ENGRD 2700: Basic Engineering Probability and Statistics
Fall 2019
Homework 6

Due **Friday Nov 15** at 11:59 pm. Submit to Gradescope by clicking the name of the assignment. See `https://people.orie.cornell.edu/yudong.chen/engrd2700_2019fa.html#homework` for detailed submission instructions.

The same stipulations from Homework 1 (e.g., independent work, computer code, etc.) still apply.

1. Compute the covariance and correlation of X and Y, when these continuous random variables have joint density function given by

$$f_{X,Y}(x,y) = \begin{cases} 1/8 & x \in [0,1), y \in [0,1) \\ 1/8 & x \in [1,2), y \in [1,2) \\ 3/8 & x \in [0,1), y \in [1,2), \text{ or } x \in [1,2), y \in [0,1) \\ 0 & \text{otherwise.} \end{cases}$$

2. Suppose there are two stocks X and Y. The annual returns of the stocks are normally distributed, with the same mean $\mu_1 = \mu_2 = 12$ and variances $\sigma_1^2 = 9$ and $\sigma_2^2 = 16$, respectively.
   (a) Assume that X and Y are independent. Suppose you hold one share of X and one share of Y. Compute the probability that your total annual return is greater than 25.
   (b) Now assume that X and Y are negatively correlated with correlation $\rho = -0.6$. It turns out that the sum of two (possibly correlated) normal random variables is still normallly distributed. Compute the probability in (a).
   (c) Still assume that $\rho = -0.6$. Suppose you can only purchase one share of X and Y in total. That is, your returns from the two stocks are wX and $(1-w)Y$, where $0 \le w \le 1$. What choice of w minimizes the variance of your total return?
   (d) Now assume that X and Y are perfectly negatively correlated (i.e., $\rho = -1$), what value of w minimizes the variance of your total return?
   (e) Would you prefer to invest under the conditions of part (c) or part (d)? Why?

3. When we have sample data we can compute the *sample* covariance and the *sample* correlation. In particular, suppose we have data pairs $((X_i, Y_i), i = 1, 2, \ldots, n)$. The sample covariance is defined as

$$q_{X,Y} = \frac{1}{n-1} \sum_{i=1}^{n} [(X_i - \bar{X})(Y_i - \bar{Y})],$$

   and the sample correlation is defined as

$$r = \frac{q_{X,Y}}{s_X s_Y},$$

   where s_X^2 and s_Y^2 are the sample variances of the X_i's and Y_i's, respectively.
   (a) Consider the dataset `DataForSunglasses.csv`. Generate a scatter plot for ice cream sales vs temperature, and compute their sample correlation. Do the same thing for sunglasses sales vs temperature, and for sunglasses sales vs ice cream sales. (That is, you need to generate three plots and compute three correlations.)
   (b) You should see a positive correlation between sales of sunglasses and ice cream sales. Does this mean that ice cream makes people sensitive to sunlight? Explain in one sentence.

4. Suppose you flip a fair coin 100 times
   (a) Use the Central Limit Theorem to approximate the probability that heads appears at most 46 times.

(b) Write down an expression for the exact probability in part (a), and compute it. (In R, you can use `pbinom`.)

(c) Our approximation from part (a) is not very accurate. This is primarily because we are using a continuous random variable (the normal) to approximate a discrete random variable (the binomial). To improve our approximation, we can use the fact that

$$P(S_{100} \leq 46) = P(S_{100} \leq 46 + c)$$

for any constant $0 \leq c < 1$, since S_{100} can only take integer values. It turns out that $c = 0.5$ works very well in practice, and so to approximate the probability of seeing at most 46 heads, we can apply the Central Limit Theorem to $P(S_{100} \leq 46.5)$ instead. Do this, and compare the approximation you obtain here to the one in part (a). The procedure is referred to as a *continuity correction.*

5. An insurance company looks at the records for millions of homeowners and conclude the probability of fire in a year is 0.01 for each house and the loss should a fire occur is \$10,000. Thus the expected loss from fire for each house is \$100. Assume the fires are independent. The company plans to sell fire insurance for \$120 (which is the expected loss plus \$20). If a house owner purchases the insurance and his/her house catches fire, the company will cover for the loss.

   (a) If the company sells the insurance policy to 10 houses, what is the expected total profit of the company?

   (b) Compute the probability of bankruptcy for the company, that is, when total profit is negative.

   (c) Use the Central Limit Theorem to approximately compute the probability of bankruptcy if the company sells the policy to 1 million houses.

6. A light bulb has a lifetime that is exponentially distributed with rate parameter $\lambda = 5$. Let L be a random variable denoting the sum of the lifetimes of 50 such bulbs. Assume that the bulbs are independent.

   (a) Compute $E[L]$ and $\mathrm{Var}(L)$.

   (b) Use the Central Limit Theorem to approximate $P(8 \leq L \leq 12)$.

   (c) Use the Central Limit Theorem to find an interval (a, b), *centered at* $E[L]$, such that

   $$P(a \leq L \leq b) = 0.95.$$

   That is, your interval should be of the form $(E[L] - c, E[L] + c)$, for some constant $c > 0$.