

Chapter 3 Group Work SOLUTIONS

Contents

Simulation Activity Solutions	2
Back of Packet Group Work Solutions	4

Simulation Activity Solutions

Part 1 - psuedocode

Answers may vary.

Part A

```
toss1 <- rbernoulli(1, 0.7)
toss2 <- rbernoulli(1, 0.7)
toss3 <- rbernoulli(1, 0.7)
```

Part B

```
X <- toss1 + toss2 + toss3 # number of heads
Y <- toss2 # logical for whether toss2 is heads
tails <- number of tosses - number of heads
Z <- number of heads - number of tails

#alternative way of calculating Z
Z <- X - (3 - X)
```

Part C

```
#just change 1st argument (# of sims)
toss1 <- rbernoulli(10000, 0.7)
toss2 <- rbernoulli(10000, 0.7)
toss3 <- rbernoulli(10000, 0.7)

#X, Y, Z are then calculated in same way as above
X <- toss1 + toss2 + toss3 # number of heads
Y <- toss2 # logical for whether toss2 is heads
Z <- X - (3 - X)
```

Part D

For X, count the number of times the result is 0, and divide by 10,000. This proportion will be a good estimate of $P(X = 0) = p(0)$. The same is done for $X = 1, 2, 3$. This will define the probability distribution. Plotting the 10,000 values of X would give you a visual of the distribution.

For Y, follow the same logic to count the number of times $Y = 0$ and $Y = 1$ and divide each by 10,000. These two proportions specify the pmf.

For Z, follow the same logic to count the number of times $Z = -3, -1, 1, 3$ and divide by 10,000. These 4 proportions specify the pmf.

Part 2 - pmfs

Distribution of X

Support of X is

$$\mathbb{S} = \{0, 1, 2, 3\}$$

- 1 way to have 0 heads: *ttt*, with probability $(0.25)^3$
- 3 ways to have 1 head: *htt, tht, tth*, each with probability $(0.7)(0.3)^2$
- 3 ways to have 2 heads: *hht, hth, thh*, each with probability $(0.7)^2(0.3)$
- 1 way to have 3 heads: *hhh*, with probability $(0.75)^3$

So the pmf of X is given by:

$$f(x) = \begin{cases} 0.027, & x = 0 \\ 0.189, & x = 1 \\ 0.441, & x = 2 \\ 0.343, & x = 3 \end{cases}$$

Distribution of Y

Support of Y is

$$\mathbb{S} = \{0, 1\}$$

Y is simply the result of the second coin toss, so the pmf of Y is defined simply as:

$$f(y) = \begin{cases} 0.3, & y = 0 \\ 0.7, & y = 1 \end{cases}$$

Distribution of Z

Note that since Z is the number of heads minus the number of tails, it can be written as $Z = X - (3 - X) = 2X - 3$

Therefore, the support and pmf of Z can be derived easily from the support and pmf of X.

Since X can be 0,1,2,3, the support of Z is

$$\mathbb{S} = \{-3, -1, 1, 3\}$$

Z will take on each of these values with the same probabilities as the pmf of X.

X	Z = 2X - 3	f(x) = f(z)
0	-3	0.027
1	-1	0.189
2	1	0.441
3	3	0.343

Back of Packet Group Work Solutions

Problem 1

For each of the following, determine the constant c so that $f(x)$ satisfies the conditions of being a pmf for a random variable X . That is, find c such that $\sum_{x \in S} f(x) = 1$.

- a. $f(x) = x/c, \quad x = 1, 2, 3, 4$
- b. $f(x) = cx, \quad x = 1, 2, 3, \dots, 10$
- c. $f(x) = c(1/4)^x, \quad x = 1, 2, 3, \dots$

Hint: use the infinite series identity from calculus that tells us

$$\sum_{n=1}^{\infty} a_1(r)^{n-1} = \frac{a_1}{1-r}$$

Solution part a

$$\begin{aligned} 1 &= \sum_{x=1}^4 f(x) \\ &= \sum_{x=1}^4 \frac{x}{c} \\ &= \frac{1}{c} + \frac{2}{c} + \frac{3}{c} + \frac{4}{c} \\ &= \frac{10}{c} \quad \implies c = 10 \end{aligned}$$

