Prueba de Evaluación Continua de la asignatura «Estadística (código 7190105-)» Curso 2022-2023

13 de abril de 2023

1. Descripción general de la prueba

La prueba de evauación continua (PEC) de la asignatura «Estadística (código 7190105-)» consiste en resolver uno de los dos ejercicios propuestos en este documento. Ambos ejercicios tienen como objetivo revisar de una manera práctica algunos de los contenidos que se estudian en las unidades 1, 2, 3 y 4 del programa de la asignatura mediante el uso de procedimientos sencillos de simulación estocástica.

Los ejercicios no requieren emplear funciones complejas para la simulación, con lo cual se pueden resolver utilizando las funciones estadísticas elementales de una hoja de cálculo como Excel. Aunque este equipo docente no impone la herramienta a utilizar, sí recomienda el uso de la programación en R. Este lenguaje de programación es un estándar para el análisis científico de datos, tanto en el ámbito académico como en el sector privado; junto con Python, R se ha convertido en el lenguaje de programación para la computación, la visualización y el análisis de datos en el campo de Estadística y Ciencia de Datos.

R es un software de distribución libre que puede descargarse de la página web

`https://www.r-project.org/`

Si no se opta por esta alternativa se puede emplear cualquier otro lenguaje o herramienta; en la resolución de la PEC hay que hacer constar la herramienta utilizada.

Recordamos que, según figura en el documento de organización del curso, la evaluación continua supone un 10 % de la nota final y que el peso del examen en la calificación final es del 90 %. También recordamos que, según aparece en la información específica del Grado sobre la asignatura y de acuerdo con la decisión tomada en la reunión de las dos Comisiones de Grado de la Escuela de Informática, celebrada el día 31 de Mayo de 2010, la nota final no puede ser 10 si no se ha realizado esta prueba de evaluación continua.

2. Entrega de la prueba

La prueba se entregará en el plazo establecido en el curso virtual, se considerará el horario peninsular para determinar la hora límite a efectos de entrega. Dado que el sistema bloquea la entrega una vez que ha pasado la hora límite establecida, es recomendable no esperar al último momento para remitir la prueba porque no habrá tiempo de reacción ante cualquier eventualidad que pueda surgir; no se admitirán entregas fuera del plazo y hora establecidos. A fin de facilitar la labor de los tutores en la corrección de la prueba, ésta se debe entregar en un único fichero con formato *pdf*.

Se ruega encarecidamente disciplina en observar las indicaciones anteriores.

3. Ejercicios propuestos

El alumno debe elegir uno de los dos siguientes ejercicios.

3.1. Ejercicio 1

Se considera una variable aleatoria continua X que tiene función de distribución $F(x)$ definida por

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{x^2+x}{2} & \text{si } 0 \leq x < 1 \\ 1 & \text{si } x \geq 1 \end{cases}$$

1. Calcular el valor teórico de $E\{X\}$ (*máximo* 1 *página*).
2. Hallar la función de cuantiles de F y utilizarla para simular valores de la variable X a partir de los valores de una variable U con distribución uniforme en el intervalo $(0, 1)$. Construir tres escenarios de simulación a partir de $N = 10^3, 10^5, 10^7$ valores de U. Utilizar los resultados de las simulaciones para aproximar $E\{X\}$ y analizar la precisión de la aproximación en función del número N de valores simulados (*máximo* 2 *páginas*).

Recomendación. Si se desea resolver este ejercicio utilizando programación con código R es aconsejable leer algún manual introductorio de R antes de comenzar a programar. La siguiente función resultará de gran utilidad.

- `runif` para generar valores de una variable uniforme en $(0, 1)$.

3.2. Ejercicio 2

Sea X una variable aleatoria con función de densidad exponencial dada por

$$f(x; \theta) = \begin{cases} \theta e^{-\theta x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

A fin de estimar θ se toma una muestra aleatoria simple, $(X_1, X_2, \ldots, X_n)$, de tamaño n de la variable anterior y se consideran los siguientes estimadores:

$$T_1 = \frac{n}{\sum_{i=1}^{n} X_i} = \frac{1}{\overline{X}} \qquad \text{y} \qquad T_2 = \frac{n-1}{\sum_{i=1}^{n} X_i}.$$

donde $\overline{X}$ es la media muestral de las n observaciones muestrales.

El ejercicio consiste en simular $B = 10000$ valores de las distribuciones de ambos estimadores, a partir de muestras de tamaño n de la variable X cuando $\theta = 2$, con el objetivo de aproximar el sesgo de ambos estimadores mediante el método de Monte Carlo. A partir de los resultados de las simulaciones, se debe responder a las siguientes cuestiones.

1. Calcular los valores aproximados de $E\{T_1\}$ y $E\{T_2\}$ para $n = 10$, $n = 100$ y $n = 1000$. Describa brevemente el procedimiento de cálculo y comente los resultados obtenidos (*máximo* 1 *página*).

2. Aproximar el sesgo de ambos estimadores. ¿Cómo varía el sesgo en función del tamaño muestral? ¿Cuál de los dos estimadores emplearía teniendo en cuenta el criterio del sesgo? (*máximo* 1 *página*).

Recomendación. Si desea resolver este ejercicio utilizando programación con código R es aconsejable leer algún manual introductorio antes de comenzar a programar. Las siguientes funciones le serán de utilidad.

- `numeric` para crear un vector donde guardar los valores de T_1 y T_2.
- `for` si necesita programar el bucle con las simulaciones. Para quienes quieran profundizar en los aspectos computacionales de la programación en R, un atajo es utilizar la función `apply` sobre la matriz de B filas y n columnas que guarda los datos simulados; las funciones de la clase `apply` utilizan la aritmética vectorizada del R evitando ciertos bucles, lo cual conlleva el consiguiente ahorro computacional.
- `rexp` para generar muestras de una distribución exponencial.
- `sum` para sumar los valores de un vector numérico.
- `mean` para calcular el promedio.

La ayuda sobre la sintaxis de cualquiera de las anteriores funciones se obtiene tecleando `?nombre de función` en la ventana de comandos.