Estadística I
Tercero del grado en Matemáticas, UAM, 2017-2018

Examen convocatoria extraordinaria, 15-6-2018

1. ($1\frac{1}{3}$ puntos) La función de masa de la variable X viene dada por

valores	1	2	3
probabilidades	p_1	p_2	p_3

El modelo tienen únicamente dos parámetros, p_1 y p_2, puesto que se debe cumplir que $p_1 + p_2 + p_3 = 1$. Se tiene que $p_1 \in (0,1)$, $p_2 \in (0,1)$ y además que $p_1 + p_2 < 1$.

En una muestra de X de tamaño 100 se han obtenido los siguientes resultados:

valores	1	2	3
número de apariciones	20	25	55

Halla la *estimación* de p_1 y p_2 por máxima verosimilitud.

2. (2 puntos) a) El vector $\mathbb{X} = (X_1, X_2, X_3)^\intercal$ sigue una normal tridimensional, con parámetros:

vector de medias: $\mathbf{m} = (1,1,0)$, matriz de varianzas/covarianzas: $V = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{pmatrix}$.

Consideramos el vector $\mathbb{Z} = (Z_1, Z_2, Z_3)^\intercal$ dado por

$$\begin{cases} Z_1 &= X_1 + X_2 \\ Z_2 &= X_1 + X_2 + X_3 \\ Z_3 &= 2X_1 + X_2 \end{cases}$$

Calcula las medias de Z_1, Z_2 y Z_3, la varianza de Z_3, y la covarianza entre Z_1 y Z_2.

b) ¿Cuál es la probabilidad de que, en una muestra aleatoria de tamaño 35 de una normal con parámetros $\mu = 1$ y $\sigma^2 = 2$, la media muestral y la cuasidesviación típica muestral sean, simultáneamente, inferiores a 1.2?

Sugerencia. Puedes/debes dejar la respuesta en términos de funciones de distribución de las variables habituales: normal estándar, chi cuadrado, t de Student, etc.

3. (2 puntos) Considera la variable X con función de densidad

$$f(x;a) = \frac{2}{a^2}\,x \qquad \text{para } x \in [0,a],$$

donde a es un parámetro positivo. Consideramos muestras de X de tamaño n.

(a) Comprueba que el estimador del parámetro a dado por

$$T_1(X_1, \ldots, X_n) = \frac{3}{2}\,\overline{X}$$

es un estimador insesgado de a. Calcula su varianza.

(b) El estimador

$$T_2(X_1, \ldots, X_n) = \text{máx}(X_1, \ldots, X_n)$$

no es insesgado. Calcula su sesgo.

4. ($1\frac{1}{3}$ puntos) La variable X sigue una cierta distribución que depende de un parámetro $a > 0$. Los valores de sus primeros momentos son

$$\mathbf{E}_a(X) = a^{3/2}, \qquad \mathbf{E}_a(X^2) = 2a^3, \qquad \mathbf{E}_a(X^3) = 5a^{9/2}, \qquad \mathbf{E}_a(X^4) = 15\,a^6.$$

Para muestras de tamaño n de X, consideramos el siguiente estimador del parámetro a:

$$T(X_1, \ldots, X_n) = \Big(\frac{\overline{X^2}}{2}\Big)^{1/3}.$$

Escribe un resultado de normalidad asintótica para T usando el método delta.

5. (2 puntos) En los dos apartados de este ejercicio, sugerimos hacer los cálculos con tres cifras decimales significativas.

a) Interesa estudiar la proporción p de jugadores de Fortnite entre los jóvenes entre 15 y 20 años.

a1) Determina el mínimo tamaño de la muestra necesario para estimar el valor de p con un error menor del 2 % (y un nivel de confianza del 95 %). Justifica la respuesta.

a2) Se ha encuestado a 3200 jóvenes entre 15 y 20 años. El 54 % de ellos declararon jugar al Fortnite. Calcula el intervalo de confianza al 97.5 % para la proporción p.

Datos:

- $z_{10\,\%} = 1.282$, $z_{5\,\%} = 1.645$, $z_{2.5\,\%} = 1.960$, $z_{1.25\,\%} = 2.241$.

b) Interesa investigar el tiempo que tardan en ejecutar un cierto algoritmo dos modelos de procesador: el Letni-8086 y el DMA-300. Se ha realizado el siguiente experimento:

- se ha ejecutado el algoritmo en 100 ordenadores idénticos equipados con procesador Letni-8086. El promedio de los tiempos de ejecución ha sido 37 segundos, y la varianza muestral ha sido 41.
- se ha ejecutado el algoritmo en 130 ordenadores idénticos equipados con procesador DMA-300. El promedio de los tiempos de ejecución ha sido 29 segundos. El promedio de los cuadrados de esos tiempos ha sido 900.

Asumimos normalidad en la variable “tiempo de ejecución” en ambos casos.

A la vista de estos datos, ¿hay suficiente evidencia estadística como para concluir que la variabilidad en el tiempo de ejecución del algoritmo para el procesador Letni-8086 es *menor* que para el procesador DMA-300? Justifica la respuesta (la hipótesis planteada, su contraste, y la conclusión obtenida).

Datos:

- Algunos percentiles de una F de Fisher:

α	5 %	4.5 %	4.0 %	3.5 %	3.0 %	2.5 %	2.0 %	1.5 %	1.0 %	0.5 %
$F_{\{99,129,\alpha\}}$	1.361	1.374	1.389	1.405	1.423	1.444	1.470	1.503	1.547	1.622
$F_{\{129,99,\alpha\}}$	1.371	1.385	1.399	1.416	1.435	1.457	1.484	1.518	1.566	1.643

6. ($1\frac{1}{3}$ puntos) Modelamos el número de días que sobrevive una cierta bacteria con una variable aleatoria X que sigue una distribución geométrica de parámetro $p \in (0, 1)$.

Se pretende contrastar la hipótesis $H_0 : p > 0.2$.

Para ello, se diseña el siguiente test:

- se analizarán 10 bacterias,
- se medirá cuántos días sobrevive cada bacteria,
- y si todas ellas duran 3 o más días, entonces se rechaza la hipótesis.

Halla la función de potencia del test y su nivel de significación.

Nota: si X sigue una distribución geométrica con parámetro $p \in (0, 1)$, entonces su función de masa viene dada por

$$\mathbf{P}(X = j) = p(1-p)^{j-1}, \qquad \text{para } j = 1, 2, 3, \ldots$$