Universidad de la República
Facultad de Ingeniería - IMERL

Probabilidad y Estadística
Segundo Semestre 2016

PRIMER PARCIAL
SÁBADO 26 DE NOVIEMBRE DE 2016.

Número de Parcial	Cédula	Nombre y Apellido

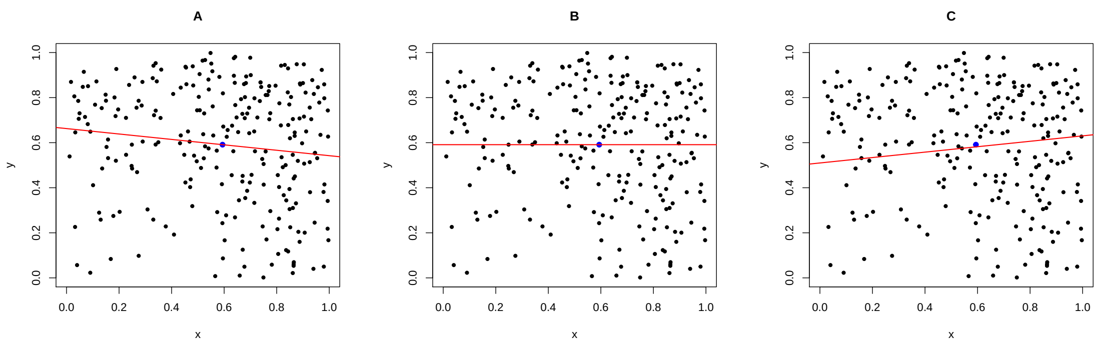
PARA USO DOCENTE			
Ej. 1	Ej. 2	Ej.3	TOTAL

Ejercicio 1. [**20 puntos**]

1. Sea (X, Y) un vector aleatorio en $\mathbb{R}^2$ con densidad $f_{(X,Y)}$. Probar que la densidad marginal de X es $f_X(x) = \int\limits_{\mathbb{R}} f_{(X,Y)}(x, y)dy$.
2. Sea (X, Y) un vector aleatorio en $\mathbb{R}^2$ con densidad $f_{(X,Y)}$ dada por:

$$f_{(X,Y)}(x,y) = \begin{cases} x+y & \text{si } x \in [0,1], y \in [0,1], \\ 0 & \text{en otro caso.} \end{cases}$$

   (a) Hallar las densidades marginales de X e Y. Verificar que son densidades.
   (b) Hallar la probabilidad de que $X + Y$ sea menor o igual que $\frac{1}{2}$.
   (c) i. ¿Son X e Y independientes?
       ii. Hallar Cov(X, Y) la covarianza entre X e Y y el coeficiente de correlación r. Indicar si la correlación es leve, moderada o fuerte.
   (d) Se considera una muestra $\{(x_i, y_i)\}_{i=1,\ldots,n}$ con densidad conjunta $f_{(X,Y)}$ definida en la parte 2. A continuación se presentan tres posibles diagramas de dispersión con la recta de regresión asociada. Indica cuál diagrama se ajusta mejor a la densidad dada. Justifique su respuesta.



Solución Ejercicio 1.

1.

$$F_X(z) = P(X \leq z) = P((X,Y) \in (-\infty, z] \times \mathbb{R})$$
$$= \int_{\mathbb{R}} \int_0^z f_{(X,Y)}(x,y)dxdy = \int_0^z \left(\int_{\mathbb{R}} f_{(X,Y)}(x,y)dy \right) dx$$

y por definición de densidad resulta entonces que $f_X(x) = \int\limits_{\mathbb{R}} f_{(X,Y)}(x,y)dy$.

2. (a) Utilizando la parte anterior se tiene que, si:

$$f_X(x) = \int\limits_{\mathbb{R}} f_{(X,Y)}(x,y)dy = \begin{cases} 0 & \text{si } x \notin [0,1] \\ \int\limits_0^1 (x+y)dy = x + \frac{1}{2} & \text{si } x \in [0,1]. \end{cases}$$

Dada la simetría de la densidad conjunta, es fácil ver que X e Y tienen la misma densidad, es decir:

$$f_Y(y) = \int\limits_{\mathbb{R}} f_{(X,Y)}(x,y)dx = \begin{cases} 0 & \text{si } y \notin [0,1] \\ \int\limits_0^1 (x+y)dx = y + \frac{1}{2} & \text{si } y \in [0,1]. \end{cases}$$

Para verificar que la función hallada es densidad, dado que es positiva, basta verificar que $\int\limits_0^1 x + \frac{1}{2}dx = 1$.

