PRÁCTICO 6

ANÁLISIS ESTADÍSTICO DE DATOS SIMULADOS.

Ejercicio 1. Genere n valores de una variable aleatoria normal estándar de manera tal que se cumplan las condiciones: $n \geq 100$ y $S/\sqrt{n} < 0{,}1$, siendo S el estimador de la desviación estándar de los n datos generados.

a) ¿Cuál es el número de datos generados efectivamente?

b) ¿Cuál es la media muestral de los datos generados?

c) ¿Cuál es la varianza muestral de los datos generados?

Ejercicio 2. Estime mediante el método de Monte Carlo la integral

$$\textbf{i)} \int_0^1 \frac{e^x}{\sqrt{2x}}\,dx, \qquad \textbf{ii)} \int_{-\infty}^{\infty} x^2 \exp(-x^2)\,dx.$$

a) Indique cómo se obtiene mediante simulación el valor de la integral.

b) Genere al menos 100 valores y deténgase cuando la desviación estándar muestral S del estimador sea menor que 0,01.

Ejercicio 3. Calcule mediante un método de Monte Carlo las siguientes integrales:

$$\textbf{i)} \int_{\pi}^{2\pi} \left(\frac{\sin(x)}{x}\right) dx \qquad \textbf{ii)} \int_0^{\infty} \frac{3}{3+x^4}$$

a) Obtenga mediante simulación en computadora el valor de la integral deteniendo la simulación cuando el semi-ancho del intervalo de confianza del 95 % sea justo inferior a 0,001.

c) Indique cuál es el número de simulaciones N_s necesarias en la simulación realizada para lograr la condición pedida y complete con los valores obtenidos la siguiente tabla (usando **4** decimales):

N^o de sim.	$\bar{I}$	S	IC(95 %)
1 000			
5 000			
7 000			
$N_s =$			

Ejercicio 4. Para $U_1, U_2, \ldots$ variables aleatorias uniformemente distribuidas en el intervalo $(0,1)$, se define:

$$N = \text{Mínimo}\left\{ n : \sum_{i=1}^{n} U_i > 1 \right\}$$

Esto es, N es igual a la cantidad de números aleatorios que deben sumarse para exceder a 1.

a) Observe que $E[N] = e$ por lo cual puede aproximar e con la media muestral $\bar{N}$.

b) Derive una expresión de la varianza del estimador $\bar{N}$ y aproxímela con 1000 simulaciones. Dar su estimador de máxima verosimilitud.

c) Dé el valor obtenido de la varianza muestral de $\bar{N}$ correspondiente a 1000 ejecuciones de la simulación y dar una estimación de e mediante un intervalo de confianza de 95 % con longitud a lo sumo 0.025.

Ejercicio 5. Considere una sucesión de números aleatorios $\{U_i\}_i$ y sea M el primer n tal que la variable U_n es menor que su variable predecesora. Es decir,

$$M = n \quad \text{tal que} \quad U_1 \leq U_2 \leq \cdots \leq U_{n-1} \text{ y } U_n < U_{n-1}$$

a) Justifique que $P(M > n) = 1/n!$, $n \geq 0$.

b) Utilice la identidad

$$E[M] = \sum_{n=0}^{\infty} P(M > n)$$

para mostrar que $E[M] = e$.

c) Utilice el resultado del item anterior para dar un estimador de $E[M]$, calcule el valor de su varianza muestral. Mediante una simulación estime el valor de e deteniéndose cuando la varianza muestral sea menor que 0,01.

d) Dé una estimación de e mediante un intervalo de ancho menor que 0,1 y con una confianza del 95 %

Ejercicio 6. Estime π sorteando puntos uniformemente distribuidos en el cuadrado cuyos vértices son: $(1,1)$, $(-1,1)$, $(-1,-1)$, $(1,-1)$, y contabilizando la fracción que cae dentro del círculo inscrito de radio 1.

a) Utilice un algoritmo para estimar la proporción de puntos que caen dentro del círculo y deténgase cuando la desviación estandar muestral del estimador sea menor que 0,01.

b) Obtenga un intervalo de ancho menor que 0,1, el cual contenga a π con el 95 % de confianza. ¿Cuántas ejecuciones son necesarias?

TÉCNICAS DE VALIDACIÓN ESTADÍSTICA

Para estos ejercicios, estudiar

`https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.chi2.html`

Otra bibliografia interesante es

`https://relopezbriega.github.io/blog/2016/06/29/distribuciones-de-probabilidad-con-python/`

No utilice implementaciones personales de densidades a menos que el ejercicio se lo pida exactamente.

Convención general: al obtener valores decimales suponemos una distribución continua.

Ejercicio 7. De acuerdo con la teoría genética de Mendel, cierta planta de guisantes debe producir flores blancas, rosas o rojas con probabilidad $1/4$, $1/2$ y $1/4$, respectivamente. Para verificar experimentalmente la teoría, se estudió una muestra de 564 guisantes, donde se encontró que 141 produjeron flores blancas, 291 flores rosas y 132 flores rojas. Aproximar el p-valor de esta muestra:

a) utilizando la prueba de Pearson con aproximación chi-cuadrada,

b) realizando una simulación.

