\mathbf{1}(0 < x < 1)dx \\ &= \int_{-1}^{1} Cx^2(1-x)dx \\ &= C\left[\frac{x^3}{3} - \frac{x^4}{4}\right]\Big|_0^1 \\ &= C\left[\frac{1}{3} - \frac{1}{4}\right] \\ &= \frac{C}{12} \end{aligned}$$

which gives

$$C = 12.$$

(b) Find $P(1/2 \le X < 1)$.

We have

$$P(1/2 \le X < 1) = \int_{1/2}^{1} 12x^2(1-x)dx = 12\left[\frac{x^3}{3} - \frac{x^4}{4}\right]\Big|_{1/2}^{1} = \frac{11}{16}.$$

(c) Find $P(X \le 1/2)$.

$$P(X \le 1/2) = 1 - P(1/2 \le X < 1) = 1 - \frac{11}{16} = \frac{5}{16}.$$

(d) Find $P(X = 1/2)$.

This is equal to 0 because X is continuous.

(e) Find $P(1 \le X \le 2)$.

This is equal to zero because the interval lies outside of the support of X.

(f) Find $\mathbb{E}X$.

We have

$$\mathbb{E}X = \int_{0}^{1} x \cdot 12x^2(1-x)dx = 12\left[\frac{x^4}{4} - \frac{x^5}{5}\right]\Big|_{0}^{1} = \frac{3}{5}.$$

4. For any rv X, verify the identity $\operatorname{Var} X = \mathbb{E}X^2 - (\mathbb{E}X)^2$.

We have

$$\begin{aligned}
\operatorname{Var} X &= \mathbb{E}(X - \mathbb{E}X)^2 \\
&= \mathbb{E}(X^2 - 2X\mathbb{E}X + (\mathbb{E}X)^2) \\
&= \mathbb{E}X^2 - 2(\mathbb{E}X)(\mathbb{E}X) + (\mathbb{E}X)^2 \\
&= \mathbb{E}X^2 - (\mathbb{E}X)^2.
\end{aligned}$$

5. Let X be a rv such that $\mathbb{E}X = \mu$ and $\operatorname{Var} X = \sigma^2$.

(a) Show that $\operatorname{Var}(aX + b) = a^2\sigma^2$.

(b) Show that $\mathbb{E}(aX + b) = a\mu + b$.

6. Consider rolling a single K-sided die and let X be the rv defined as the number on the up-face.

   (a) Give the pmf of the rv X.

   We have

$$p_X(x) = \begin{cases} \frac{1}{K}, & x = 1, 2, \ldots, K \\ 0, & \text{otherwise} \end{cases}$$

   (b) Find $\mathbb{E}X$.

   We have

$$\mathbb{E}X = \sum_{x=1}^{K} x \cdot \frac{1}{K} = \frac{1}{K}\sum_{x=1}^{K} x = \frac{1}{K}\frac{K(K+1)}{2} = \frac{K+1}{2}.$$

   (c) Find $\operatorname{Var} X$.

   We have

$$\mathbb{E}X^2 = \sum_{x=1}^{K} x^2 \cdot \frac{1}{K} = \frac{1}{K}\sum_{x=1}^{K} x^2 = \frac{1}{K}\frac{K(K+1)(2K+1)}{6} = \frac{(K+1)(2K+1)}{6},$$

   so

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

   (d) Give the expected value of the roll of a 6-sided die and interpret this quantity in your own words.

   The expected value is $(6+1)/2 = 3.5$. We can never roll the value 3.5, so it is *not a value we expect to observe*—rather, if we were to roll a die many many times, we would expect the average of these many many rolls to be very close to 3.5.

7. Consider rolling two K-sided dice with faces numbered $1, \ldots, K$ and let X be the rv defined as the maximum of the two up-faces. The pmf of X is given by

$$p_X(x) = (2x-1)/K^2, \quad x = 1, \ldots, K.$$

(a) Show that $p_X(x)$ sums to 1 over the support of X.

We have

$$\sum_{x=1}^{K} \frac{2x-1}{K^2} = \frac{1}{K^2}\left[2\frac{K(K+1)}{2} - K\right] = 1$$

(b) Find $\mathbb{E}X$.

We have

$$\begin{aligned}
\mathbb{E}X &= \sum_{x=1}^{K} x \cdot \frac{2x-1}{K^2} \\
&= \frac{1}{K^2}\left[2\sum_{x=1}^{K} x^2 - \sum_{x=1}^{K} x\right] \\
&= \frac{1}{K^2}\left[2\frac{K(K+1)(2K+1)}{6} - \frac{K(K+1)}{2}\right] \\
&= \frac{K+1}{K}\left[\frac{2(2K+1)-3}{6}\right] \\
&= \frac{(K+1)(4K-1)}{6K}
\end{aligned}$$

(c) Find $\mathbb{E}X^2$.

We have

$$\begin{aligned}
\mathbb{E}X^2 &= \sum_{x=1}^{K} x^2 \cdot \frac{2x-1}{K^2} \\
&= \frac{1}{K^2}\left[2\sum_{x=1}^{K} x^3 - \sum_{x=1}^{K} x^2\right] \\
&= \frac{1}{K^2}\left[2\frac{K^2(K+1)^2}{4} - \frac{K(K+1)(2K+1)}{6}\right] \\
&= \frac{K(K+1)}{K^2}\left[\frac{3K(K+1)-(2K+1)}{6}\right] \\
&= \frac{(K+1)}{K}\left[\frac{3K^2+K-1}{6}\right] \\
&= \frac{3K^3+4K^2-1}{6K}.
\end{aligned}$$

(d) Give an expression for $\mathrm{Var}\, X$ (does not have to be simplified) and compute $\mathrm{Var}\, X$ for $K = 10$.

We have

$$\mathrm{Var}\, X = \frac{3K^3 + 4K^2 - 1}{6K} - \left[\frac{(K+1)(4K-1)}{6K}\right]^2.$$

(e) Let $K = 10$ and use Chebyschev's theorem to find an interval within which X will fall with probability at least $8/9$.

If $K = 10$, we have $\mathrm{Var}\, X = 56.65 - (7.15)^2 = 5.528$. From Chebychev's theorem, X should fall within 3 standard deviations of its mean with probability at least $8/9$ (note that $1 - (1/3)^2 = 8/9$). So X will lie in the interval

$$7.15 \pm 3 \times \sqrt{5.528} = (0.0965, 14.20)$$

with probability at least $8/9$.

To double check our expressions for $\mathbb{E}X$, $\mathbb{E}X^2$, and $\mathrm{Var}\, X$, we can run a simulation in R. We can compute the value of X for many many rolls of two dice and then compute the mean of the values, the mean of the squared values, and the variance of the values. The following R code does this (try running it!):

```
S <- 100000 # set number of times to roll two dice
K <- 10
r1 <- sample(1:K,S,replace = TRUE) # roll 1st die S times
r2 <- sample(1:K,S,replace = TRUE) # roll 2nd die S times

# make a matrix with columns r1 and r2, and then take the maximum of each row
X <- apply(cbind(r1,r2),1,max)

mean(X)
mean(X^2)

var(X)
```

(f) Comment on the usefulness of the interval computed in the previous part.

Since the interval contains the entire support of the random variable X, it is not useful; we know that X must take one of the values $1, 2, \ldots, 10$.

Optional (do not turn in) problems for additional study from Wackerly, Mendenhall, Scheaffer, 7th Ed.:

- 3.12, 3.24, 3.30

- 4.6 (*Hint: Shown in Lec 04*), 4.7, 4.8, 4.11, 4.13, 4.16