(b) Sea $\mathcal{R} = \left\{(x,y) \in \mathbb{R}^2 : x + y \leq \frac{1}{2}\right\}$. Entonces:

$$P(X+Y \leq \frac{1}{2}) = P((X,Y) \in \mathcal{R}) = \int \int_{\mathcal{R}} f_{(X,Y)}(x,y)dy = \int_0^{1/2} \int_0^{1/2-x} x + ydydx$$
$$\int_0^{1/2} dx \int_0^{1/2-x} x + ydy = \int_0^{1/2} -\frac{x^2}{2} + \frac{1}{8}dx = \frac{1}{24}.$$

(c) i. X e Y no son independientes pues $f_{(X,Y)}(x,y) = x + y \neq (x+1/2)(y+1/2) = f_X(x)f_Y(y)\ \forall\, x,y \in [0,1]$.

ii. La covarianza entre X e Y se define como $\mathrm{Cov}(X,Y) = \mathrm{E}\,(XY) - \mathrm{E}\,(X)\,\mathrm{E}\,(Y)$ donde:

$$\mathrm{E}\,(X) = \int_{\mathbb{R}} x f_X(x)dx = \int_0^1 x(x + \frac{1}{2})dx = \frac{7}{12}.$$

Dado que X e Y tienen la misma distribución, se tiene que $\mathrm{E}\,(Y) = \mathrm{E}\,(X) = \frac{7}{12}$.

$$\mathrm{E}\,(XY) = \int_{\mathbb{R}} \int_{\mathbb{R}} xy f_{(X,Y)}(x,y)dxdy = \int_0^1 \int_0^1 xy(x+y)dx = \frac{1}{3}.$$

Por lo tanto $\mathrm{Cov}(X,Y) = \mathrm{E}\,(XY) - \mathrm{E}\,(X)\,\mathrm{E}\,(Y) = \frac{1}{3} - (\frac{7}{12})^2 = -\frac{1}{12^2}$.
El coeficiente de correlación r se calcula como

$$r = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}\,(X)}\sqrt{\mathrm{Var}\,(Y)}}$$

Por lo tanto, necesitamos calcular $\mathrm{Var}(X) = \mathrm{Var}(Y) = \mathrm{E}\left(Y^2\right) - \mathrm{E}(Y)^2 = \mathrm{E}\left(Y^2\right) - (\frac{7}{12})^2$:

$$\mathrm{E}\left(Y^2\right) = \int_{\mathbb{R}} x^2 f_X(x)dx = \int_0^1 x^2(x+1/2)dx = \frac{5}{12}$$

Por lo tanto $\mathrm{Var}(X) = \mathrm{Var}(Y) = \frac{11}{12^2}$, de donde $r = -\frac{1}{11} = -0.091$. La correlación es entonces leve.

(d) La recta de regresión tiene pendiente $\alpha = r\frac{s_Y}{s_X} = r$ pues ya vimos que las variables X e Y tienen la misma varianza.

Dado que el coeficiente de correlación es negativo, la opción C que muestra una recta de regresión con pendiente positiva queda descartada. Además la opción B muestra una recta horizontal, esto es una pendiente igual a cero, y por lo tanto también queda descartada. Por lo tanto, la opción que más se aproxima a la densidad conjunta dada es la opción A.

Ejercicio 2. [**16 puntos**]

1. Sean $X_1, X_2, \ldots, X_k$ variables aleatorias independientes con distribución exponencial de parámetro λ. Se define $T = \min\{X_1, X_2, \ldots, X_k\}$. Calcular valor esperado, mediana y varianza de T.

2. Un reconocido magnate inmobiliario de iniciales DT, desea testear a su proveedor de lamparitas. Para hacerlo encarga a su gerente de operaciones la tarea de estimar el tiempo medio de vida de las lamparitas que se utilizan en sus edificios. Puede asumirse que un edificio tiene 1000 lamparitas. Sin embargo, esperar a la rotura de las 1000 lamparitas para obtener su promedio no es viable pues llevaría demasiado tiempo. El gerente, conociendo la impaciencia de su jefe, decide cambiar la estrategia y considerar 100 edificios nuevos (de modo de que todas las lamparitas comienzan a funcionar todas al mismo tiempo) y registra el tiempo en que se rompe la primer lamparita de cada edificio. Se asume que el tiempo de vida de cada lamparita es una variable aleatoria exponencial de parámetro λ y que son independientes entre sí. Sea T_i el tiempo, medido en meses, en que se rompe la **primer** lamparita en el edificio i. Se tiene por tanto una muestra $T_1, T_2, \ldots, T_{100}$ de la cuál se presenta a continuación un resumen numérico:

