**HOMEWORK 1**

**1.** Basic notions in Statistics: Population, Statistical Units, Distribution;

**2.** Notion of average. Computational problems with floating point rapresentation ( errors, catastrophical cancellation) and numerical solution ( Knuth ) ;

**3.** (For exercise you can use Visual Studio: C# or VB, or other IDE for Javascript)

We have n servers with m attackers. The hacker has probability p to penetrate each server. Make a graphical representation (line flat if hacker doesn’t penetrate and a jump to 1 if he penetrates), try different n,m,p.

At time n we want to count how many reached each level.

Useful funcions: random, drawlines, drawrectangle

**1**

Population

* **Definition**: In statistics, a population refers to the entire set of individuals, objects, or events that are of interest in a particular study. The population encompasses all possible observations or data points that meet a certain criterion.
* **Example**: If you are studying the height of students in a school, the population would include every student enrolled in the school.
* **Types of Population**:
  + **Finite Population**: When the total number of individuals or units in the population is countable, such as the students in a school.
  + **Infinite Population**: When it is not possible to count all units or when the population is theoretically infinite, such as the number of times a coin could be tossed.

Statistical Units

* **Definition**: A statistical unit is an individual element or member of the population that is the focus of statistical measurement or analysis. These units are the entities about which data are collected.
* **Example**: If the population is "students in a school," each individual student is a statistical unit. If the study is about cars in a city, each car is a statistical unit.
* **Characteristics**:
  + **Homogeneous**: All units should be comparable and share the same characteristics that define the population.
  + **Distinguishable**: Each unit should be clearly identifiable and separable from other units.

Distribution

* **Definition**: A distribution describes how values or observations are spread across a set of possible outcomes or the range of a dataset. In statistics, distributions are used to understand the patterns and characteristics of data.
* **Types of Distributions**:
  + **Frequency Distribution**: Displays how frequently each value appears in a dataset.
  + **Probability Distribution**: Provides the likelihood of different outcomes in an experiment or process.
* **Example**: In a class, if the test scores are spread out from 60 to 100, the way these scores are arranged (i.e., how many students scored between 60-70, 70-80, etc.) represents the frequency distribution of the scores.

**2**

**Notion of Average (Mean)**

The average (or mean) is a central measure in statistics that represents the sum of all elements in a data set divided by the number of elements in the set. Mathematically, for a data set with *n* values x1,x2,…,xn the average is given by:

Average =

There are various types of averages, such as:

* **Arithmetic Mean**: The simple sum of values divided by the count.
* **Geometric Mean**: The nth root of the product of values.
* **Harmonic Mean**: The reciprocal of the arithmetic mean of the reciprocals of the data points.

**Computational Problems with Floating Point Representation**

In computing, numbers are typically represented using **floating point** arithmetic, which allows for a wide range of values. However, this representation is not perfect due to limited precision, leading to several types of errors:

**A. Rounding Errors**

Floating point numbers have limited precision, so many real numbers cannot be represented exactly. When a number is stored as a floating point value, it is often rounded to the nearest representable value, introducing a small error.

**B. Precision Loss**

As the number of digits increases, floating point numbers lose precision, especially in the least significant digits. This issue becomes particularly problematic when performing operations on numbers of widely varying magnitudes.

**C. Catastrophic Cancellation**

Catastrophic cancellation occurs when subtracting two nearly equal floating point numbers. In this case, most significant digits cancel out, and the result retains only the least accurate digits. This leads to a large relative error in the result, despite the operands being represented accurately. For example:

*x=1.000000000001* and *y=1.000000000000*

The result of x−y should be 1×10−12, but due to floating point precision, the result may be inaccurate.

**D. Overflow and Underflow**

* **Overflow** occurs when a number exceeds the largest value that can be represented in a floating point system.
* **Underflow** occurs when a number is smaller than the smallest representable value, often leading to it being rounded to zero.

**Numerical Solutions (Knuth's Perspective)**

Donald Knuth, a pioneer in computer science, has addressed the issue of floating point arithmetic in his work "The Art of Computer Programming." To minimize errors in floating point computations, Knuth suggests:

* **Avoiding Catastrophic Cancellation**: When subtracting nearly equal numbers, reformulate the problem to avoid the direct subtraction of large values that are close to each other.
* **Kahan Summation Algorithm**: This is a method for minimizing the error when adding a sequence of floating point numbers. It compensates for rounding errors by keeping a separate running total of the small errors that occur during the addition process.
* **Rearranging Computations**: Sometimes, performing calculations in a different order can reduce rounding errors. For example, summing numbers from smallest to largest may reduce the total error.