Grado en Ingeniería Informática

Examen de Matemáticas III (Convocatoria de junio, 2017)

Ejercicio 1: (0.75 pto) Las diagonales de un polígono se obtienen uniendo dos vértices no consecutivos del mismo. Usando técnicas de combinatoria, determinar razonadamente el número de diagonales del cuadrado y el hexágono así como el número de diagonales para el caso general de un polígono de n lados. ¿Existe algún polígono en el que el número de lados sea igual al número de diagonales?

Ejercicio 2. En la red informática de una empresa hay tres sistemas multiusuario: S_1, S_2 y S_3. Las peticiones de conexión que se realizan a estos equipos se reparten de manera que el 50% se efectúan sobre S_1, el 30% sobre S_2 y el 20% sobre S_3. Los tiempos de respuesta a estas peticiones son variables aleatorias X_1, X_2 y X_3 respectivamente, expresadas en segundos, de las que se sabe lo siguiente:

- La función de distribución del tiempo de respuesta de S_1 es

$$F(x) = \begin{cases} 1 - e^{-x/3} & \text{si} \quad x > 0 \\ 0 & \text{si} \quad x \leq 0 \end{cases}$$

- El tiempo de respuesta de S_2 es cualquier valor real entre 4 y 8, siendo constante la correspondiente función de densidad en dicho intervalo.
- Sólo hay tres posibles tiempos de respuesta para S_3: 5, 6 ó 7, y todos ellos con la misma probabilidad.

Se pide:

a) (0.75 pto) Determinar la función de densidad o de probabilidad, según corresponda, de los tiempos de respuesta de S_1, S_2 y S_3.

b) (1 pto) Si el tiempo de respuesta a una petición de conexión ha superado los 6 segundos, determinar la probabilidad de que se haya hecho sobre el servidor S_1.

Ejercicio 3. El método utilizado por una compañía para la creación de un DVD consta de dos etapas independientes. El tiempo que se requiere para realizar la primera etapa es una variable aleatoria X que sigue una distribución normal de media 60 minutos y desviación típica 10 minutos. El tiempo que se requiere para realizar la segunda etapa es una variable aleatoria Y que sigue una distribución normal de media 40 minutos y desviación típica 5 minutos.

a) (0.75 pto) Determinar la probabilidad de que una y sólo una de las etapas requiera más de 45 minutos.

b) (0.75 pto) Calcular la probabilidad de que al crear un DVD se tarde más tiempo en completar la primera etapa que la segunda.

c) (0.75 pto) Si la compañía debe crear 20 DVD's distintos, determinar la probabilidad de que en, al menos 3, la segunda etapa tarde en completarse más de 45 minutos.

Ejercicio 4: La distribución de Rayleigh es adecuada para modelar la velocidad del viento cuando se conoce su valor medio. Si X sigue una distribución de Rayleigh de parámetro $\alpha > 0$ su función de densidad viene dada por

$$f(x) = \begin{cases} \frac{2x}{\alpha^2} e^{-x^2/\alpha^2} & \text{si} \quad x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Se sabe que $E(X) = \frac{\alpha\sqrt{\pi}}{2}$ y que $Var(X) = \frac{\alpha^2(4-\pi)}{4}$. Sea $X_1, \ldots, X_n$ una muestra aleatoria simple de X. Responder razonadamente a las siguientes cuestiones:

(a) (1 pto) Construir la función de probabilidad conjunta de la muestra. ¿Es suficiente para estimar $\frac{1}{\alpha}$, el estadístico definido por $T_1(X_1, \ldots, X_n) = \dfrac{n}{\sqrt{\sum_{i=1}^n X_i^2}}$?

(b) (0.75 pto) Sea $T_2 = \frac{2}{n\sqrt{\pi}} \sum_{i=1}^n X_i$. Demostrar que T_2 es un estimador consistente para α.

Ejercicio 5. Como parte de un estudio de calidad del servicio, se desea estudiar los tiempos de acceso a dos servidores ftp. Se recoge una muestra del servidor A, obteniendo los siguientes resultados (en ms.):

Tamaño muestral	media	desviación típica
11	73.82	5.15

Posteriormente, se recoge una muestra de tiempos de accesos al servidor B que arroja los siguientes resultados:

Tamaño muestral	media	desviación tpica
9	70.22	3.49

Supuesto que los tiempos de acceso a los servidores siguen distribuciones normales independientes, plantear y resolver contrastes de hipótesis adecuados para responder a las siguientes cuestiones (utilizar nivel de significación $\alpha = 0.05$):

(a) (0.75 pto) ¿ Existe evidencia significativa para afirmar que las varianzas de los tiempos de acceso son distintas?