m	$Q1$	m_T	Q_3	M
0.00006	0.026	0.061	0.122	0.890

donde $\mathrm{m} = \min_i T_i$, M $= \max_i T_i$, m_T es la mediana empírica y Q_1 y Q_3 son el primer y tercer cuartil de la muestra respectivamente. Además:

s_T^2	$\overline{T}$	σ_T^2
0.0083	0.0897	0.0081

A partir de la muestra anterior:

(a) Calcular el estimador de máxima verosimilitud de λ.

(b) Hallar dos estimadores diferentes al calculado en la parte anterior para el parámetro λ.

(c) Hallar un intervalo de confianza aproximado al nivel 0.95 para el parámetro λ.

Solución Ejercicio 2

1. La variable aleatoria T tiene distribución dada por

$$F_T(x) = 1 - \prod_{i=1}^{k}(1 - F_{X_i}(x)) = 1 - (e^{-\lambda x})^k = 1 - e^{-k\lambda x} \quad \forall\, x \geq 0.$$

   Por lo tanto T tiene distribución Exponencial de parámetro $k\lambda$.

   Por lo tanto, se tiene que el valor esperado de T es $\mathrm{E}(T) = \frac{1}{k\lambda}$, la mediana es $m_T = \frac{log(2)}{k\lambda}$ y la varianza es $\mathrm{Var}(T) = \frac{1}{(k\lambda)^2}$.

2. Utilizando la parte anterior con $k = 1000$, resulta que T_i es exponencial de parámetro $1000\times\lambda$, $\forall\, i = 1, \ldots, 100$. Por lo tanto $f_{T_i}(x) = 1000\lambda e^{-1000\lambda x}$ $\forall\, x \geq 0$.

   (a) La función de verosimilitud está dada por:

$$L(\lambda) = \prod_{i=1}^{100} f_T(T_i) = \prod_{i=1}^{100} 1000\lambda e^{-1000\lambda T_i} = (1000\lambda)^{100} e^{-1000\lambda \sum_{i=1}^{100} T_i}$$

   Tomando logaritmo, resulta que:

$$\log(L(\lambda)) = 100 \times \log(1000\lambda) - 1000\lambda \sum_{i=1}^{100} T_i.$$

   Para buscar el valor de λ que maximiza la verosimilitud, derivamos la expresión respecto a λ y obtenemos:

$$\frac{\partial \log(L(\lambda))}{\partial\lambda} = 100 \times \frac{1}{\lambda} - 1000 \times \sum_{i=1}^{100} T_i = 0 \quad \text{sii} \quad \lambda = \frac{100}{1000 \sum_{i=1}^{100} T_i} = \frac{1}{1000\overline{T}}.$$

   Por lo tanto el estimador de máxima verosimilitud $\hat{\lambda}$ de λ es $\hat{\lambda} = \frac{1}{1000\overline{T}}$.

   (b) El estimador por momentos $\hat{\lambda}_1$ en este caso coincide con el estimador de máxima verosimilitud ya que la ley de los grandes números nos asegura que el promedio $\overline{T}$ es un estimador del valor esperado $\mathrm{E}(T) = \frac{1}{1000\lambda}$. Por lo tanto:

$$\hat{\lambda}_1 = \frac{1}{1000\overline{T}_{100}} = \frac{1}{1000 \times 0.0897} = 0.0111.$$

   Otro posible estimador para λ se obtiene de utilizar que s_X^2 es un estimador de la varianza de T, donde $\mathrm{Var}(T) = (\frac{1}{1000\lambda})^2$. Por lo tanto $\hat{\lambda}_2 = \frac{1}{1000 s_X} = \frac{1}{1000\sqrt{0.0083}} = 0.0109$ es también un estimador de λ.

