Statistics - Homework 5 (Due Dec. 13, 2017)

1. Let *Z* be a standard normal random variable. It can be shown that $X = Z^2$ has a chi-square distribution with degree of freedom 1.
   (1) Prove the above result by considering the probability density function of *X*.
   (2) Prove the above result by considering the moment generating function of *X*.
2. A scientist is given an assignment of determining the average *soil moisture content* (SMC) in a farm. Assuming soil moisture contents at different locations in the farm are identically distributed with expected value μ and standard deviation σ. The scientist plans to take SMC measurements at *n* locations and all measurements ($x_i,\ i = 1, \cdots, n$) can be considered as independent of others. The average SMC in the farm (μ) is then estimated by $\bar{x}_n$, the average of all SMC measurements.
   (1) If SMC are measured at 25 locations, what is the probability that $\bar{x}_n$ will fall within the range of $(\mu - 0.1\sigma,\ \mu + 0.1\sigma)$?
   (2) At least how many measurements need to be taken in order for the probability that $\bar{x}_n$ falls in the range of $(\mu - 0.1\sigma,\ \mu + 0.1\sigma)$ to be higher than 0.95?
3. Assuming the soil moisture content in Problem 3 can be characterized by a gamma density with mean and standard deviation being 15% and 9%, respectively. The *n* SMC measurements taken by the scientist can be considered as a random sample of size *n*. Conduct the following simulations using R:
   (1) Simulate 10,000 random samples, each of sample size *n*, for *n* = 100, 200, 300, 400, 500, 1000, and 2000, respectively.
   (2) For any of the above sample sizes, calculate the average soil moisture content ($\bar{x}_n$) of each individual random sample and determine the percentage (*p*) of random samples satisfying $\mu - 0.1\sigma \le \bar{x}_n \le \mu + 0.1\sigma$.
   (3) *Based on the Weak Law of Large Numbers*, calculate the probabilities that $\bar{x}_n$ falls within the range of $(\mu - 0.1\sigma,\ \mu + 0.1\sigma)$ for *n* = 100, 200, 300, 400, 500, 1000, and 2000, respectively.
   (4) Show scatter plots of (*n*, *p*) based on the results of (2) and (3), respectively. [Show values of *n* and *p* in the x-axis and y-axis, respectively.]
   (5) Derive *p versus n* relationships for the two scatter plots in (4). [**Note**: You can derive the relationships *empirically or theoretically*.]
4. Let *Z* be a standard normal random variable.
   (1) Simulate 10,000 random samples, each of sample size 30, of *Z*. For each sample, calculate the sample mean ($\bar{x}$) and the sample variance (s^2).
   (2) Use R to calculate the correlation coefficient of the sample mean and sample variance and show the scatter plot of $(\bar{x},\ s^2)$.
5. Let X be a normal random variable with expected value 200 and standard deviation 120.
   (1) Generate 1000 random samples, each of sample size n=50, of X. Find the order statistic Y_n of each random sample. [The 1000 values of Y_n is referred to as a random sample of size 1000 of Y_n.]
   (2) Plot the histogram of Y_n and calculate the sample mean, standard deviation, and coefficient of skewness of Y_n.
   (3) Assuming the order statistic Y_n can be characterized by a Gumbel distribution Y^*. Use the sample mean and sample standard deviation obtained in (2) to generate a random sample of size 1000 of the Gumbel distribution.
   (4) Plot the histogram of the Gumbel random sample and compare with the histogram in (2).
   (5) On the same plot, show the empirical cumulative distribution functions of the random samples of Y_n and Y^*.