Problem Section 5 KEY

Binomial distribution

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:

- Identify whether the conditions for a binomial experiment are satisfied.

- Calculate binomial probabilities

- Calculate expected value

- Back up and support work with relevant explanations

Exercises

1. Consider the following random variables. Which ones are binomial? Explain your answer.

a. X is the number of "good" rolls when a die is rolled four times. We call a roll "good" if the number rolled is greater than the roll number. So the first roll is good if it is a 2 or higher. The fourth roll is good if it is a 5 or a 6.

   The four conditions of a binomial experiment are:

   i. The number of trials must be fixed
   ii. Each trial must result in a success or failure
   iii. The probability of a success must be the same for each trial
   iv. The trials must be independent.

   Let's check these conditions for the example. Here a trial is a roll of a fair die.

   The number of rolls is fixed, therefore condition (i) is satisfied. Each roll results in a "good" outcome or not, hence condition (ii) is satisfied. The probability of a "good" roll however varies from trial to trial. For example, the probability of a good roll on trial 1 is given by $\frac{5}{6}$ while it is $\frac{2}{6}$ for roll 4.

   Hence X is not a binomial random variable.

b. X is the number of tosses for the first head to appear when tossing a coin. You may assume repeated tosses are made in identical fashion.

   The number of trials is not fixed and hence X is not a binomial random variable.

2. A multiple choice test has 20 questions, each with four answers one of which is correct. Suppose a student randomly guesses the answer to each question. Let the random variable X denote the number of questions they guess correctly.

a. What is the distribution of X? State the name of the distribution and also the values of its parameters. Be sure to state (in context) any assumptions you need to make.

We can think of each question as a trial in a binomial experiment. Three conditions for this scenario to be modeled as a binomial experiment are satisfied: there are a fixed number of trials, each trial has two outcomes (correct answer/wrong answer) and the probability of a correct answer is the same - $\frac{1}{4}$ - for each trial. If we are willing to assume that the trials are independent – meaning that whether we guess correctly on one question has no impact on another - then all four conditions of a binomial experiment are met. This may not be a reasonable assumption if the student gets instant feedback. Knowing they answered something incorrectly may result in them getting flustered and botching the next question.

Under the assumptions of the binomial experiment, the random variable X will then have a $Binom(n = 20, \pi = \frac{1}{4})$ distribution.

b. Give an expression for the probability that they will correctly guess 12 questions. In a code chunk, write the R function for calculating this probability and save it to a variable and print. You can additionally report your answer (rounded to four digits) using inline code.

This is given by $P(X = 12) = \binom{20}{12}\left(\frac{1}{4}\right)^{12}\left(\frac{3}{4}\right)^{8}$ and can be calculated in R.

```
prob_x <- dbinom(x = 12,size = 20, prob = 1/4)    #P(X=12)
prob_x
```

```
## [1] 0.0007516875
```

The probability that $X = 12$ is given by 0.0008.[1]

c. How many questions should they *expect* to guess correctly? Write the formula and report the answer.[2]

Since $X \sim Binom(n, \pi)$ they should expect to get $n\pi = 5$ questions correct.

d. Give an expression for the probability that they will correctly guess more questions than expected. In a code chunk, write the R function for calculating this probability, save it to a variable and print it. You can additionally report its value using inline code.

Here we want to calculate

$$P\left(X > \frac{20}{4}\right) = P(X > 5)$$

```
prob_more <- 1 - pbinom(q = 4, size = 20, prob = 1/4 )    #P(X > q)
prob_more
```

```
## [1] 0.5851585
```

The probability is 0.5852.

3. The playoff series in the National Hockey league now are all best of seven games, but the early rounds used to be best of three. If a team thinks that on any given day they have probability 0.55 to beat their first round opponent, should they prefer the three game or the seven game format?

In the best of three format, the team needs to win 2 games to beat their first round opponent. This can happen in any of the following ways: (win, win), (win, lose, win), (lose, win, win). Since these outcomes are disjoint we can simply add their probabilities.

event	probability
win, win	0.55×0.55
win, lose, win	$0.55 \times 0.45 \times 0.55$
lose, win, win	$0.45 \times 0.55 \times 0.55$

Note we are assuming that the games are independent and this is why we can multiply the probabilities.