Solution part b

$$\begin{aligned} 1 &= \sum_{x=1}^{10} f(x) \\ &= \sum_{x=1}^{10} cx \\ &= c \sum_{x=1}^{10} x \\ &= c(1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10) \\ &= c55 \\ &\implies c = \frac{1}{55} \end{aligned}$$

Solution part c

$$\begin{aligned} 1 &= \sum_{x=1}^{\infty} f(x) \\ &= \sum_{x=1}^{\infty} c \left(\frac{1}{4}\right)^x \\ &= \sum_{x=1}^{\infty} c \frac{1}{4} \left(\frac{1}{4}\right)^{x-1} \end{aligned}$$

Then $a_1 = \frac{c}{4}$ and $r = \frac{1}{4}$, and we can use the fact that the geometric series $\sum_{n=1}^{\infty} a_1(r)^{n-1} = \frac{a_1}{1-r}$. So we have

$$\begin{aligned} 1 &= \frac{c/4}{1 - 1/4} \\ &= \frac{c/4}{3/4} \\ &= \frac{c}{3} \\ &\implies c = 3 \end{aligned}$$

Problem 2

Recall the non-uniform example where we rolled a fair four-sided die twice and let X be the maximum of the two outcomes. We determined that the pmf of X was given by

$$f(x) = \frac{2x-1}{16}, \quad x = 1, 2, 3, 4$$

. Define the cdf of X . That is define $F(x) = P(X \leq x)$ for each value of x in the support.

Solution

First recall:

- $f(1) = P(X = 1) = 1/16$
- $f(2) = P(X = 2) = 3/16$
- $f(3) = P(X = 3) = 5/16$
- $f(4) = P(X = 4) = 7/16$

To define the cdf, we define $F(x) = P(X \leq x)$ for each value $x = 1, 2, 3, 4$:

$$\begin{aligned} F(1) &= P(X \leq 1) \\ &= P(X = 1) \\ &= 1/16 \end{aligned}$$

$$\begin{aligned} F(2) &= P(X \leq 2) \\ &= P(X = 1) + P(X = 2) \\ &= 1/16 + 3/16 \\ &= 4/16 \end{aligned}$$

$$F(3) = P(X \leq 3) = P(X = 1) + P(X = 2) + P(X = 3) = 1/16 + 3/16 + 5/16 = 9/16$$

$$F(4) = P(X \leq 4) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = 1/16 + 3/16 + 5/16 + 7/16 = 16/16 = 1$$

Note that this can be written as a general expression $F(x) = \frac{x^2}{16}$, $x = 1, 2, 3, 4$

Problem 3

$$\begin{aligned} E(X) &= 0(.027) + 1(.189) + 2(.441) + 3(.343) \\ &= 2.1 \end{aligned}$$

$$\begin{aligned} E(Y) &= 0(.3) + 1(.7) \\ &= 0.7 \end{aligned}$$

$$\begin{aligned} E(Z) &= -3(.027) - 1(.189) + 1(.441) + 3(.343) \\ &= 1.2 \end{aligned}$$

Problem 4

Let $E(X) = 4$. Find

- $E(Y)$ when $Y = 2X + 3$
- $E(Z)$ when $Z = 7 - 5X$
- $E(32)$

Solution part a

$$\begin{aligned} E(Y) &= E(2X + 3) \\ &= E(2X) + E(3) \\ &= 2E(X) + 3 \\ &= 2(4) + 3 \\ &= 11 \end{aligned}$$

Alternatively, could show it from definition of Expected value:

$$\begin{aligned} E(Y) &= E(2X + 3) \\ &= \sum_{x \in S} (2x + 3)f(x) \\ &= \sum_{x \in S} (2xf(x) + 3f(x)) \\ &= \sum_{x \in S} 2xf(x) + \sum_{x \in S} 3f(x) \\ &= 2 \sum_{x \in S} xf(x) + 3 \sum_{x \in S} f(x) \\ &= 2E(X) + 3(1) \\ &= 2(4) + 3 \\ &= 11 \end{aligned}$$

