STAT 5113: Statistical Inference

Homework 6 – Due April 13, 2018

1. Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{E}xp(1/\theta)$. Note that we are taking the parameter θ to be the expected value of the distribution. The sufficient statistic for this model is the sample mean $\bar{X}$, which has distribution $\mathcal{G}am(n, n/\theta)$. Consider testing the hypothesis

$$\mathrm{H}_0: \ \theta \leq \theta_0 \qquad \text{versus} \qquad \mathrm{H}_1: \ \theta > \theta_0.$$

   (a) Based on the distribution of the sufficient statistic, determine the Likelihood Ratio Test (LRT) statistic, $\lambda(\bar{X})$.

   (b) Show that λ is a decreasing function. Use this fact to express the general form of the rejection region of the LRT as explicitly as possible in terms of $\bar{x}$.

   (c) Determine the power function $\pi(\theta)$ of the test.

   (d) Consider the following data set, providing survival times in weeks of patients who were diagnosed with leukemia:

   $$65,\ 156,\ 100,\ 134,\ 16,\ 108,\ 121,\ 4,\ 39,\ 143,\ 56,\ 26,\ 22,\ 1,\ 1,\ 5,\ 65.$$

   Assuming independent, exponentially distributed survival times, with expected value θ, use the framework above to test the hypothesis the the average survival time is more than one year. That is, consider $\theta_0 = 52$. Select the threshold of the rejection region so that $\pi(\theta_0) = \alpha$.

   (e) Plot the graph of the power function. What can you deduce about the size of the test?

   (f) Carry out the test with the data provided and draw your conclusions.

   (g) What is the *p-value* of the test?

   (h) [Extra Credit] Show that for any value of the threshold and any sample size, the power function of the test is increasing in θ.

2. Consider two independent samples $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ from Normal populations with means μ_X and μ_Y and variances σ_X^2 and σ_Y^2.

   (a) Determine the LRT statistic to test

$$\mathrm{H}_0: \ \sigma_X^2 = \sigma_Y^2 \qquad \text{versus} \qquad \mathrm{H}_1: \ \sigma_X^2 \neq \sigma_Y^2.$$

   (b) In a packing plant, two machines pack cartons with jars. The times it takes each machine to pack 20 cartons are recorded. Assume that the model specified above holds for these data. A summary of the data, in seconds, is give below:

   $$\bar{x} = 42.14, \ \ \hat{\sigma}_X^2 = 0.4664, \ \ \bar{y} = 43.23, \ \ \hat{\sigma}_Y^2 = 0.5625.$$

   Using the asymptotic distribution of the LRT statistics, calculate an approximate *p-value* to test the equality of the variances.y