STARFISH SCHOOL

A Virtual Bootcamp for Astronomy Graduate Students

WEEK 3 EXERCISES

Version 1.0

Mini-Exercise 1

1. Plot the PDF for the χ^2 distribution, for different values of the degrees of freedom (N) of that distribution.
2. Compare this to the normal distribution.
3. What do you notice about the two distributions?
4. Now compare the normal distribution to the log normal distribution for a range of values of the mean and variance. How do the mean and variance of the log normal distribution map onto the mean, variance of the normal distribution?

Mini-Exercise 2

1. Use the accept-reject approach to transform numbers generated from a uniform distribution into those following the distribution P(x) = (1/(e-1))exp(x) for 0<x<1 and 0 elsewhere
    a. Draw two random samples x1, x2 from the U(0,1) distribution
    b. If x1 < P(x2), keep x1
    c. If not, draw another two random samples from the distribution
    d. Continue until you have 100 samples
    e. Histogram the samples and over plot the PDF
2. Use CDF sampling to do the same thing above.
    a. To do this, compute the CDF F(X) by integrating the PDF P(x) from -∞ to X
    b. Then find the inverse F-1(X)of the CDF.
    [HINT: Remember an inverse function F-1(x) is such that F(F-1(x)) = x]
    c. Draw a random samples x1 from the U(0,1) distribution
    d. Then the variable y = F-1(x1) will have the probability distribution you seek
    e. Continue until you have 100 samples
    f. Histogram the samples and over plot the PDF

Main Exercise

1. Download the data set xvalues.csv from the website
2. Generate a histogram for these values using bin widths of 2, from -8 to 4. Before going to part b), what do you notice about this distribution? Would you hypothesize what distribution the data came from?
3. Generate a new histogram for these values using bin widths of 2, starting instead from -7.
4. Make a boxplot of these data and find the summary statistics
5. Make a kernel density estimate plot of the distribution. How does this compare to the other options?
6. Based on your figures, comment on the pros and cons of each estimate of the distribution (histogram, boxplot, KDE)

7. Standardize the data from question 1, and make a new histogram and boxplot. Compare these to your histogram and boxplot in question 1.
8. What are the mean and standard deviation of the standardized data?
9. Check the 68-95-99 rule using the standardized data. Is the empirical rule applicable here? Why or why not?

Stretch Goals:

1. Make an empirical CDFs of the data and compare to the CDF of a normal.
2. Make a Q-Q plot (look up what this is)!