Facultad de Matemática, Astronomía y Física

Universidad Nacional de Córdoba

Probabilidad y Estadística – Licenciatura en Computación

Práctico Especial 2001

Problema de Simulación

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

$$f_L(l) = \begin{cases} a \cos\left(\frac{\pi}{2}\, l\right) & , \text{ si } l \in (0,1) \\ 0 & , \text{ si } l \notin (0,1) \end{cases}$$

a) Calcular el valor de la constante a.

b) Construir la función de distribución acumulada de la variable L.

c) Calcular exactamente $E[L]$ y $\mathrm{Var}[L]$.

Se escogen ahora 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 $B = 2$.

Con el fin de realizar simulaciones numéricas, desarrollar un programa que reproduzca el algoritmo de almacenamiento de los objetos en las cajas. Probar su funcionamiento reproduciendo el ejemplo anterior.

Desarrollar un generador para la variable aleatoria L, empleando la idea de la transformación integral desarrollada en la sección 13.5 del libro de referencia ‡.

d) Generar 100000 valores de L empleando el generador construído y calcular el promedio y varianza de la muestra obtenida. Comparar estos resultados con los calculados en el item **(c)**.

Asumir ahora $n = 10$ y mediante simulaciones numéricas en la computadora (utilizar 100000 simulaciones), estimar:

e) el valor medio y la varianza de B

f) $P(B = t)$ para $t = 1, 2, \ldots, n$,

g) el valor medio y la varianza de W,

h) el valor medio del cociente W/B, es decir la proporción esperada de espacio vacío, y la correspondiente varianza de esta cantidad.

Problema de Estadística

Para el estudio del clima y su cambio se requiere de la acumulación de mediciones de varias magnitudes meteorológicas durante períodos de tiempo que superen el siglo. Sobre estas series temporales de mediciones se implementa el análisis estadístico. En particular, la estación meteorológica Córdoba funciona de manera continuada desde el año 1872, siendo una de las primeras en establecerse en América.

Con el fin de implementar un primer estudio con base estadística del clima en Córdoba, en el archivo `lluvias.dat` se compilan los registros pluviométricos anuales (en mm) correspondientes al período 1873–1992. Utilizando la muestra de valores de lluvia caída en Córdoba:

a) Identificar los valores máximos, mínimos y la mediana.

b) Calcular la media y la varianza.

c) Confeccionar un histograma de frecuencias relativas tomando 20 particiones en el intervalo $[400, 1400]$ mm.

d) Realizar el estudio de cuantiles en la muestra y confeccionar el correspondiente “box plot”.

Con el fin de establecer si existe cambio climático se propone:

e) Graficar la lluvia caída (en mm) en función del tiempo (en años) para el período 1876–1990.

(Excluímos adrede los valores extremos de la muestra para el subsiguiente análisis).

f) Considerar las submuestras: **A** correspondiente al intervalo de años [1876,1935] y **B** correspondiente al intervalo de años [1936,1990]. Realizar un ajuste lineal mediante el método de los mínimos cuadrados (desarrollado en la sección 14.5 ‡) de ambas submuestras. Destacar los valores de las pendientes (α) obtenidas y sus correspondientes varianzas. Superponer las rectas calculadas sobre la gráfica del item anterior.

g) Calcular el p-valor de la prueba de hipótesis: $\alpha_{\mathrm{A}} < \alpha_{\mathrm{B}}$.

h) ¿Qué conclusión puede extraerse del análisis realizado?

Referencia:

‡ *Probabilidad y Aplicaciones Estadísticas.* Paul L. Meyer (Addison Wesley Longman, edición revisada, 1998)