(b) (1 pto) ¿Se puede afirmar que el tiempo medio de acceso al servidor A supera en más de 2 ms. al tiempo medio de acceso al servidor B?

Ejercicio 6. La siguiente tabla recoge el tiempo de respuesta (en nanosegundos) de un circuito lógico en frío (x) y el tiempo de respuesta (en nanosegundos) tras una hora de funcionamiento intensivo (y), para un conjunto de 8 máquinas:

x	6	5	8	13	7	4	5	9
y	7	8	11	15	10	6	9	12

(a) (0.75 pto) Ajustar a los datos un modelo del tipo $y = \beta_0 + \beta_1 x + \varepsilon$. Calcular e interpretar una medidad de la bondad del ajuste realizado.

(b) (1 pto) Estimar, mediante un intervalo de confianza al 95%, el tiempo esperado que tras una hora de funcionamiento intensivo tardará en responder un determinado circuito, si en frío tuvo un tiempo de respuesta de 10 nanosegundos.

DURACIÓN: HASTA LAS 12:50H.

Ejercicio 1:

Hay una diagonal por cada pareja de vértices, siempre que el par de vértices no determine un lado del polígono. Además, hay tantos lados como vértices. En consecuencia, el número de aristas del cuadrado viene dado por $C_{4,2}-2 = \binom{4}{2}-2 = 6-2 = 2$.

Para el exágono: $C_{6,2}-6 = \binom{6}{2}-6 = 15-6 = 9$.

Para un polígono de n lados: $C_{n,2}-n = \binom{n}{2}-n = \frac{n(n-1)}{2}-n = \frac{n^2-3n}{2}$

Si el número de lados es igual al número de diagonales, debe ser $\frac{n^2-3n}{2} = n \iff n^2-3n = 2n \iff n^2-5n=0$ → $n=0$ (solución no válida), $\underline{n=5}$

Por tanto el único polígono en el que el número de lados es igual al número de diagonales es el pentágono.

Ejercicio 2:

a) • Función de densidad del tiempo de respuesta de S_1:

$$f_1(x) = F_1'(x) = \begin{cases} \frac{1}{3}\cdot e^{-x/3} & \\ 0 & \text{si } x \leq 0 \end{cases}$$

que corresponde a una distribución Exp(1/3).

• Función de densidad del tiempo de respuesta de S_2:

De los datos que nos dan se deduce que la función de densidad es de la forma

$$f_2(x) = \begin{cases} c & \text{si } x \in (4,8) \\ 0 & \text{si } x \notin (4,8) \end{cases}$$

siendo $c>0$ un valor constante.

Entonces debe ser $1 = \int_{-\infty}^{+\infty} f_2(x)\,dx = \int_4^8 c\,dx = 4c \Rightarrow c = \frac{1}{4}$

Por tanto, $f_2(x) = \begin{cases} \frac{1}{4} & \text{si } x \in (4,8) \\ 0 & \text{si } x \notin (4,8) \end{cases}$ que corresponde a una distribución $U(4,8)$.

- Nos dicen que, para S_3, sólo hay tres posibles tiempos de respuesta. En consecuencia, X_3 es una variable aleatoria discreta. Además, los 3 posibles tiempos de respuesta son equiprobables por lo que la función de probabilidad de X_3 es:

$$P(X_3 = k) = \begin{cases} \frac{1}{3} & \text{si } k \in \{5, 6, 7\} \\ 0 & \text{en otro caso} \end{cases}$$

b) Si T es el tiempo que tarda en producirse una conexión y denotamos por S_i, $i=1,2,3$, al suceso que ocurre cuando la petición de conexión se realiza sobre el sistema S_i, tenemos que calcular

$$P(S_1|_{T>6}) = \frac{P(T>6 \cap S_1)}{P(T>6)} = \frac{P(T>6|_{S_1}) \cdot P(S_1)}{P(T>6|_{S_1}) \cdot P(S_1) + P(T>6|_{S_2}) \cdot P(S_2) + P(T>6|_{S_3}) \cdot P(S_3)}$$

De los datos del problema, $P(S_1) = 0'5$, $P(S_2) = 0'3$, $P(S_3) = 0'2$.