Solution part b

$$\begin{aligned}E(Z) &= E(7 - 5X) \\&= 7 - 5E(X) \\&= 7 - 5(4) \\&= -13\end{aligned}$$

Solution part c

$E(32) = 32$ by part 1 of Theorem 2.2-1

Problem 5

Let $f(x) = \frac{x}{10}$, $x = 1, 2, 3, 4$. Find:

- a. $E(X)$
- b. $E(X^2)$
- c. $E(X(5 - X))$

Solution part a

$$\begin{aligned}E(X) &= \sum x f(x) \\&= \sum x \frac{x}{10} \\&= \frac{1}{10} \sum x^2 \\&= \frac{1}{10} (1^2 + 2^2 + 3^2 + 4^2) \\&= \frac{1}{10} (30) \\&= 3\end{aligned}$$

Solution part b

$$\begin{aligned}E(X^2) &= \sum x^2 f(x) \\&= \sum x^2 \frac{x}{10} \\&= \frac{1}{10} \sum x^3 \\&= \frac{1}{10} (1^3 + 2^3 + 3^3 + 4^3) \\&= \frac{1}{10} (100) \\&= 10\end{aligned}$$

Solution part c

$$\begin{aligned}
E[X(5 - X)] &= E[5X - X^2] \\
&= E(5X) - E(X^2) \\
&= 5E(X) - E(X^2) \\
&= 5(3) - 10 \\
&= 5
\end{aligned}$$

Problem 6

Let $E(X) = 5$ and $V(X) = 36$. Find:

- a) $V(3X + 7)$
- b) $V(2 - X)$
- c) $E(X^2)$
- d) $E(5X + 2X^2)$

Solution part a

$$V(3X + 7) = 3^2 V(X) = 9 * 36 = 324$$

Solution part b

$$V(2 - X) = V(-X + 2) = (-1)^2 V(X) = V(X) = 36$$

Solution part c

Note

$$V(X) = E(X^2) - [E(X)]^2 \implies E(X^2) = V(X) + [E(X)]^2$$

Therefore,

$$E(X^2) = 36 + 5^2 = 61$$

Solution part d

$$\begin{aligned}
E(5X + 2X^2) &= E(5X) + E(2X^2) \\
&= 5E(X) + 2E(X^2) \\
&= 5(5) + 2(61) \\
&= 147
\end{aligned}$$

Problem 7

In a lab experiment involving inorganic syntheses of molecular precursors to organometallic ceramics, the final step of a five-step reaction involves the formation of a metal-metal bond. The probability of such a bond forming is $p = 0.2$. Let X equal the number of successful reactions out of $n = 25$ such experiments.

- a) What distribution is appropriate to model X ?
- b) Write out the pmf of X

- c) Find the probability that $X = 1$
- d) Find the probability that X is at least 1
- e) Find the mean, variance, and standard deviation of X

Solution part a

$$X \sim \text{binomial}(25, 0.2)$$

Solution part b

$$p(x) = \binom{25}{x} (0.2)^x (0.8)^{25-x}$$

Solution part c

$$\begin{aligned} P(X = 1) &= f(1) = \binom{25}{1} (0.2)^1 (0.8)^{25-1} \\ &= \binom{25}{1} (0.2)^1 (0.8)^{24} \\ &= 25(.2)(.8)^{24} \\ &= 0.0236 \end{aligned}$$

Solution part d

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) \\ &= 1 - \binom{25}{0} (0.2)^0 (0.8)^{25-0} \\ &= 1 - 1(1)(.8)^{25} \\ &= 0.9962 \end{aligned}$$

Solution part e

Recall that $E(X) = np$ and $V(X) = np(1-p)$ for a binomial distribution. Therefore in this case, we have

$$\begin{aligned} E(X) &= 25(0.2) = 5 \\ \sigma^2 = V(X) &= 25(0.2)(0.8) = 4 \end{aligned}$$

$$\sigma = \sqrt{V(X)} = \sqrt{4} = 2$$

Problem 8

A random variable X has a binomial distribution with mean 6 and variance 3.6. Find $P(X = 4)$.

