Probabilidad y Estadística – Licenciatura en Computación

Práctico de Simulación

Problema 1: Un matemático del siglo dieciocho llamado Buffon, sugirió la siguiente técnica para calcular el valor de π: Tomar una gran hoja de papel, la cual se encuentra rayada con finas líneas paralelas separadas entre sí por dos centímetros y arrojar al azar sobre el papel una fina aguja de un centímetro de largo, de forma que caiga acostada, un gran número de veces. Se calcula la proporción de veces que la aguja interseca alguna línea del papel y el valor de π se estima como la inversa de dicha proporción.

a) Escribir un programa que simule en la computadora la técnica mencionada de arrojar la aguja. Esto resultará mucho más cómodo que efectivamente dejar caer una aguja sobre una hoja rayada. Notar que el problema es equivalente a sortear uniformente la posición del centro de la aguja en un intervalo de ancho 2, y sortear el ángulo que forma la aguja de largo 1 con la dirección de la recta que contiene al intervalo de interés, uniformemente entre 0 y π. Para cada prueba verificar a continuación si la proyección de la aguja sobre la recta mencionada, interseca algún extremo del intervalo.

Realizar una tabla con los valores de π estimados para un número n de pruebas, con $n = 1000, 2000, \ldots,$ $10000, 20000, \ldots, 100000$. Estimar el valor de n necesario para asegurar el valor de π con cinco cifras significativas (es decir que el error afecte recién al sexto decimal).

Nota: Para realizar la simulación el programa utilizará el valor de π que quiere estimarse. Esto sin embargo, no constituye un fraude dado que sólo se lo utiliza para simular el proceso de arrojar la aguja, no en el cálculo del número en sí.

b) En base a los conocimientos de la Teoría de la Probabilidad aprendidos en el curso, elabore un argumento que explique porqué el método de arrojar la aguja funciona para estimar el valor de π.

Problema 2: Presentaremos un versión simplificada de un problema de interés en el diseño de bases de datos. Consideremos objetos unidimensionales caracterizados por su longitud L, la cual es una variable aleatoria continua cuya densidad de probabilidad está dada por:

$$f_L(l) = \begin{cases} 2\,(1-l)\,, & \text{si } l \in (0,1) \\ 0\,, & \text{si } l \notin (0,1) \end{cases}$$

Supongamos se escogen al azar de manera independiente n objetos, cuyas longitudes resultan: $L_1, L_2, \ldots L_n$. Estos objetos deben ser guardados consecutivamente en cajas unidimensionales de longitud 1 y en el orden de sus correspondientes índices. Además, ningún objeto puede ser partido en dos cajas.

Esta última condición implica que puede generarse una gran cantidad de espacio desperdiciado en las cajas utilizadas. Sea W la cantidad de espacio desperdiciado y B el número de cajas empleadas para almacenar los n objetos. Obviamente se cumple que $B \leq n$. Para clarificar la notación y las condiciones de almacenamiento consideremos el siguiente ejemplo:
Supongamos $n = 3$, y $L_1 = 0.62$, $L_2 = 0.25$ y $L_3 = 0.44$. Entonces los objetos 1 y 2 se acomodan en la caja 1 y el objeto 3 se acomoda en la caja 2. Se tiene que $1 - (0.62 + 0.25) = 0.13$ es el espacio vacío en la primer caja y $1 - 0.44 = 0.56$ es el espacio desperdiciado en la caja 2. Por lo tanto, $W = 0.13 + 0.56 = 0.69$ y $B = 2$.

Asumir $n = 10$ y mediante simulaciones estimar:

a) el valor medio y la varianza de B.

b) $P(B \leq t)$ para $t = 1, 2, \ldots, n$.

c) el valor medio del cociente W/B, es decir la proporción esperada de espacio vacío.

Problema 3: Consideremos un conjunto de n variables aleatorias independientes e idénticamente distribuídas $X_1, X_2, \ldots, X_n$, cuya densidad de probabilidad es $P_X(x) = \exp(-x)$. Consideremos ahora la variable aleatoria:

$$\overline{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Esto es el promedio de las n variables dadas.

a) Construir un histograma para las frecuencias relativas de la variable aleatoria X cuya densidad de probabilidad es $P_X(x)$. Sortear 10000 valores de la variable.

b) Construir los histogramas de las variables aleatorias $\overline{X}_2$, $\overline{X}_3$, $\overline{X}_5$ y $\overline{X}_{10}$. Para el calculo de la variable $\overline{X}_n$ primero debe sortearse un conjunto de n valores de la variable X, calcular un valor de $\overline{X}_n$, sortear un nuevo conjunto de n valores de la variable X para calcular otro valor de $\overline{X}_n$ y así sucesivamente hasta tener el número de valores deseados en el histograma de para la variable $\overline{X}_n$.

c) A partir de los datos generados en el item anterior, calcular el valor medio y la varianza de las variables $\overline{X}_2$, $\overline{X}_3$, $\overline{X}_5$ y $\overline{X}_{10}$.

d) Superponer sobre la gráfica de cada uno de los histogramas construídos, la curva correspondiente a la densidad de probabilidad de una variable normal cuya media y varianza son iguales a las de la variable $\overline{X}_n$ del correspondiente histograma.