Problem Section 6

Expected Value and Variance

Learning Outcomes

The problems are designed to build conceptual understanding and problem-solving skills. The emphasis is on learning to find, evaluate and build confidence. The specific tasks include:

- Calculate expected value and variance using the PMF
- Use Chebychev's inequality to make probability statements
- Back up and support work with relevant explanations

Exercises

1. (Law of the Unconscious Probabilist) Weird Wally offers you the following choice. You may have a flat amount of $\frac{1}{3.5}$ dollars or you may roll a fair die and he will give you $\frac{1}{X}$ dollars where X is the value on the roll. The PMF of X is shown in the table below.

x	1	2	3	4	5	6
$f(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Which is the better deal? Calculate $E\left[\frac{1}{X}\right]$ to decide. (Hint: use the Law of the Unconscious Probabilist - Lemma 7.2)

We can find $E\left[\frac{1}{X}\right]$ as follows by using the Law of the Unconscious Probabilist:

$$
\begin{aligned}
E\left[\frac{1}{X}\right] &= \sum_x \frac{1}{x} \cdot f(x) \\
&= \frac{1}{6} \sum_{x=1}^{6} \frac{1}{x}
\end{aligned}
$$

We may calculate this using R.

```
inv_x <- 1/(1:6)
E_1_x <- sum(inv_x)/6
E_1_x
```

```
## [1] 0.4083333
```

Using R we see that $E\left[\frac{1}{X}\right] = 0.4083333$ versus Wally's offer of 0.2857143. Thus rolling the die has a better expected value, and this is the better deal.

2. (Mean and variance)

x	-\$5	\$170
$f(x)$	37/38	1/38

a. Suppose a random variable X has PMF as shown below. Find $E[X]$. Also calculate $SD[X]$.

We may solve for $E(X)$ and $Var(X)$ using the definitions of these quantities:

$$E(X) = \sum_x xf(x) = -5 \times \frac{37}{38} + 170 \times \frac{1}{38}.$$

Since $Var(X) = E(X^2) - E(X)^2$. We know $E(X)$, so we can now find $E(X^2)$

$$E(X^2) = \sum_x x^2 f(x) = -5^2 \times \frac{37}{38} + 170^2 \times \frac{1}{38}$$

These are calculated below.

```
E_X <- -5*37/38 + 170/38
Var_X <- (25*37/38 + 170^2/38 - E_X^2)
SD_X <- sqrt(Var_X)    #SD = sqrt(Var(X))
```

Thus we have that $E(X) =$ -\$0.39 and $SD(X) =$ \$28.01.

The SD tells us the typical deviation in earnings from the average earnings ($E(X)$). It is really high in this example. It tells us that typically our net gain will be in the interval $[E(X) - SD(X), E(X) - SD(X)] = [-28.41, 27.62]$

b. If X denotes a temperature of a randomly selected day recorded in degrees Fahrenheit, then $Y = \frac{5}{9}X - \frac{160}{9}$ is the corresponding temperature in degrees Celsius. If the standard deviation for X is $15.7°F$, what is the standard deviation of Y?

By Lemma 7.4, we have the result that

$$Var(Y) = \left(\frac{5}{9}\right)^2 Var(X).$$

Since $Var(X) = 15.7^2$ we can say that

$$Var(Y) = \frac{25 \times 15.7 \times 15.7}{81}$$

and

$$SD(Y) = \sqrt{Var(Y)} = \frac{5 \times 15.7}{9}.$$

3. (Chebychev) Suppose X is a random variable with mean and variance both equal to 20. What can be said about $P(0 < X < 40)$?

Hint: Chebychev's inequality says that

$$P(|X - 20| \geq k\sqrt{20}) \leq \frac{1}{k^2}.$$

What would you choose for k here so you can say something about $P(0 < X < 40)$?

If we choose $k = \sqrt{20}$ then by Chebychev's inequality, we can say

$$P(|X - 20| \geq 20) \leq \frac{1}{20}. \tag{1}$$

Now it is important to note that the event $0 < X < 40$ is actually the complement of the event $|X - 20| \geq 20$ since

$$
\begin{aligned}
0 < X < 40 &= (0 - 20) < X - 20 < (40 - 20), \\
&= -20 < X - 20 < 20, \\
&= |X - 20| < 20
\end{aligned}
$$

Therefore

$$
\begin{aligned}
P(0 < X < 40) &= 1 - P(|X - 20| \geq 20), \\
&\geq 1 - \frac{1}{20}
\end{aligned}
$$

where the last inequality follows by reversing the one in (1)

4. Suppose we wish to generate $X \sim Binom(n = 10, \pi = \frac{2}{3})$ subject to the constraint $X \leq 3$. Say we use the following naive algorithm to accomplish this task:

   - Generate an x from a $Binom(10, \frac{2}{3})$
   - Accept the value x if $x <= 3$. Otherwise reject it.

a. Calculate the acceptance probability. That is, what is the probability we will accept a value x that is generated?

We accept x if $x \leq 3$. Thus the acceptance probability is $P(X \leq 3)$. We can find this using `pbinom`

```
pi <- pbinom(q = 3, size = 10, prob = 2/3)
```

The acceptance probability is 0.0197.

b. Put your calculation from part a. to the test by generating 1,000 values from a $Binom(10, \frac{2}{3})$ and calculate how often you would generate an acceptable value. Bump up the number of simulations to 5,000. What do you notice?

```
set.seed (131)

nrep <- 1000
x <- rbinom(n = nrep, size = 10, prob = 2/3)

sum(x <= 3)/nrep          #fraction of accepted values
```

```
## [1] 0.015
```

c. Define a new random variable Y as the number of times we have to generate a binomial variable before we find an acceptable one. For example, if on our first try, we get an acceptable x, then $y = 0$. Write the PMF of Y.

The random variable Y represents the number of failures before we find an acceptable value. It therefore takes values $0, 1, 2, \ldots$.

If $Y = 0$, then it means there were 0 failures and we accepted the first value generated. Therefore $P(Y = 0) = \pi$ where π is the acceptance probability we found in part a.

If $Y = 1$, then it means there was 1 failure and then we generated an acceptable value. Therefore $P(Y = 1) = (1 - \pi) \times \pi$. We are of course relying on the fact that the outcomes of each random generation from the binomial PMF are independent of each other.

Similarly if $Y = 2$, then it means there were 2 failures and then we generated an acceptable value. Therefore $P(Y = 2) = (1 - \pi)^2 \times \pi$.

Hence the pattern is as follows.

$$f(y) = (1-\pi)^y \times \pi, \quad y = 0, 1, 2, \dots$$