Solution

We are given that $E(X) = np = 6$ and $V(X) = np(1 - p) = 3.6$. Therefore, we can plug in and set $6(1 - p) = 3.6$. Solving for p gives $p = 0.4$. Plugging back into $np = 6$ gives $n = 15$. We can then write the pmf of X as

$$f(x) = \binom{15}{x} p^x (1 - p)^{15-x}$$
$$P(X = 4) = f(4) = \binom{15}{4} (0.4)^4 (0.6)^{11} = 0.1268$$

Problem 10

Return to the Joel Embiid example (he makes 87% of his free throws, he's fouled shooting a 3-pointer). Describe in detail how you could, in principle, perform a by hand simulation involving physical objects (e.g. coins, dice, spinners, etc.) to estimate the probability that the Lakers lose. Be sure to describe (1) what one repetition of the simulation entails, and (2) how you would use the results of many repetitions. Note: you do NOT need to compute any numerical values.

Solution

[Answers may vary]. One possible solution:

- (1) To simulate one repetition of the experiment, we could use a spinner where 87% of the area is shaded as “success” and 13% of the area is shaded as “failure”. We spin the spinner 3 times and record the number of successes. This gives us one value of the random variable X
- (2) We could repeat the process in part (1) 100,000 times. This would give us 100,000 values of the random variable X . Note that the Sixers lose if $X = 0$ or $X = 1$, they tie if $X = 2$, and they win if $X = 3$. So to find the probability that they lose, we count up the number of times X equals 0 or 1 and divide by 100,000.

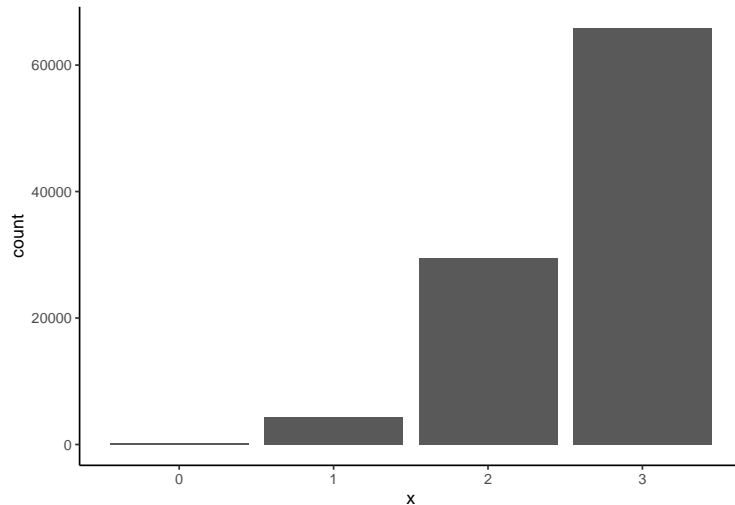
Additional fun

The `rbinom()` function in R simulates values from the binomial distribution, so we could run the above simulations (and save them into a data frame called `data`) with the following code:

```
data <- data.frame(x = rbinom(100000, 3, 0.87))
```

A bar plot of the 100,000 simulated x values shows us an approximate distribution of X

```
ggplot(data = data, aes(x = x)) +  
  geom_bar() +  
  theme_classic()
```



To estimate the probability the Lakers lose from our simulations, we have:

```
data %>%
  summarize(prop = sum(x == 0 | x == 1)/100000)
```

```
##      prop
## 1 0.04592
```

We can compare this to the theoretical probability from the pmf:

```
dbinom(1, 3, .87)
```

```
## [1] 0.046306
```

Here is the true pmf (navy outline) plotted on top of the simulated distribution (grey fill), and we can see they are nearly identical.

```
#convert simulated counts into proportions & compute pmf for each value of x
data_summary <- data %>%
  count(x, name = "count") %>%
  mutate(rel_freq = count/100000,
         pmf = dbinom(x, 3, .87))

ggplot(data_summary) +
  geom_col(aes(x = factor(x), y = rel_freq),
           alpha = 0.8) +
  geom_col(aes(x = factor(x), y = pmf),
           fill = NA, color = "navyblue") +
  labs(x = "x", y = "Density f(x)") +
  theme_classic()
```

