

Stat 343 Final

What is your name? _____

Problem I: Likelihood Ratio Test Concepts

In about 2-3 sentences, explain why in a hypothesis test involving composite hypotheses, the p-value is calculated as the probability of obtaining a likelihood ratio statistic at least as small as the value of the likelihood ratio statistic that was observed in our data set given that the null hypothesis is true. (Why at least as small rather than at least as large?)

Problem II: Confidence Level Concepts

(a) Define the *coverage rate* of a frequentist confidence interval in a sentence or two.

(b) Define the *nominal coverage rate* of a frequentist confidence interval in a sentence or two. It's ok if your definition for this part is informal as long as you get the idea across.

(c) Explain why the coverage rate of an interval might not be equal to the nominal coverage rate of the interval in a sentence or two.

Problem III: Bootstrap Concepts

(a) Describe in detail how you could use the bootstrap percentiles interval method to generate a 95% confidence interval for the population value of the coefficient of determination (R^2) for a multiple regression model. You may assume that you have available a function that will calculate the value of R^2 for a given data set (for example, in R this could be done with `mod <- lm(...)` and then `r_sq <- rsquared(mod)`). You do not need to write any actual code, but your description of the procedure should be detailed enough that someone who knew how to program but didn't know statistics could code it up.

(b) What assumptions are needed for this interval to have appropriate coverage?

Problem IV: Likelihood Ratio Tests in Practice

Suppose that $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Poisson}(\lambda)$. The pmf of X_i is $f_{X_i|\lambda}(x_i|\lambda) = \frac{\lambda^{x_i} \exp(-\lambda)}{x_i!}$

(a) Find the likelihood ratio statistic for testing $H_0 : \lambda = \lambda_0$ vs. $H_A : \lambda = \lambda_A$, where $\lambda_0 < \lambda_A$.

(b) Show how the p-value for the likelihood ratio test can be calculated based on a comparison of $\sum_{i=1}^n X_i$ to $\sum_{i=1}^n x_i$. Use the case that $\lambda_0 < \lambda_A$, and note that if $a < 1$, $\log(a) < 0$.

(c) A fact you may have seen in probability is that if $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Poisson}(\lambda)$ and we define Y to be the sum of the X_i ($Y = \sum_{i=1}^n X_i$), then $Y \sim \text{Poisson}(n\lambda)$. Suppose that in the hypothesis statements from part (a), $\lambda_0 = 10$ and $\lambda_A = 2$. Show how the p-value for the test could be calculated either using an R command or as a suitable sum of the pmf of a Poisson distribution (you don't need to run the R code or do the summation, just show how to set it up). You only need to show the R code or the summation, but in either case the set up should involve only known numbers like 10, 2, n , and x_1, \dots, x_n . Note that a Poisson distribution is discrete.

Problem V: Confidence Intervals in Practice

Suppose that the random variables Y_1, \dots, Y_n satisfy $Y_i = \beta x_i + \varepsilon_i$, where x_1, \dots, x_n are fixed **constants** and $\varepsilon_i \stackrel{\text{i.i.d.}}{\sim} \text{Normal}(0, \sigma^2)$, where σ^2 is known. With this set up, the pdf of Y_i is

$$f_{Y_i|\beta, \sigma^2, x_i}(y_i|\beta, \sigma^2, x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-1}{2\sigma^2}(y_i - \beta x_i)^2\right]$$

Note that the ε_i are iid, but the Y_i are not. Each Y_i has a different mean that depends on the corresponding x_i . This is a simple linear regression model with only a slope, no intercept.

(a) Find the maximum likelihood estimator of β , $\hat{\beta}^{MLE}$. Be sure to check the second order conditions to ensure you have a maximum.

(b) Find the bias and variance of the maximum likelihood estimator $\hat{\beta}^{MLE}$.

(c) Find the Fisher information about β from the full data set. Since Y_1, \dots, Y_n are not identically distributed, it will not work to find the information from one observation and multiply it by n ; instead, you should base this on the full log-likelihood.

(d) Is $\hat{\beta}^{MLE}$ an efficient estimator of β ? Explain briefly in a sentence or two based on your answers to the previous parts.

(e) Suppose we observe y_1, \dots, y_n for a large sample size n . Show how you could find a $(1 - \alpha)100\%$ credible interval for β based on a large-sample normal approximation to the posterior distribution of β . You don't have any numbers, so you won't be able to actually calculate the interval. But your answer should be in the form of a formula such that if I had observed x_i 's, y_i 's, and knew my sample size n and variance σ^2 , I could easily calculate the interval bounds using R.