ST 501 R project

For this project you will be using R to simulate random data, approximate quantities, and create graphs. Hopefully this will help you to understand the convergence concepts of chapter 5.

You will turn in an R file to moodle that adheres to the R file submission guidelines. You should also turn in a PDF corresponding to the derivations.

## Part I - Convergence in Probability

1. Consider the 'double exponential' or Laplace distribution. A RV $Y \sim Laplace(\mu, b)$ has PDF given by
$$f_Y(y) = \frac{1}{2b} e^{-\left(\frac{|y-\mu|}{b}\right)}$$
for $-\infty < y < \infty$, $-\infty < \mu < \infty$, and $b > 0$. If you are interested, some applications of the distribution are given here.

   We will consider having a random sample of Laplace RVs with $\mu = 0$ and $b = 5$. We'll look at the limiting behavior of $L = \frac{1}{n}\sum_{i=1}^{n} Y_i^2$ using simulation.

   (a) Give a derivation of what $L$ converges to in probability. You should show any moment calculations and state the theorem(s) you use.

   (b) Explain what $K = \sqrt{L}$ converges to and why.

   (c) Derive the CDF of $Y$. Note you'll have two cases and you should show your work (the answer can be found on wikipedia).

   (d) We want to see what happens to these two quantities as our sample size ($n$) increases. We will generate data as follows (different from the notes but you can use that to get started!):

      i. We'll generate $N = 50$ values of $L$ and $K$ for every value of $n$ from 1 to 250. That is, consider $n = 1$. We want to generate $N = 50$ datasets with $n = 1$. For each data set we want to find $L$ and $K$. Now, for $n = 2$ we want to generate 50 datasets, find $L$ and $K$ for each one. All the way until we get to $n = 250$.

      ii. You should use proposition D to generate the values (that is, use only random uniform values directly from R).

   (e) Next we will plot the values of $L$ and $K$ that we found to see what the values are converging to! Create two separate plots, one for each statistic.

i. Plot the values with $n$ on the xaxis and the observed $L$ (or $K$) values on the yaxis.

ii. Add a solid horizontal line for the value we know theoretically the statistics are converging toward.

iii. Add dashed lines corresponding to $\epsilon = 20$ away in either direction for $L$ and $\epsilon = 3$ away in either direction for $K$.

iv. Make sure all plots have a useful title and labeling.

v. In a comment, explain how graphically we are seeing the theoretical result of convergence in probability.

## Part II - Convergence in Distribution

2. This time we'll look at the limiting distribution of our statistics.

   (a) Take the $N$ values of $L$ you have for each of $n = 10, 50, 100, 250$ and create a histogram to inspect the shape of the sampling distribution of $L$ for those values of $n$.

   (b) Make sure the graph(s) have appropriate titles and labels.

   (c) In a comment, does it appear as though $L$ or $K$ is converging to a normal distribution?

3. Theoretically, since $L$ is an average of *iid* RVs ($Y_i^2$ are each RVs) with finite variance we know that, properly standardized, $L$ should have a standard normal limiting distribution by the CLT.
   Derive the appropriate standardization that will converge to a standard normal distribution for $\mu = 0$ and an arbitrary $b$. Show your work. Note: the kurtosis of the Laplace distribution is 6 and we have $\mu = 0$. Find the formula for kurtosis (book or wikipedia) and the calculation of the fourth moment won't be too bad!

4. Redo your above 4 plots using the standardization.

5. Now let's see if convergence is occuring for larger values of $n$. Generate data using the same method as above except do so for $n = 1000$ and $n = 10,000$. Use $N = 10,000$ data sets as well.

6. Create similar plots for these two $n$ values. In a comment discuss how the CLT is manifesting for this problem. Does $n > 30$ work?