Además,
- $P(T>6|_{S_1}) = P(X_1 > 6) = 1 - P(X_1 \leq 6) = 1 - F_1(6) = 1 - (1 - e^{-6/3}) = e^{-2}$
- $P(T>6|_{S_2}) = P(X_2 > 6) = \int_6^{+\infty} f_2(x)\,dx = \int_6^8 \frac{1}{4}\,dx = \frac{1}{2}$
- $P(T>6|_{S_3}) = P(X_3 > 6) = P(X_3 = 7) = \frac{1}{3}$

Entonces,

$$P(S_1|_{T>6}) = \frac{e^{-2} \cdot 0'5}{e^{-2} \cdot 0'5 + \frac{1}{2} \cdot 0'3 + \frac{1}{3} \cdot 0'2} \simeq 0'238$$

Ejercicio 3:

X = tiempo necesario para completar la 1ª etapa

Y = " " " " " 2ª "

$X \sim N(60, 10^2)$, $Y \sim N(40, 5^2)$; X, Y son independientes

a) $\cdot\; P(X>45) = P\left(\frac{X-60}{\sqrt{100}} > \frac{45-60}{\sqrt{100}}\right) \overset{Z\sim N(0,1)}{=} P(Z>-1'5) = P(Z<1'5) = 0'9332$

$\cdot\; P(Y>45) = P\left(\frac{Y-40}{\sqrt{25}} > \frac{45-40}{\sqrt{25}}\right) = P(Z>1) = 1-P(Z\leq 1) = 1-0'8413 = 0'1587$

La probabilidad de que una y sólo una de las etapas requiera más de 45' será entonces:

$$P((X>45 \cap Y<45)\cup(X<45\cap Y>45)) = P(X>45\cap Y<45)+P(X<45\cap Y>45) =$$

$$\underset{X,Y \text{ independientes}}{=} P(X>45)\cdot P(Y<45) + P(X<45)\cdot P(Y>45) = 0'9332\cdot(1-0'1587) +$$

$$+ (1-0'9332)\cdot 0'1587 = 0'7957$$

b) Tenemos que calcular $P(X>Y) = P(X-Y>0)$. De los datos del problema se deduce que $X-Y\sim N(20,125)$. Entonces,

$$P(X>Y) = P(X-Y>0) = P\left(\frac{X-Y-20}{\sqrt{125}} > \frac{0-20}{\sqrt{125}}\right) = P(Z>-1'79) =$$

$Z\sim N(0,1)$

$$= P(Z<1'79) = 0'9633$$

---

c) Denotaremos por W a la variable aleatoria definida como:

W = nº de DVD's de los 20 creados en los que la segunda etapa tarda más de 45'.

Entonces $W\sim B(20, 0'1587)$. Tenemos que calcular

$$P(W\geq 3) = 1-P(W<3) = 1-P(W=0)-P(W=1)-P(W=2) =$$

$$= 1-\binom{20}{0}0'1587^0\cdot 0'8413^{20} - \binom{20}{1}0'1587^1\cdot 0'8413^{19} - \binom{20}{2}0'1587^2\cdot 0'8413^{18} \simeq$$

$$\simeq 1-0'3639 = 0'6361$$

Ejercicio 4:

a) Función de densidad conjunta:

$$f(x_1,\dots,x_n) = \prod_{i=1}^{4} f_{X_i}(x_i) = \prod_{i=1}^{4}\frac{2x_i}{\alpha^2}\cdot e^{-\frac{x_i^2}{\alpha^2}} = \left(\prod_{i=1}^{4}x_i\right)\cdot\frac{2^4}{\alpha^{2n}}\cdot e^{-\frac{\sum x_i^2}{\alpha^2}}$$

Por otra parte, $T_1 = \dfrac{n}{\sqrt{\sum_{i=1}^{n} X_i^2}} \Rightarrow \sum_{i=1}^{n} x_i^2 = \dfrac{n^2}{T_1^2}$. Sustituyendo en la función de densidad conjunta de la muestra,

$$f(x_1,\dots,x_n) = \underbrace{\left(\prod_{i=1}^{n} x_i\right)}_{h(x_1,\dots,x_n)} \cdot \underbrace{\frac{2^n}{\alpha^{2n}} \cdot e^{-\frac{n^2}{T_1^2}\cdot\frac{1}{\alpha^2}}}_{= g\left(T_1, \frac{1}{\alpha}\right)}$$

