Stat 343 Final Review Problems

Final Structure and Coverage

The final will have 2 kinds of questions:

1) conceptual questions to make sure you understand what's going on.
2) some problems where you're asked to do some math.

There will not be any R coding questions.

Conceptual Topics

- All the things from the midterm
- Things about confidence intervals, likelihood ratios, and p-values

Here are some examples:

Problem 1

If n = 1,000,000,000, is there an unbiased estimator with lower variance than the MLE? You could answer "yes", "no", or "maybe". Justify your answer.

Problem 2.

A 90% confidence interval for the average number of children per household based on a simple random sample is found to be (0.7, 2.1). Because the average number of children per household, μ, is some fixed number in the population (at least, at a particular moment in time when we conduct the study), it doesn't make any sense to claim that $P(0.7 \leq \mu \leq 2.1) = 0.90$. What do we mean, then, by saying that this is a "90% confidence interval"? Can we ever make probability statements about confidence intervals?

Problem 3. TRUE or FALSE?

(a) TRUE or FALSE: A 95% confidence interval contains 95% of the population.

(b) The central limit theorem states that as the sample size becomes large, the distribution of the sample mean $\bar{X}$ approaches a normal distribution. In class, we developed an approach to deriving a confidence interval based on this result. TRUE or FALSE: For this interval estimation procedure, the actual coverage probability may not be exactly equal to the nominal coverage probability.

Example Worked Problems

The midterm will have problems roughly similar in content to the examples below.

Problem 1

Suppose that $Y_1, Y_2, \ldots, Y_n$ are i.i.d. samples from a distribution with pdf given by $f_Y(y|\theta) = \frac{1+\theta y}{2}$ for $-1 < y < 1$ and $-1 < \theta < 1$. An investigator is interested in testing $H_0 : \theta = 0$ vs. $H_A : \theta = 0.5$.

(a) Plot a picture of the pdfs of Y_1 under the null and alternative hypotheses. (One plot, two lines).

(b) Consider a sample of size n. Derive a general structure of the rejection region of the most powerful test of size $\alpha = 0.05$ for testing $H_0 : \theta = 0$ vs. $H_A : \theta = 0.5$. By "general structure", I mean that you should find the test statistic and indicate in general what the rejection region of the test looks like, but you do not need to derive the exact critical value for the test.

(c) Suppose now that $n = 1$. Calculate the p-value of the test if you observe $y = 0.5$. You should be able to get to a number.

(d) Still working with $n = 1$, find the rejection region of the test.

Problem 2

We have previously shown that if $X \sim \text{Binomial}(n, \theta)$, then the maximum likelihood estimator $\hat{\theta} = X/n$ has expected value θ and variance $\theta(1-\theta)/n$. Regarding X as a sum of results of iid Bernoulli trials, find an approximation to the distribution of $\hat{\theta}$ if n is large. Use your approximation to find an approximate 95% confidence interval for θ. This is the confidence interval for θ that is typically taught in introductory statistics courses.

Problem 3

Consider a situation in which we've observed $x_1 = 1$, $x_2 = 2$, and $x_3 = 4$. Calculate the bootstrap distribution of the median. Your answer should be in the form of a table listing possible values of the median and their probabilities. (Hint: there are 27 possibilities, and they are all equally likely). Use the distribution to find a 90% confidence interval for the median based on the bootstrap percentile method. Is it appropriate to use this approach? Why or why not?