Introduction to Probability and Statistics

Session Exercises

Israel Leyva-Mayorga

Exercise 1: Use the data in the moodle materials folder to plot the absolute, relative, and normalized histograms of

a) The latency for communication in our Starlink setup (file: Starlink_latency.csv)
b) The temperature measured by a sensor during a PhD course at Oulu University (file: temperature.csv)

Are these datasets approximately normal?

Exercise 2: Simulate rolling a fair dice n times.

a) Calculate the sample mean $\overline{X}_i$ and variance S_i^2 for all $i = 1, 2, \ldots, n$.
b) Plot the sample mean $\overline{X}_i$ and $\overline{X}_i + S_i$.
c) Plot a normalized histogram with the n outcomes. How many times do you need to roll the dice so the histogram resembles the pmf of a uniform RV?

Tip: You can use the code dice_loln.py in the moodle page as a base

Exercise 3: Consider the case of flipping a fair coin n times and let H_n be the random variable (RV) of the number of *heads*. Also recall that the sum of n Bernoulli RVs is a Binomial random variable with mean np and variance $np(2-p)$. Calculate and compare the probability of observing $H_{10} \geq 7$ and $H_{100} \geq 70$ using each of the following methods.

a) Summing over the formula for the probability mass function (pmf) of Binomial RVs

$$P(H_n \geq k) = \sum_{i=1}^{n} \binom{n}{k} p^k (1-p)^{n-i} \tag{1}$$

b) Using the central limit theorem

Exercise 4: A football team will play 60 games this year. Thirty-two of these games are against teams playing the champions league, denoted as class A teams, and 28 are against other teams, denoted as class B teams. The outcomes of the games are independent. The team will win each game against a class A team with probability 0.5, and it will win each game against a class B team with probability 0.7. Let X denote its total number of victories in the year.

a) Is X a Binomial RV?
b) Let X_A and X_B denote, respectively, the number of victories against class A and class B teams. What are their distributions?
c) What is the relationship between X, X_A and X_B?
d) Approximate the probability that the team wins 40 or more games this year

Exercise 5: The room temperature during a PhD course was recorded using a Raspberry Pi and a temperature sensor. The collected values are in the file temperature.csv (the temperature is the second column).

a) Calculate the sample mean $\overline{X}_n$ and variance S_n^2 of the temperature with all the $n = 253$ measurements
b) Assume that $\overline{X}_n$ is the true temperature μ and that S_n^2 is the real variance of the Gaussian noise that affects each measurement of the sensor, denoted by σ^2. Consequently, assume that the temperature measurements have a distribution $X_i \sim \mathcal{N}(\mu, \sigma^2)$. Plot the likelihood or log-likelihood function for the first 10 and with the first 100 measurements
c) Calculate the Maximum Likelihood Estimator (MLE) for μ with the first 10 and with the first 100 measurements under the previous assumption.