Generating random variables, testing the normality by Q-Q plot, and confidence interval for mean value

Background

In class, we have introduced the uniform distribution and normal distribution, including the probability density function and their properties. Another distribution we have seen in homework and exam is called exponential distribution, please see section 4.7. In this project, you will be asked to generate random variables from these distributions using the functions provided by Microsoft Excel. Please see the instructions below.

In a lot of statistics problems, the normal assumption is very critical. The recent example we saw is that we require the data from normal distribution in order to use t-distribution for the mean value. Thus, given a dataset, it is always desirable to test if the data is from normal distribution or not. In statistics, there are many methods to perform the task, but probably the easiest and most intuitive method is by probability plot, or quantile-quantile plot (Q-Q plot). In this project, we will also generate a Q-Q plot from the data.

Instruction

1. Self-study Section 4.7 and Section 4.8 (Uniform distribution), understand the pdf and cdf of uniform distribution and exponential distribution.
2. Self-study Section 4.10 in the textbook, understanding the idea of QQ plot and how to judge normality from the plot.
3. The function RAND() in Excel generates a uniform random variable between 0 and 1. Please refer to http://office.microsoft.com/en-us/excel-help/rand-HP005209229.aspx, to see the usage of the function, and understand how to generate a general uniform random variable.
4. Using the RAND() function to generate 50 uniform random variables from (5,8), and generate a QQ plot for the data. *Hint, use 3*RAND()+5* Refer to http://facweb.cs.depaul.edu/cmiller/it223/normQuant.html for the instruction for the QQ plot. On the webpage, there is also a link for a complete example of QQ plot. In your worksheet, you can save your generated random variables in the first column, but if you make any operation, the values in the first column will change. Thus, a better way is to copy the values in the first column into the second column, and the operation is “select the values in column A > Ctrl+C or right click and select copy > highlight a cell in Column B and right click, select Paste Special > Paste Values” by this way, when you do any operation, the values in column B will be the same and you can work on the data in column B directly.
5. What do you get from step 4? Based on the resulting QQ plot, are the data from normal distribution?

6. Generate 50 exponential random variables from Exp(1) distribution using the command \(-\log(RAND())\).

7. Similar to step 4, generate a QQ plot for the normal random variables you obtained in step 6. Based on the resulting QQ plot, are the data from normal distribution?

8. Generate 50 normal random variable from N(5, 2). *Hint: the Excel function NORMINV(RAND(), mu, sigma) generates a random variable from normal distribution with mean mu and standard deviation sigma.*

9. Similar to step 4, generate a QQ plot for the normal random variables you obtained in step 8. Based on the resulting QQ plot, are the data from normal distribution?

10. For the normal distribution, please generate confidence intervals at levels 90%, 95%, and 99%. Do they include the true mean value? Calculate the length of the intervals, which one is the longest and which is the shortest? What is the relation between confidence level and the length? *Hint: you can use the excel function CONFIDENCE, please refer to the website for more information http://www.excelfunctions.net/Excel-Confidence-Function.html*

**Maximum Likelihood Estimation**

In class, we briefly introduced the maximum likelihood method for parameter estimation. Please self-study section 4.9 again. Now, suppose that there is a random sample \(X_1, X_2, \cdots, X_n\) from a Poisson distribution with probability mass function

\[
f(x \mid \lambda) = e^{-\lambda} \frac{\lambda^x}{x!} \quad \text{for} \quad x = 0, 1, 2, \cdots,
\]

where \(\lambda > 0\) is the unknown parameter. Please use the maximum likelihood method estimate the parameter \(\lambda\). Please write your solution in details!

**Requirement for Report**

Please organize the ideas (or methods), the results, and your answers to the questions in a project report file. Try to organize everything as well as you can! Please also print out the excel files and the graphs, attached to the report.

**Due date:**

Nov. 23rd, Wednesday.