   Finalmente, es posible obtener un tercer estimador $\hat{\lambda}_3$ a partir de la mediana empírica ya que la mediana teórica es $\frac{\log(2)}{1000\lambda}$ de donde

$$\hat{\lambda}_3 = \frac{\log(2)}{1000 m_T} = \frac{\log(2)}{1000 \times 0.061} = 0.0113.$$

   (c) Un intervalo de confianza aproximado al nivel 0.95 para el valor esperado $\mathrm{E}(T) = \frac{1}{1000\lambda}$ es:

$$I = \left[\overline{T}_{100} - \frac{z_{0.025}\sigma_T}{\sqrt{100}}, \overline{T}_{100} + \frac{z_{0.025}\sigma_T}{\sqrt{100}}\right] = [0.072, 0.107].$$

Usamos que $z_{0.025} = 1.96$ y los datos proporcionados. Es posible calcular el intervalo utilizando s_T en lugr de σ_T.
Sin embargo, este no es un intervalo de confianza para λ sino para que $\frac{1}{1000\lambda}$. Por lo tanto, el intervalo de confianza para λ resulta ser:

$$I = \left[\frac{1}{1000 \times 0.107}, \frac{1}{1000 \times 0.072}\right] = [0.0093, 0.0138].$$

Ejercicio 3. [**24 puntos**]

Sea $X_1, X_2, \ldots, X_n$ una muestra independiente e idénticamente distribuida, con distribución Normal con parámetros μ desconocido y $\sigma^2 = 4$, es decir $X_i \sim \mathcal{N}(\mu, 4)$.

1. Se define Z_1 tal que $2Z_1 = X_1 - \mu$. Hallar la función de distribución de Z_1. ¿Reconoce alguna distribución conocida?
2. Hallar el mínimo valor para el tamaño n de la muestra, tal que la probabilidad de que el promedio $\overline{X}_n$ difiera de μ en menos de 0.1, sea mayor que 0.9.
3. (a) Dado $\alpha < 0.5$, hallar ϵ en función de n y α tal que $P(\mu \in [\overline{X}_n, \overline{X}_n + \epsilon]) = \frac{1}{2} - \alpha$.
   (b) Hallar ϵ para $\alpha = 0.05$ y $n = 16$.
4. Se consideran los siguientes datos que corresponden a la muestra anterior:

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}
4.97	1.6	−0.84	0.55	−0.95	−5.14	2.34	1.51	0.35	−3.6	0.01	0.38	−0.34	−0.78	1.71	2.2

Se plantea el siguiente test de hipótesis sobre el valor esperado:

$$\begin{cases} H_0 : \mu = \mu_0 = 0 \\ H_1 : \mu = \mu_1 > 0. \end{cases}$$

(a) Probar que $\mathcal{R}_\alpha = \{\overline{X}_n \geq \frac{2z_\alpha}{\sqrt{n}}\}$ es una región crítica para el test planteado.
(b) ¿Cuál es la decisión para $\alpha = 0.05$, $\alpha = 0.1$ y $\alpha = 0.5$?
(c) Indicar cuál de las siguientes afirmaciones sobre el p-valor α^* es verdadera. Justifque su respuesta:

i. $\alpha^* < 0.1$ ii. $\alpha^* \in [0.1, 0.5]$ iii. $\alpha^* > 0.5$

(d) Asumiendo $\mu_1 = 0.5$ y $\alpha = 0.05$, hallar β la probabilidad de error tipo II.

Solución Ejercicio 3

1.

$$F_{Z_1}(z) = \mathbf{P}(Z_1 \leq z) = \mathbf{P}(2Z_1 \leq 2z) = \mathbf{P}(X_1 - \mu \leq 2z) = \mathbf{P}(X_1 \leq \mu + 2z)$$
$$= \int_0^{\mu+2z} \frac{1}{2\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{2})^2} dx$$

Haciendo el cambio de variable $u = \frac{x-\mu}{2}$, resulta que:

$$F_{Z_1}(z) = \int_0^z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du$$

de donde se tiene que Z_1 tiene distribuciíon Normal de parámetros $\mu = 0$ y $\sigma = 1$.

2. Se pide hallar el mínimo n tal que $\mathbf{P}\left(|\overline{X}_n - \mu| < 0.1\right) > 0.9$:

$$\mathbf{P}\left(|\overline{X}_n - \mu| < 0.1\right) = \mathbf{P}\left(-0.1 \leq \overline{X}_n - \mu \leq 0.1\right) = \mathbf{P}\left(-\frac{0.1\sqrt{n}}{2} \leq \sqrt{n}\frac{(\overline{X}_n - \mu)}{2} \leq \frac{0.1\sqrt{n}}{2}\right)$$
$$= 2\Phi\left(\frac{0.1\sqrt{n}}{2}\right) - 1 > 0.9 \quad \text{sii} \quad \Phi\left(\frac{0.1\sqrt{n}}{2}\right) > 0.95 \quad \text{sii} \quad \frac{0.1\sqrt{n}}{2} > z_{0.05} = 1.65$$

Por lo tanto, tenemos que n debe ser tal que $n > (\frac{2\times 1.65}{0.1})^2 = 1089$. Por lo tanto el valor mńimo de $n = 1090$.

