Homework 1

PM522b Introduction to the Theory of Statistics Part 2

Due: January 23, 2018

Order Statistics

1. Let $X_1, X_2, ..., X_n \overset{iid}{\sim}$ U[0, θ].

   a. Find the density function of $X_{(j)}$ where j is an integer $1 \leq j \leq n$.

   b. Given your result in a, find $\mathrm{E}(X_{(j)})$ and $\mathrm{Var}(X_{(j)})$.

   c. Find the mean difference between two successive order statistics, namely $\mathrm{E}(X_{(j)}) - X_{(j-1)})$ and interpret this result.

   d. Find the joint density function of $X_{(j)}$ and $X_{(k)}$ where j and k are integers $1 \leq j < k \leq n$.

   e. Given your result in d, find $\mathrm{Cov}(X_{(j)}, X_{(k)})$.

   f. Find the variance between the difference of the two order statistics, $\mathrm{Var}(X_{(k)} - X_{(j)})$.

2. Suppose a machine uses 10 batteries that have U[1/2, 1] distribution (in years), and it shuts off when 1/2 of the batteries are dead.

   a. What is the expected time when the 5th battery will die?

   b. What is the probability that the machine will shut off before 3/4 of a year?

   c. The machine's efficiency is lost when there are 3 dead batteries. It costs $1 per day to run the machine at this point. How much money will the company spend before the machine shuts off? (Hint: This is a joint probability problem!)

3. For $X_1, X_2, ..., X_n \overset{iid}{\sim}$ Unif[0, 1], write a function in R to draw random samples of size 15, and take the maximum of each sample.

   a. Graphically examine the sampling distribution of the maximum and describe your results.

   b. Derive the pdf of the sampling distribution of the maximum.

4. Using kids.RData posted on Blackboard (this is the CHS data used in PM511a), construct Q-Q plots in R comparing the wt variable (weight in lbs) to the normal, log-normal, gamma, and Weibull distributions. Do not use a package to create these plots. Label all axes accordingly. To what distribution do the data follow most closely?

5. The median absolute deviation (MAD) of a random variable X is defined as the median of the absolute deviations between the observations and their median, i.e. MAD=median($|x_i-$ median$(x_i)|$). The interquartile range (IQR) of a random variable X is the difference between the upper quartile and the lower quartile.

   a. Show that for a symmetric continuous random variable with a strictly positive pdf that MAD = IQR/2

   b. Calculate the MAD for a standard normal distribution

   c. Calculate the MAD for a standard Cauchy distribution

   d. Using R, generate a random sample from a uniform distribution and calculate the MAD.

6. Complete the proof that $\bar{X}$ and S^2 are independent by looking at the n-1 deviations $(X_1 - \bar{X}, X_2 - \bar{X}, ..., X_n - \bar{X})$.