Por tanto, T_1 es un estadístico suficiente para $\frac{1}{\alpha}$.

b) Vamos a comprobar que (i) $\lim_{n\to\infty} E(T_2) = \alpha$ y (ii) $\lim_{n\to\infty} Var(T_2) = 0$.

$$E(T_2) = E\left(\frac{2}{n\sqrt{\pi}} \cdot \sum_{i=1}^{n} X_i\right) = \frac{2}{n\sqrt{\pi}} \cdot \sum_{i=1}^{n} E(X_i) \underset{\substack{X_1,\dots,X_n \text{ es} \\ \text{m.a.s. de } X}}{=} \frac{2}{n\sqrt{\pi}} \cdot \sum_{i=1}^{n} E(X) = \frac{2}{n\sqrt{\pi}} \cdot \sum_{i=1}^{n} \frac{\alpha\sqrt{\pi}}{2} =$$

$$= \frac{2}{n\sqrt{\pi}} \cdot n \cdot \frac{\alpha\sqrt{\pi}}{2} = \alpha.$$

Por tanto, $\lim_{n\to\infty} E(T_2) = \lim_{n\to\infty} \alpha = \alpha$ y se cumple (i).

$$Var(T_2) = Var\left(\frac{2}{n\sqrt{\pi}} \cdot \sum_{i=1}^{n} X_i\right) \underset{X_1,\dots,X_n \text{ independientes}}{=} \frac{4}{n^2 \cdot \pi} \cdot \sum_{i=1}^{n} Var(X_i) \underset{\substack{X_1,\dots,X_n \text{ es} \\ \text{m.a.s. de } X}}{=} \frac{4}{n^2\pi} \cdot \sum_{i=1}^{n} Var(X) =$$

$$= \frac{4}{n^2\pi} \cdot \sum_{i=1}^{n} \frac{\alpha^2(4-\pi)}{4} = \frac{4}{n^2 \cdot \pi} \cdot n \cdot \frac{\alpha^2(4-\pi)}{4} = \frac{\alpha^2(4-\pi)}{n\pi}$$

Entonces $\lim_{n\to\infty} Var(T_2) = \lim_{n\to\infty} \dfrac{\alpha^2(4-\pi)}{n\pi} = 0$, luego se cumple (ii).

Como se cumplen (i) y (ii), T_2 es un estimador consistente de α.

Ejercicio 5

Consideramos las variables X_A, X_B : tiempos de acceso a los servidores A y B, respectivamente. $X_A \sim N(\mu_A, \sigma_A^2)$ y $X_B \sim N(\mu_B, \sigma_B^2)$

a) Tenemos que resolver el contraste $\begin{cases} H_0 : \sigma_A^2 = \sigma_B^2 \\ H_1 : \sigma_A^2 \neq \sigma_B^2 \end{cases}$ con $\alpha = 0'05$

Se rechazará la hipótesis nula si $\dfrac{S_{c_A}^2}{S_{c_B}^2} \geq f_{n_A-1, n_B-1, 1-\alpha/2}$

o bien $\dfrac{S_{c_A}^2}{S_{c_B}^2} \leq \dfrac{1}{f_{n_B-1, n_A-1, 1-\alpha/2}}$

$S_{c_A}^2 = \dfrac{n_A \cdot S_A^2}{(n_A-1)} = \dfrac{11 \cdot 5'15^2}{10} = 29'17475$

$S_{c_B}^2 = \dfrac{n_B \cdot S_B^2}{(n_B-1)} = \dfrac{9 \cdot 3'49^2}{8} = 13'7026125$

De las tablas de la distribución F-Snedecor:

$f_{n_A-1, n_B-1, 1-\alpha/2} = f_{10,8,0'975} = 4'295$

$f_{n_B-1, n_A-1, 1-\alpha/2} = f_{8,10,0'975} = 3'855$

Entonces se rechazará Ho si $\dfrac{S_{c_A}^2}{S_{c_B}^2} \geq 4'295$ o bien $\dfrac{S_{c_A}^2}{S_{c_B}^2} \leq \dfrac{1}{3'855} = 0'259403$

Puesto que $\dfrac{S_{c_A}^2}{S_{c_B}^2} = \dfrac{29'17475}{13'7026125} \simeq 2'1291$, no se verifica la regla de rechazo, y no tenemos evidencia para afirmar que las varianzas son distintas.

b) Tenemos que resolver el contraste $\begin{cases} H_0: \mu_A = \mu_B + 2 \\ H_1: \mu_A > \mu_B + 2 \end{cases}$

o lo que es igual, $\begin{cases} H_0: \mu_A - \mu_B = 2 \\ H_1: \mu_A - \mu_B > 2 \end{cases}$ con $\alpha = 0'05$. Teniendo en cuenta el resultado obtenido en el apartado anterior, se trata de un contraste sobre la diferencia de medias de dos poblaciones normales con varianzas desconocidas pero iguales.