3. (a) Se busca ϵ tal que $\mathbf{P}\left(\mu \in [\overline{X}_n, \overline{X}_n + \epsilon]\right) = \frac{1}{2} - \alpha$, donde:

$$\mathbf{P}\left(\mu \in [\overline{X}_n, \overline{X}_n + \epsilon]\right) = \mathbf{P}\left(-\epsilon \leq \overline{X}_n - \mu \leq 0\right) = \mathbf{P}\left(-\frac{\sqrt{n}\epsilon}{2} \leq \sqrt{n}\frac{(\overline{X}_n - \mu)}{2} \leq 0\right)$$
$$= \Phi(0) - \Phi\left(-\frac{\sqrt{n}\epsilon}{2}\right) = \frac{1}{2} - (1 - \Phi\left(\frac{\sqrt{n}\epsilon}{2}\right)) = \Phi\left(\frac{\sqrt{n}\epsilon}{2}\right) - \frac{1}{2}$$

Por lo tanto $\mathbf{P}\left(\mu \in [\overline{X}_n, \overline{X}_n + \epsilon]\right) = \frac{1}{2} - \alpha$ sii $\Phi\left(\frac{\sqrt{n}\epsilon}{2}\right) = 1 - \alpha$ sii $\frac{\sqrt{n}\epsilon}{2} = z_\alpha$ sii $\epsilon = \frac{2z_\alpha}{\sqrt{n}}$.

(b) Si $\alpha = 0.05$ entonces $z_\alpha = 1.65$ y para $n = 16$ resulta que $\epsilon = \frac{2\times 1.65}{4} = 0.825$.

4. (a) $\mathcal{R}_\alpha$ es una región crítica para el test planteado si $P_{H_0}(\mathcal{R}_\alpha) = \alpha$:

$$P_{H_0}(\mathcal{R}_\alpha) = P_{H_0}\left(\overline{X}_n \geq \frac{2z_\alpha}{\sqrt{n}}\right) = P_{H_0}\left(\sqrt{n}\frac{\overline{X}_n}{2} \geq z_\alpha\right) = 1 - \Phi(z_\alpha) = \alpha,$$

pues bajo H_0, $\mu = 0$ y por lo tanto $\overline{X}_n \sim \mathcal{N}(0, \frac{4}{n})$.

(b) En este caso, tenemos que $\overline{X}_n = 0.248$, por lo tanto la región crítica puede escribirse como $\mathcal{R}_\alpha = \{2\overline{X}_n \geq z_\alpha\} = \{0.496 \geq z_\alpha\}$

- Si $\alpha = 0.05$, entonces $z_{0.05} = 1.65$, por lo tanto el estadístico no pertenece a $\mathcal{R}_{0.05}$ y no se rechaza H_0.
- Si $\alpha = 0.1$, entonces $z_{0.05} = 1.29$, por lo tanto el estadístico no pertenece a $\mathcal{R}_{0.1}$ y no se rechaza H_0.
- Si $\alpha = 0.5$, entonces $z_{0.5} = 0.5$, por lo tanto el estadístico pertenece a $\mathcal{R}_{0.5}$ y se rechaza H_0.

(c) De acuerdo a la parte anterior, la opción correcta es la opción ii., pues tenemos que para $\alpha = 0.1$ no se rechaza H_0 y para $\alpha = 0.5$ se rechaza H_0. Por lo tanto el p-valor α^* es algún valor intermedio.

(d) Para $\alpha = 0.05$ y $n = 16$ la región crítica es $\mathcal{R}_{0.05} = \{2\overline{X}_n \geq 1.65\}$. Por lo tanto,

$$\beta = P_{H_1}(\mathcal{R}^c_{0.05}) = P_{H_1}(2\overline{X}_n < 1.65) = P_{H_1}(\overline{X}_n < 0.825) =$$
$$P_{H_1}(2(\overline{X}_n - 0.5) < 2(0.825 - 0.5)) = \Phi(0.65) = 0.7422.$$

Donde en el último paso usamos que bajo H_1 la verdadera media es $\mu_1 = 0.5$.