Homework 1 (due Friday, September 27, 2024)

Remark Please also turn-in your code and worksheet for any program that you used for this homework.

1. **Birthday Problem**

   In a set of n, $n \geq 2$, randomly chosen people, we want to find out the probability that there is some pair of them having the same birthday. (Assume that one year has exactly 365 days, and that the n birthdays are mutually independent uniform random variables.)

   Write a program that takes two inputs: the first input is the number of people, n, and second input is the number N of times you run the simulation. The output would be p_n, the (empirical) probability that there is some pair of them having the same birthday; that is, p_n equals the number of simulations in which there exist at least a pair with the same birthday divided by the total number of simulations (which is N).

   Answer the following questions.

   a) Based on your simulations, find an approximate 95% confidence interval for p_{10}, the probability that there is some pair of people having the same birthday when the number of randomly chosen people is 10.

   b) Based on your simulations, find an approximate 95% confidence interval for p_{20}, the probability that there is some pair of people having the same birthday when the number of randomly chosen people is 20.

   c) Based on your simulations, find an approximate 95% confidence interval for p_{30}, the probability that there is some pair of people having the same birthday when the number of randomly chosen people is 30.

   d) Run the program to observe empirically the least value of n such that the probability of having a pair of people with the same birthday is more than $\frac{1}{2}$.

   e) **(extra credit)** Derive the formula of p_n to back up the result of your experiment.

2. **Alice and Bob Play a Game.**

   Alice and Bob play the following game. Bob has two buttons. When the first button is pressed, the button says one number, uniformly at random chosen from $\{1, 2, \ldots, 99\}$, independently from all other presses by either button so far. Similarly, the second button, when pressed, says a number, uniformly at random chosen from $\{2, 3, \ldots, 100\}$, independently from all other presses by either button so far.

   The rule of this game is as follows. Bob selects one button from the two and gives it to Alice. Bob can select whatever button he wants to, and Alice does not know Bob's selection when Bob picks. (Without pressing them, Alice cannot distinguish the two buttons from the outside.) Alice is allowed to press the button given to her and record the outputs as many times as she wants. Alice wins if she says correctly if the button given to her is the first or the second. Otherwise, she loses.

Suppose that the rule of this game is known to both Alice and Bob. Describe a strategy Alice might take that guarantees at least 99% probability of winning for Alice, and show that the strategy works.

3. **Practice with Uniform and Geometric Distributions**

   Let $A_1, A_2, \ldots$ be a sequence of mutually independent tosses of a fair die so each A_i is a random variable following the distribution $\text{Unif}\{1, 2, \ldots, 6\}$. Let X be the smallest index n such that $A_n = 5$. (In other words, X is the number of tosses it takes until we see the number 5 for the first time, counting also the toss that 5 appears.)

   Suppose that a and b are two integers with $a < b$. Let $Y \sim \text{Unif}\{a, a+1, \ldots, b\}$.

   If it turns out that X and Y have the same means and the same variances (i.e. $\mathbb{E}(X) = \mathbb{E}(Y)$ and $\text{Var}(X) = \text{Var}(Y)$), what are a and b?