SIMULACION ESTOCASTICA 2021-2

Taller 2: Análisis estadístico de datos simulados, bootstrap y simulación de modelos

Indicaciones: Subir al link en AVATA un archivo *pdf* con el procedimiento analítico y análisis, y un archivo en `R` con los códigos de los puntos que lo requieran.
Fecha de entrega: Martes 19 de Abril de 2022

Ejercicios:

1. **Stopping generating new simulation data (1 point)** *(Ross)*

   Write a program to generate standard normal random variables until you have generated n of them, where $n \geq 100$ is such that $S/\sqrt{n} < 0.01$, where S is the sample standard deviation of the n data values. Note that this is the "Method for Determining When to Stop Generating New Data".

   a) How many normals do you think will be generated?
   b) How many normals did you generate?
   c) What is the sample mean of all the normals generated?
   d) What is the sample variance?
   e) Comment on the results of (c) and (d). Were they surprising?

2. **Gaining confidence with confidence intervals (1 point)** *(Jones, et al.)*

   We know that the $U(-1, 1)$ rv has mean 0. Use a sample of size 100 to estimate the mean and give a 95 % confidence interval. Does the confidence interval contain 0? Repeat the above a large number of times. What percentage of time does the confidence interval contain 0? Write your code so that it produces output similar to the following:

```
Number of trials: 10

Sample mean  lower bound  upper bound  contains mean?

    -0.0733      -0.1888       0.0422               1
    -0.0267      -0.1335       0.0801               1
    -0.0063      -0.1143       0.1017               1
    -0.0820      -0.1869       0.0230               1
    -0.0354      -0.1478       0.0771               1
    -0.0751      -0.1863       0.0362               1
    -0.0742      -0.1923       0.0440               1
     0.0071      -0.1011       0.1153               1
     0.0772      -0.0322       0.1867               1
    -0.0243      -0.1370       0.0885               1

100 percent of CI's contained the mean
```

3. **Bootstrap (1 point)** *(from Robert and Casella)*

   The code "*bootstrap.basic.example.RobertCasella.r*çalculates a bootstrap estimation (of size 2500) of the distribution of the mean $\overline{y}$ of the sample:

$$y = \{4.313, 4.513, 5.489, 4.265, 3.641, 5.106, 8.006, 5.087\},$$

   and compare it with the normal approximation from the Central Limit Theorem. It also shows the bootstrap estimation $\hat{q}_{.95}(\overline{y})$ of the 95 % quantile of $\overline{y}$, and the estimation $\hat{q}_{.5}(\overline{y})$ of its median.

   Modify the code to calculate bootstrap estimations of:

   a) The distribution of the sample standard deviation S of y. Plot it as an histogram.

   b) The mean $E[S]$ of S.

   c) The median $\hat{q}_{.5}(S)$ of the distribution of S.

   d) The variance $Var[S]$ of S.

4. **Simulation of the two dices (3 points)**

   a) Write a code to simulate the two fair dices in exercise 2 from Homework 1 and to calculate M, the smallest of the face values.

   b) Modify your code by running $n = 10^4$ independent simulations to estimate: (i) the expectation of M, (ii) the variance of M, and (iii) the probability that M is at least 3.

   c) Estimate the 95 % confidence intervals (CI's) of the three estimators and verify if they contain the exact values (from Exercise 2 in Homework 1).

   d) Modify the code to include a plot of each estimator and their 95 % CI's in terms of the size of the simulation $n = 1, 2, \ldots, 10^5$. Plot the horizontal lines corresponding to the exact values.