STATS 210 **Semester 2** **Assignment 2**

Due: 11:30pm, Thursday 3rd September, 2020

1. A coach is training two Waka Ama teams, Team **A** and Team **B**, and needs to decide which team is quicker in a 500m race so that he can enter that team into the Waka Ama National Sprint Championships. He races the two teams against each other and observes how many races each team wins. Assume that the results of each race are independent of one another.

   (a) Let X be the number of races that Team **A** wins out of the fixed number of 15 races observed in total. Under the assumption that Team **A** and Team **B** are equally likely to win a race, write down the distribution of X, with parameters. (2)

   (b) Team **A** wins 11 out of 15 races. The coach wishes to check his hunch that Team **A** and **Team B** do not have the same probability of winning a race. Formulate the null and alternative hypotheses in terms of the distribution of X and its parameters. Specify the full distribution of X and use a two-sided alternative hypothesis. (2)

   (c) Sketch as a curve the shape of the probability function of X under the *null* hypothesis. Your sketch should have axes labelled x and $P(X = x)$. Mark on the sketch the upper and lower limits of x, and the approximate value of x where the curve peaks under the null hypothesis. Also mark the observed value $x = 11$ so that you can see the tail probabilities required for the p-value, and shade under the curve the area represented by the p-value. (4)

   (d) Write down the *R* command required to find the p-value for the hypothesis test, and run this command in *R* to find the p-value. Interpret the result in terms of the strength of evidence against the null hypothesis. Is there evidence that Team **A** is quicker than team **B**? (4)

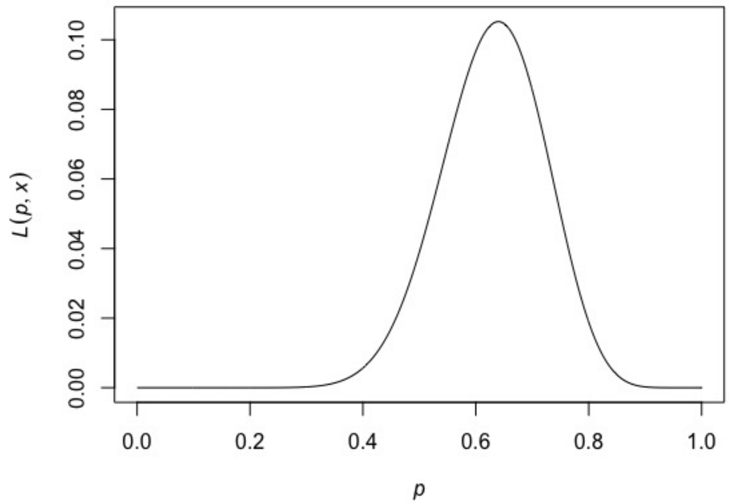
   (e) Imagine that the coach raced the teams against each other 150 times instead of 15 times, and that Team **A** won 110 times instead of 11 times. Team **A** has won the same proportion of races in both cases. If the coach conducted the same test as above, what can you say about the p-value from this new hypothesis test? Does it give **more**, **less**, or **the same amount** of evidence against H_0 as the p-value calculated in part (d)? Explain your answer.
   You do not need to perform any calculations. (2)

---

2. *Deja Brew* is a coffee bar on campus that sells hot chocolate. Catriona, who works at the coffee bar, thinks that customers may prefer to have a chocolate fish with their hot chocolate rather than the marshmallow she gives them now. To investigate her theory, she buys a box of 16 chocolate fish. She asks each customer who buys a hot chocolate whether they would like to switch to a chocolate fish rather than a marshmallow, until all of the 16 chocolate fish have been used. For each customer, the probability of them switching to a chocolate fish is p. Assume that no-one can have more than one chocolate fish.

   (a) Let X be the number of customers who do *not* switch to a chocolate fish before Catriona has used all 16 chocolate fish in the box. State the distribution of X, with parameters, leaving your answer in terms of p. (2)

(b) Catriona has to ask 25 customers before she has used all 16 chocolate fish. Write down the likelihood for this problem, $L(p; x)$, being careful with the observed value of x. Remember to state the range of values of p for which the likelihood is defined. Find the maximum likelihood estimate of p. You should make reference to the graph of the likelihood function, shown below. Show all your working in your answer. (5)



(c) What would be the maximum likelihood estimator for p if Catriona has X customers who do not switch to a chocolate fish before she has used all 16 chocolate fish? Find the answer by following your working for part (b). (4)

Catriona is convinced that the probability of each customer switching to a chocolate fish is not the same as the probability of them not switching to a chocolate fish, i.e., she suspects that $p \neq 0.5$. She decides to carry out a hypothesis test.