---

[1] set options(scipen = 999) in setup to discourage scientific notation

[2] Note: this is a simple enough calculation that you may not even need R to do the multiplication for you. In cases when you need a calculator, you should always do the work in a code chunk first. You can always save the value to a variable, and then report the value using inline code

The probability of beating the first round opponent in the 3 game format is therefore:

```
0.55^2 + 2*0.55*0.45*0.55
```

```
## [1] 0.57475
```

Another way to justify this is in the following manner. The team can either win in 2 games or 3 games. The probabilities of each event are calculated below.

$$P(\text{win in 2 games}) = 0.55^2$$

$$P(\text{win in 3 games}) = \binom{2}{1}0.55 \times 0.45 \times 0.55$$

The first calculation is fairly self explanatory. They need to win games 1 and 2.

However, in order to win in 3 games, they must lose one of the first two games and then win the third. If we define X as the number of games won in the first two played, then $X \sim Binom(n = 2, \pi = 0.55)$. Then we simply need to find $P(X = 1)$ which is given by $\binom{2}{1}0.55^2 0.45$.

Adding these together gives the probability of winning in a best of three format.

————————————————————————

Let's use this second approach to find the probability of winning in the best of seven format. Of course, now they need to win 4 games in order to beat their opponent. Therefore

$$P(\text{win}) = P(\text{win in 4}) + P(\text{win in 5}) + P(\text{win in 6}) + P(\text{win in 7})$$

Again, using independence of games, we can write:

$$P(\text{win in 4}) = 0.55^4 = 0.0915063$$

$$P(\text{win in 5}) = \underbrace{\binom{4}{3}0.55^3 \times 0.45}_{\text{win 3 of 4}} \times \underbrace{0.55}_{\text{win fifth game}} = 0.1647112$$

$$P(\text{win in 6}) = \underbrace{\binom{5}{3}0.55^3 \times 0.45^2}_{\text{win 3 of 5}} \times \underbrace{0.55}_{\text{win sixth game}} = 0.1853002,$$

$$P(\text{win in 7}) = \underbrace{\binom{6}{3}0.55^3 \times 0.45^3}_{\text{win 3 of 6}} \times \underbrace{0.55}_{\text{win seventh game}} = 0.1667701$$

Therefore the probability of winning in a seven game format is:

```
prob_win_in_4 <- 0.55^4;
prob_win_in_5 <- dbinom(x = 3, size = 4, prob = 0.55 )*0.55
prob_win_in_6 <- dbinom(x = 3, size = 5, prob = 0.55 )*0.55
prob_win_in_7 <- dbinom(x = 3, size = 6, prob = 0.55 )*0.55

prob_win_in_4 + prob_win_in_5 + prob_win_in_6 + prob_win_in_7
```

```
## [1] 0.6082878
```

They have a slightly better chance of winning with the seven game format.

4. A lottery randomly picks a 3-digit number from 000 to 999 (so there are 1,000 possible numbers). You pick your lucky number and bet \$1. If your number wins, you get \$700 (including your \$1 bet back). If

your number loses, you get nothing. What is your expected winning (average gain or loss) from playing once?

Let W be the random variable which denotes the winning from playing once. The distribution of W is shown below.

w	-\$1	\$700
$f(w)$	$\frac{999}{1000}$	$\frac{1}{1000}$

we could either win \$700 if my lucky number is the one chosen randomly. This occurs with probability $\frac{1}{1000}$ since there are 1000 numbers all of which are equally likely to be selected, of which your lucky number is one. On the other hand, we could lose our ante of \$1 which corresponds to winning of -\$1. This event occurs the balance of the time, in other words with probability $\frac{999}{1000}$.

Using the definition of expected values:

$$E\left[W\right] = \sum_x x \cdot f(x)$$
$$= -1 \times \frac{999}{1000} + 700 \times \frac{1}{1000}.$$

The expected winning is calculated below.

```
#calculate E(W)

E_w <- -1*999/1000 + 700*1/1000
E_w
```

```
## [1] -0.299
```

Therefore, we should expect to lose 30¢on the average.