Estadística I
Grado en Matemáticas, UAM, 2018-2019

Hoja 3. Muestreo aleatorio y estadísticos

Nota: para calcular en excel valores de funciones de distribución $F_X(x) = \mathbf{P}(X \leq x)$, usa

- `=distr.norm.estand(x)` cuando X sea una normal estándar,
- `=distr.chicuad(x;n;verdadero)`, cuando X sea una χ^2 con n grados de libertad,
- `=distr.t.n(x;n;verdadero)`, cuando X sea una t de Student con n grados de libertad.

Sobre estadísticos

1. Consideremos una muestra aleatoria de tamaño $n = 50$ de una variable X que sigue una distribución $\mathcal{N}(4, 1)$ (una normal con media 4 y varianza 1).

a) Utiliza la desigualdad de Chebyshev para obtener una cota superior para $\mathbf{P}(|\overline{X} - 4| > 0.3)$.

b) Calcula numéricamente la probabilidad $\mathbf{P}(|\overline{X} - 4| > 0.3)$ utilizando que $\overline{X}$ es una normal. Compara el resultado con el del apartado a). Medita al respecto.

2. Definimos la asimetría y la curtosis de una variable Z con media μ y desviación típica σ como

$$\operatorname{asim}(Z) = \frac{\mathbf{E}((Z-\mu)^3)}{\sigma^3} \qquad \text{y} \qquad \operatorname{curt}(Z) = \frac{\mathbf{E}((Z-\mu)^4)}{\sigma^4}.$$

Sea $X_1, \dots, X_n$ una muestra aleatoria de X. Comprueba que

$$\operatorname{asim}(\overline{X}) = \frac{\operatorname{asim}(X)}{\sqrt{n}} \qquad \text{y que} \qquad \operatorname{curt}(\overline{X}) = \frac{\operatorname{curt}(X)}{n} + 3\,\frac{n-1}{n}.$$

3. La variable X toma los valores $+1$ y -1 con probabilidad $1/2$ cada uno de ellos. Consideramos muestras aleatorias de X de tamaño 5. Determina cómo son (qué valores toman, y con qué probabilidades) las variables $\overline{X}$ y S_X^2.

4. La variable X toma los valores 1, 2 y 3 con probabilidad $1/3$ cada uno de ellos. Consideramos muestras aleatorias de X de tamaño 3. Determina cómo son (qué valores toman, y con qué probabilidades) las siguientes variables: Z_1 es el valor más grande de la muestra, Z_2 es el valor más pequeño, y Z_3 es el segundo valor más pequeño (es decir, y en este caso de muestras de tamaño 3: tras ordenar la muestra de menor a mayor, se toma el valor que queda en medio).

5. Sea $X \sim \text{EXP}(\lambda)$, con $\lambda > 0$. Consideramos el estadístico $m_n = \text{mín}(X_1, \dots, X_n)$. Comprueba que $m_n \sim \text{EXP}(n\lambda)$ y calcula $n\mathbf{E}(m_n)$.

6. Sea $X \sim \text{UNIF}([0, a])$, con $a > 0$. Consideramos el estadístico $M_n = \text{máx}(X_1, \dots, X_n)$.

a) Determina las funciones de distribución y densidad de M_n, y calcula $\mathbf{E}(M_n)$ y $\mathbf{V}(M_n)$.

b) Calcula, para un $\varepsilon > 0$ dado,

$$\mathbf{P}(a - M_n \geq \varepsilon\, a).$$

Muestreo de normales

Nota: en los ejercicios que siguen puedes dejar escrita la respuesta en términos de valores de la función de distribución de la normal estándar y/o de las funciones de distribución de una χ^2 o una t de Student con un cierto número de grados de libertad. Si esos valores se puedan calcular explícitamente (con Excel), ¡hazlo!

7. Consideramos muestras aleatorias de tamaño 35 de una normal con parámetros $\mu = 1$ y $\sigma^2 = 2$.

a) ¿Cuál es la probabilidad de que la media muestral y la cuasidesviación típica muestral sean, simultáneamente, inferiores a 1.2?

b) Suponiendo que la cuasidesviación típica de la muestra es menor que 1.2, ¿cuál es la probabilidad de que la media muestral sea también menor que 1.2?

8. Sea $(X_1, \dots, X_{100})$ una muestra aleatoria de X normal con parámetros $\mu = 3$ y $\sigma^2 = 4$. Calcula $\mathbf{P}\big(|\overline{X} - 2| > 1/2\,,\ S^2 > 4.2\big)$.

9. Sea $(X_1, \dots, X_{20})$ una muestra aleatoria de normales estándar. Calcula $\mathbf{P}(\overline{X} > S)$.

10. Sea $(X_1, \dots, X_{100})$ una muestra aleatoria de una variable $X \sim \mathcal{N}(0, \sigma^2)$.

a) Calcula $\mathbf{P}\big(|\overline{X}| > \sigma\,,\ S^2 > 2\sigma^2\big)$.

b) Calcula la probabilidad de que, o bien ocurra que $|\overline{X}| > \sigma$, o bien que $S^2 > 2\sigma^2$ (estos "o" no son excluyentes).

11. Sea $(X_1, \dots, X_n)$ una muestra aleatoria de una variable $X \sim \mathcal{N}(\mu, \sigma^2)$. Para $t > 0$ fijo, calcula la probabilidad de que $n\,(\overline{X} - \mu)^2 + (n-1)\,S^2$ sea $\geq t$.

Ejercicio adicional

12. Dada una muestra aleatoria $X_1, X_2, \dots, X_n$ de X, el *estadístico de orden* $X_{(r:n)}$ se define, para cada entero $1 \leq r \leq n$, como sigue: se ordena la muestra $(X_1, X_2, \dots, X_n)$ de menor a mayor,

$$X_{(1:n)} \leq X_{(2:n)} \leq \cdots \leq X_{(n:n)}\,,$$

y se toma el valor que ocupa la posición r-ésima en esta ordenación de menor a mayor. El primer estadístico, $X_{(1:n)}$, es el mínimo de la muestra. El último, $X_{(n:n)}$, es el máximo de la muestra.

a) Consideremos el estadístico $X_{(n-1:n)}$, es decir, el segundo término más grande de entre $X_1, X_2, \dots, X_n$. Comprueba que

$$F_{X_{(n-1:n)}}(t) = n\,(1 - F_X(t))\,F_X(t)^{n-1} + F_X(t)^n.$$

b) Suponiendo que la variable aleatoria X tiene función de densidad $f_X(t)$, comprueba que

$$f_{X_{(n-1:n)}}(t) = n(n-1)\,f_X(t)\,F_X(t)^{n-2}\,(1 - F_X(t)).$$

c) Consideremos ahora el caso general del estadístico $X_{(r:n)}$, para cierto $1 \leq r \leq n$. Comprueba que

$$F_{X_{(r:n)}}(t) = \sum_{j=r}^{n} \binom{n}{j} F_X(t)^j\,(1 - F_X(t))^{n-j}.$$

¿Qué se obtiene cuándo $r = n$ y cuando $r = 1$?

d) Suponiendo que la variable aleatoria X tiene función de densidad $f_X(t)$, comprueba que

$$f_{X_{(r:n)}}(t) = n\binom{n-1}{r-1} f_X(t)\,F_X(t)^{r-1}\,(1 - F_X(t))^{n-r}.$$