Se rechazará H_0 si $\bar{X}_A - \bar{X}_B \geq \delta_0 + S_p \cdot \sqrt{\frac{1}{n_A} + \frac{1}{n_B}} \cdot t_{n_A+n_B-2,1-\alpha}$, donde

$$S_p = \sqrt{\frac{n_A \cdot S_A^2 + n_B \cdot S_B^2}{n_A + n_B - 2}} = \sqrt{\frac{11 \cdot 5'15^2 + 9 \cdot 3'49^2}{18}} = 4'7221$$

y $t_{n_A+n_B-2,1-\alpha} = t_{18,0'95} = 1'734$

Por tanto, se rechazaría H_0 si

$$\bar{X}_A - \bar{X}_B \geq 2 + 4'7221 \cdot \sqrt{\frac{1}{11} + \frac{1}{9}} \cdot 1'734 = 5'68029$$

Puesto que $\bar{X}_A - \bar{X}_B = 73'82 - 70'22 = 3'60$, no se cumple la regla de rechazo y no podemos afirmar que el tiempo medio de acceso al servidor A supere en más de 2 ms. al tiempo medio de acceso al servidor B.

Ejercicio 6

a) De los datos del problema se deduce que

$\bar{x} = 7'125 \quad S_x^2 = 7'359375$

$\bar{y} = 9'75 \quad S_y^2 = 7'4375$

$$S_{xy} = \frac{\sum_{i=1}^{n} x_i y_i}{n} - \bar{x}\cdot\bar{y} = 76'5 - 7'125 \cdot 9'75 = 7'03125$$

Puesto que la recta de regresión viene dada por

$y - \bar{y} = \frac{S_{xy}}{S_x^2}(x - \bar{x})$ sustituyendo se obtiene la recta

$$\boxed{y = 0'955414 \cdot x + 2'942675}$$

Como medida de la bondad del ajuste realizado calculamos el coeficiente de determinación R^2. Por ser un modelo lineal simple podemos calcularlo como $R^2 = \frac{S_{xy}^2}{S_x^2 \cdot S_y^2} = \frac{7'03125^2}{7'359375 \cdot 7'4375} \simeq 0'9032$

Podemos decir, por tanto, que el modelo explica el 90'32% de la variabilidad del tiempo de respuesta tras una hora de funcionamiento intensivo.

b) El intervalo solicitado viene dado por

$$\hat{y}(x_0) \pm S \cdot \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{n \cdot S_x^2}} \cdot t_{n-2,1-\alpha/2}$$

Con $x_0 = 10$ e $\underline{\hat{y}(x_0)} = \hat{y}(10) = 0'955414 \cdot 10 + 2'9426752 = \underline{12'4968152}$

Puesto que se pide un I.C. al 95%, $\alpha = 0'05$ y $t_{n-2,1-\alpha/2} = t_{6,0'975} = 2'447$

Finalmente, calculamos los cuadrados medios del error:

$$S^2 = \frac{SS_{Tot} - SS_{Reg}}{n-2} = \frac{59'5 - 53'7420367}{8-2} = 0'9596605$$

modelo lineal simple

$$SS_{Tot} = n \cdot S_y^2 = 8 \cdot 7'4375 = 59'5$$

$$SS_{Reg} = n \cdot b_1^2 \cdot S_x^2 = 8 \cdot 0'9554142^2 \cdot 7'359375 = 53'7420367$$

Finalmente, sustituyendo en la expresión del intervalo obtenemos:

$$12'4968152 \pm \sqrt{0'9596605} \cdot \sqrt{\frac{1}{8} + \frac{(10-7'125)^2}{8 \cdot 7'359375}} \cdot 2'447$$

$$12'4968152 \pm 1'23491576 \longrightarrow \boxed{(11'26189944,\ 13'73173096)}$$