AMS-205B WINTER 2016
TAKE-HOME FINAL EXAM

Due electronically on Tuesday, March 15 at 8:00 am. Email a single .pdf with your answers to abel@soe.ucsc.edu. We will meet to discuss the exam in our regular classroom that same day between 9:30 and 10:15 am.

(1) Consider data consisting of pairs (Y_i, X_i) for $i = 1, \dots, n$ arising from a non-linear regression model

$$Y_i = \alpha + \beta \cos(\omega X_i + \delta) + \epsilon_i, \qquad \epsilon_i \sim \mathsf{N}(0, \sigma^2),$$

where α, β, ω, δ and σ^2 are all unknown parameters.

(a) Show that the expression for the maximum likelihood estimator for $(\alpha, \beta, \omega, \delta, \sigma^2)$ is not available in closed form, and write an algorithm (in your favorite language) to find the value of the estimate for a given sample. Apply your algorithm to the dataset contained in the file `orbits.txt` and report your point estimates.

(b) Construct approximate confidence intervals for δ and ω by inverting a Wald-like test based on a normal approximation for the distribution of their maximum likelihood estimators.

(c) Construct an algorithm to construct bootstrapped 95% confidence intervals for δ and ω and apply it to the data in `orbits.txt`. Report your estimates and compare them with the approximation you derived before.

(d) If you were asked to test $H_0 : \delta = 0$ versus $H_a : \delta \neq 0$, what would be your conclusion in this case?

(2) Consider two random samples $X_1, \dots, X_n$ and $Y_1, \dots, Y_n$, where each observation corresponds to the lifetime (in years) of a certain electronic component produced at two different manufacturing plants. For the purpose of this problem we will assume that the lifetimes are exponentially distributed with means λ_1 and λ_2, respectively, and that we are interested in testing whether the lifetime of the components produced in the first plant is longer than that corresponding to components from the second plant, i.e., $H_0 : \lambda_1 \leq \lambda_2$ versus $H_a : \lambda_1 > \lambda_2$.

(a) Derive the likelihood ratio test along with formulas for the exact and the asymptotic p-values.

(b) Derive a Wald test based on the statistic $T = \bar{X} - \bar{Y}$ and a formula for the corresponding p-value.

(c) Describe how you would implement a permutation test for this problem and how to compute a p-value for it.

Now, consider the data contained in the file `lifetime.txt`, where the first column corresponds to the X_is and the second to the Y_is. For this data, compute the four p-values you derived before for this dataset. What conclusions do you draw in this case? Do they all agree? Discuss.

(3) Consider a random sample $X_1, \ldots, X_n \sim \mathsf{N}(\theta, \sigma^2)$ where both θ and σ^2 are unknown. We are interested in testing $H_0 : \theta = \theta_0$ versus $H_a : \theta > \theta_0$ using a test that rejects H_0 if the statistic

$$T = \frac{\sqrt{n}(\bar{X} - \theta_0)}{S}$$

is small.

(a) Derive an exact expression for the power function.

(b) Derive an approximate expression for the power function assuming that T follows a standard normal distribution under the null hypotheses.

(c) Graph both of these functions for $n = 10$, $n = 100$ and $n = 1000$ assuming a level $\alpha = 0.1$ for the test. Discuss.

(d) Is this the uniformly most powerful test for this problem?

Assume now that you plan to carry out an experiment that will allow you to collect a random sample that will be used to test these hypotheses. What is the smallest sample size you should collect if you are want to run a level 0.05 test and be able to detect an increase in the mean (with respect to θ_0) equivalent to two times the true standard deviation of the data with probability 0.8?