

Homework 6

Spring 2023

Your Name Here

Instructions

- This homework is due in Gradescope on Wednesday May 17 by midnight PST.
- Please answer the following questions in the order in which they are posed.
- Don't forget to knit the document frequently to make sure there are no compilation errors.
- When you are done, download the PDF file as instructed in section and submit it in Gradescope.
- Rule on collaboration: You may guide someone who is stuck by giving high level advice. However, everyone is expected to write up their answers individually, which includes deciding what and how much to explain and entirely in your own words.

Exercises

1. (Starch or sugar) The results of scoring the offspring plants of as either starchy or sugary and as having either a green or a white base leaf appear below.

1) starchy-green	2) starchy-white	3) sugary-green	4) sugary-white
1997	906	904	32

According to a genetic model for these traits, the probability that a plant exhibits one of these trait combinations should be $\frac{1}{4}(2 + \theta_0)$ for the first combination, $\frac{1}{4}(1 - \theta_0)$ for the middle two combinations and $\frac{1}{4}\theta_0$ for the last where θ_0 is a probability related to linkage closeness.

- a. Determine the MLE of θ_0 . Be sure to clearly show
 - Log likelihood function $\ell(\theta)$ in terms of the counts x_1, x_2, x_3, x_4 (i.e., not just for this data)
 - First derivative equation and calculate the MLE $\hat{\theta}_0^{mle}$.
 - b. Give an expression for the asymptotic standard error of $\hat{\theta}_0^{mle}$ and calculate it.
 - c. Calculate an approximate 95% Wald confidence interval for θ_0 and report it in context.
 - d. Is there any reason to be concerned about the approximation in part c? Make a relevant plot and comment.
2. (Discrete X) Suppose X is a discrete random variable with PMF $f(x)$ indexed by a parameter θ as shown below.

θ_0	$x = 1$	$x = 2$	$x = 3$	$x = 4$
1	1/3	1/6	1/12	5/12
2	1/2	1/4	1/6	1/12

W

- a. Say we want to test $H_0 : \theta_0 = 1$ versus $H_1 : \theta_0 \neq 1$. Calculate the Likelihood Ratio Statistic W for each value of x and write it in the last row of the table. Briefly explain your work below. (Hint: The parameter θ_0 can only take two values, so you can find the MLE fairly easily)
- b. Write the sampling distribution of W below in tabular form assuming H_0 is true. (Hint: W is a discrete random variable)
- c. Suppose we observe $x = 3$. Calculate the P-value. What should we conclude at a 0.05 level of significance?
3. (Likelihood ratio) Suppose X_1, X_2, \dots, X_n are an *i.i.d.* sample from PDF

$$f(x) = (\theta_0 + 1)x^{\theta_0}, \quad 0 < x < 1,$$

where $\theta_0 > -1$.

- a. Determine the form of the Likelihood Ratio Test statistic W for testing

$$H_0 : \theta_0 = 0, \quad H_1 : \theta_0 \neq 0$$

assuming a sample x_1, x_2, \dots, x_n .

To make grading easier, please clearly indicate each of the following:

- Log likelihood function $\ell(\theta)$
 - Expression for the MLE $\hat{\theta}_0^{mle}$ (no need to verify that it is a maximum)
 - Expression for the likelihood ratio statistic simplified as much as possible
- b. The 30 values below are a random sample from this distribution for some true (unknown) value θ_0 . Calculate the likelihood ratio statistic for this data. (Write all your code for the remaining parts in the code chunk labeled `lik-ratio` but show the code in the Appendix. Report answers (rounded to 4 digits) using inline code.
 - c. Calculate the P-value for the Likelihood Ratio Test statistic using the approximate chi square distribution. Is there reason to be concerned about using the chi-squared distribution? Compare the log likelihood function with its quadratic approximation.
 - d. An alternative to using the chi squared distribution to estimate the p-value is to calculate an empirical p-value by generating a large number B of samples from the null hypothesis. Follow the steps below to calculate an empirical p-value.
 - Step 0: Set the random number seed to 414.
 - Step 1: Generate $x_1^*, x_2^*, \dots, x_{30}^* \stackrel{i.i.d.}{\sim} Unif(0, 1)$. (why?)
 - Step 2: For the generated sample, calculate the value of the MLE $\hat{\theta}_0^*$ and the likelihood ratio test statistic w^* .
 - Step 3: Repeat steps 1 and 2 a large number $B = 10000$ times. (Don't forget to cache the code chunk `lik-ratio`)
 - Step 4: Count the fraction of times that w^* from the generated samples exceeds the w we observed. Report the empirical P-value and also make a histogram of the values of w^* . (Don't forget those labels and title)