Ejercicio 8. Para verificar que cierto dado no estaba trucado, se registraron 1000 lanzamientos, resultando que el número de veces que el dado arrojó el valor i ($i = 1,2,3,4,5,6$) fue, respectivamente, 158, 172, 164, 181, 160, 165. Aproximar el $p-$valor de la prueba: “el dado es honesto”

a) utilizando la prueba de Pearson con aproximación chi-cuadrada,

b) realizando una simulación.

Ejercicio 9. Calcular una aproximación del $p-$valor de la hipótesis: “Los siguientes 10 números son aleatorios”:

`0.12, 0.18, 0.06, 0.33, 0.72, 0.83, 0.36, 0.27, 0.77, 0.74.`

Ejercicio 10. Calcular una aproximación del $p-$valor de la hipótesis: “Los siguientes 13 valores provienen de una distribución exponencial con media 50,0”:

`86.0, 133.0, 75.0, 22.0, 11.0, 144.0, 78.0, 122.0, 8.0, 146.0, 33.0, 41.0, 99.0`

Ejercicio 11. Calcular una aproximación del $p-$valor de la prueba de que los siguientes datos corresponden a una distribución binomial con parámetros $(n = 8, p)$, donde p no se conoce:

`6, 7, 3, 4, 7, 3, 7, 2, 6, 3, 7, 8, 2, 1, 3, 5, 8, 7.`

Ejercicio 12. Un escribano debe validar un juego en cierto programa de televisión. El mismo consiste en hacer girar una rueda y obtener un premio según el sector de la rueda que coincida con una aguja. Hay 10 premios posibles, y las áreas de la rueda para los distintos premios, numerados del 1 al 10, son respectivamente:

`31%, 22%, 12%, 10%, 8%, 6%, 4%, 4%, 2%, 1%.`

Los premios con número alto (e.j. un auto 0Km) son mejores que los premios con número bajo (e.j. 2x1 para entradas en el cine). El escribano hace girar la rueda hasta que se cansa, y anota cuántas veces sale cada sector. Los resultados, para los premios del 1 al 10, respectivamente, son:

`188, 138, 87, 65, 48, 32, 30, 34, 13, 2.`

(a) Construya una tabla con los datos disponibles

(b) Diseñe una prueba de hipótesis para determinar si la rueda es justa

(c) Defina el p-valor a partir de la hipótesis nula

(d) Calcule el p-valor bajo la hipótesis de que la rueda es justa, usando la aproximación chi cuadrado

(e) Calcule el p-valor bajo la hipótesis de que la rueda es justa, usando una simulación.

Ejercicio 13. Generar los valores correspondientes a 30 variables aleatorias exponenciales independientes, cada una con media 1. Luego, en base al estadístico de prueba de Kolmogorov-Smirnov, aproxime el $p-$valor de la prueba de que los datos realmente provienen de una distribución exponencial con media 1.

Ejercicio 14. Se sortean elementos de un conjunto de datos que tiene una distribución t-student de 11 grados de libertad. El investigador, que no conoce la forma verdadera de la distribución, asume que la misma es normal.

Analice la validez de esta suposición para muestras de tamaños 10, 20, 100 y 1000 elementos Para ello realice simulaciones numéricas e implemente el test de Kolmogorov-Smirnov a los datos simulados, asumiendo una distribución N(0,1). Presente los resultados en una tabla que contenga el número de elementos de la simulación, el valor del estadístico D y el $p-$valor

Ayuda: Función de probabilidad normal: Para obtener la función de probabilidad normal, se puede usar la función `math.erf`. Por ejemplo, la cantidad `math.erf(x/math.sqrt(2.))/2.+0.5` equivale a

$$\int_{-\infty}^{x} N(0,1)(t)\,dt \tag{1}$$

Ayuda: Generación de números aleatorios con una distribución t-student: Utilice el siguiente código para generar números aleatorios que siguen una distribución T-student:

```
import math
import random

def rt(df): # df grados de libertad
    x = random.gauss(0.0, 1.0)
    y = 2.0*random.gammavariate(0.5*df, 2.0)
    return x / (math.sqrt(y/df))
```

Ejercicio 15. En un estudio de vibraciones, una muestra aleatoria de 15 componentes del avión fueron sometidos a fuertes vibraciones hasta que se evidenciaron fallas estructurales. Los datos proporcionados son los minutos transcurridos hasta que se evidenciaron dichas fallas.

1.6, 10.3, 3.5, 13.5, 18.4, 7.7, 24.3, 10.7, 8.4, 4.9, 7.9, 12, 16.2, 6.8, 14.7

Pruebe la hipótesis nula de que estas observaciones pueden ser consideradas como una muestra de la distribución exponencial.

Ejercicio 16. Decidir si los siguientes datos corresponden a una distribución Normal:

91.9, 97.8, 111.4, 122.3, 105.4, 95.0, 103.8, 99.6, 96.6, 119.3, 104.8, 101.7

Calcular una aproximación del $p-$valor.