(d) Formulate Catriona's null and alternative hypotheses, H_0 and H_1, in terms of the distribution of X and its parameters. Use a two-sided alternative hypothesis. (2)

(e) The following table shows part of the probability function $f_X(x) = P(X = x)$ under the null hypothesis, with all probabilities rounded to 4 decimal places. Fill in the missing values p_1 and p_2. (2)

x	...	8	9	10	...	24	25	26	...
$f_X(x) = \mathbb{P}(X = x)$	...	0.0292	p_1	0.0487	...	0.0229	p_2	0.0144	...

(f) Catriona used the following command in R to calculate $\mathbb{P}(X \leq 9)$:

```
pnbinom(9,16,0.5) = 0.1148.
```

Use this information to calculate the p-value for Catriona's hypothesis test. Interpret the p-value. Do you think that Catriona is justified in her suspicions? (4)

3. A charity runs an annual raffle to raise money. Raffle tickets are sold for \$1 each, and the charity gives away a prize worth $\$x$. Matiu is the manager of the charity. He has noticed that the number of raffle tickets sold depends on the size of the prize, x. Over the last four years he has experimented with different values of prize money, x, and observed the following outcomes for the number of tickets sold:

Year, i	**Prize money, x_i**	**Number of tickets sold, Y_i**
1	50	329
2	100	437
3	150	590
4	200	630

Matiu sets up the following model:

Let Y_i be the number of tickets sold when the prize money is $\$x_i$, and assume that Y_1, Y_2, Y_3 and Y_4 are independent. Matiu assumes that:

$$Y_i \sim \text{Poisson}(\beta \log(x_i)),$$

where β is an unknown parameter to be estimated, satisfying $0 < \beta < \infty$ and 'log' denotes 'ln', the natural logarithm. Specifically:

$$Y_1 \sim \text{Poisson}(\beta \log(x_1)); \quad Y_2 \sim \text{Poisson}(\beta \log(x_2)); \quad Y_3 \sim \text{Poisson}(\beta \log(x_3)); \quad Y_4 \sim \text{Poisson}(\beta \log(x_4))$$

Matiu wishes to estimate β. He will then use the answer to choose the best value of $\$x$ to give away in prize money in the future.

(a) Give an expression for $E(Y_i)$ and $\text{Var}(Y_i)$. (2)

(b) Using Matiu's model, explain why we can write

$$\mathbb{P}(Y_1 = 329, Y_2 = 437, Y_3 = 590, Y_4 = 630) = \mathbb{P}(Y_1 = 329) \times \mathbb{P}(Y_2 = 437) \times \mathbb{P}(Y_3 = 590) \times \mathbb{P}(Y_4 = 630).$$

Make sure you give a full justification. (2)

(c) Let Matiu's observations be (x_i, y_i) for $i = 1, 2, 3, 4$. Show that the likelihood function for estimating β, according to his model, is:

$$L(\beta; \underline{y}, \underline{x}) = K \beta^{(y_1+y_2+y_3+y_4)} e^{-\beta \log(x_1 x_2 x_3 x_4)}$$

where K is a constant that you should specify, but not calculate. Remember to state the range of values of β for which the likelihood is defined. (5)

(d) Show that the maximum likelihood estimator for β is

$$\widehat{\beta} = \frac{Y_1 + Y_2 + Y_3 + Y_4}{\log(x_1 x_2 x_3 x_4)}$$

You may assume that the likelihood attains a maximum in the range $0 < \beta < \infty$. (4)

(e) Find the maximum likelihood estimate of β for Matiu's data shown in the table above. (2)

(f) If $E(\widehat{\beta}) = \beta$, we say that $\widehat{\beta}$ is an *unbiased* estimator of β. Using the expression for $\widehat{\beta}$ given in part (d), find an expression for $E(\widehat{\beta})$. Is $\widehat{\beta}$ an unbiased estimator? (5)

(g) Let Z_i be the amount of money raised in year i, which can be calculated as $Z_i = Y_i - x_i$, i.e., the number of tickets sold, minus the value of the prize money. Give an expression for $E(Z_i)$. (2)

Total: (55)
