ENGRD 2700: Basic Engineering Probability and Statistics

Spring 2023

Prof. M. Chazal and S. Henderson

Homework 7

Due Thursday, April 20th at 11:59pm.

- The general guidelines for this homework are the same as for Homework 1.
- The main topics of this homework are Covariance, Correlation, Sampling Distributions, and the Central Limit Theorem. A good strategy before starting working on it is:
  - To review the lecture notes on Covariance&Correlation and on Sampling Distributions.
  - To make sure you have a good understanding of the problems treated in Recitation 10.

Questions

Please, justify all your answers. In particular, cite any result/theorem you use.

1. (20 points) The gain from one share of stock in company i over the coming year is random, and given by S_i, where S_i is uniformly distributed on $[-2, 6]$, $i = 1, 2, \ldots, 30$. (Negative values represent losses.) Also, the gains from the 30 companies are independent.

   i) Suppose that your portfolio consists of 1 share of stock from each company. What is the mean and variance of the overall gain of your portfolio over the coming year?

   ii) Suppose that your portfolio consists instead of 30 shares of stock S_1. What is the mean and variance of the overall gain of your portfolio over the coming year?

   iii) Compute if possible, or give an approximation for, the probability your portfolio makes an overall loss in both cases (i) and (ii) above.

   iv) Assume now that there are only 4 companies. The gains still have the same uniform distribution on $[-2, 6]$ but are possibly dependent, with the following correlation coefficients: $\text{Corr}(S_1, S_2) = 0$, and $\text{Corr}(S_1, S_3) = -0.5$ and $\text{Corr}(S_1, S_4) = -0.9$. Your portfolio must consist of one share of stock in company 1 and of one share of stock in only one of the other three companies. Which company would you choose? Explain why, comparing relevant expectations and variances.

2. (21 points) *For this question, please use Python for the computational parts and provide relevant pieces of code, label and give titles to your plots.*
   For each of the distributions (a), (b), and (c)

   $$\text{(a)} \quad X \sim N(\mu = 0, \sigma^2 = 4); \qquad \text{(b)} \quad X \sim \text{Exp}(\lambda = 1); \qquad \text{(c)} \quad X \sim \text{Pois}(\lambda = 5)$$

   answer the following questions.

   i) What are the mean and variance of the sample mean $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$?
   Compute them for $n = 5$, 10, 20 and 40.

   ii) Make a histogram of relative frequencies for each of the sample means $\overline{X}_5$, $\overline{X}_{10}$, $\overline{X}_{20}$ and $\overline{X}_{40}$. To make each histogram, generate 500 observations of $\overline{X}_n$.
   Comment on what you observe from the plots (mean, spread, skewness...)

   iii) Provide (the exact if possible or) an approximation of the sampling distribution of $\overline{X}_{40}$? Add the density plot of that distribution to the histogram of $\overline{X}_{40}$.

3. (7 points) An administrator of the Finger Lakes Medical and Wellness Center knows that of all patients 30% generate bills that become at least 3 months overdue.

   i) Suggest a statistic that represents the proportion of individuals who will generate bills that are overdue by at least 3 months, out of the next 200 patients? Assuming that these latter pay their bills independently, give the distribution of that statistic (by stating its PMF).

   ii) Find an approximation of the probability that, of all these 200 patients, less than 20% will generate bills that become at least 3 months overdue.

4. (7 points) A Flexible Spending Account (FSA) is a special account you put money into that you use to pay for out-of-pocket health care costs. You don't pay taxes on this money. This means you'll save an amount equal to the taxes you would have paid on the money you set aside. You must use the money in an FSA within the plan year or you lose the remaining balance. Assume your monthly out-of-pocket health care costs are normally distributed with mean \$100 and standard deviation \$10. Suppose that your monthly costs are independent. What amount M in dollars should you deposit on your FSA at the beginning of the year to ensure that with probability 0.99 the end-of-the-year remaining balance is less than half the taxes you would have paid on M? Assume that your tax rate is 10%.