**Universidad de la República** **Probabilidad y Estadística**
**Facultad de Ingeniería - IMERL** **Curso 2do semestre 2017**

| Número de Parcial | Cédula | Nombre y Apellido |
|---|---|---|
| | | |

# Segundo Parcial – 25 de Noviembre 2017

*El parcial dura 3 horas y se puede utilizar material (no digital).*

**Ejercicio 1 (20 puntos)** Sea $a > 0$ y $X \sim U[-a, a]$. Definimos la variable aleatoria $Y = X^2 - a^2$.

*a*) Halle la densidad de $Y$.

*b*) Halle la esperanza de $Y$.

*c*) Sea $Y_1, \ldots, Y_n$ una muestra aleatoria simple con la misma distibución que la v.a. $Y$.

1) Halle un estimador de $a$ por el método de los momentos.
2) Halle un estimador de $a$ por el método de máxima verosimilitud.

**Ejercicio 2 (10 puntos)**

Pruebe que si $X_1 \sim \mathbf{N}(\mu_1, \sigma_1^2)$, $X_2 \sim \mathbf{N}(\mu_2, \sigma_2^2), \ldots, X_n \sim \mathbf{N}(\mu_n, \sigma_n^2)$ y son independientes entonces

$$X_1 + X_2 + ... + X_n \sim \mathbf{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

Se sugiere utilizar la función generatriz de momentos.

**Ejercicio 3 (20 puntos)** Se considera una muestra proveniente de una variable aleatoria:

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $x_i^*$ | 2.0 | 2.2 | 2.7 | 3.0 | 3.1 | 3.8 | 4.7 | 4.9 | 5.2 | 5.3 |

*a*) Suponiendo que esta muestra proviene de una variable aleatoria con distribución normal, halle un intervalo de confianza al 95 % para $\mu$.

*b*) Realice un test de hipótesis para decidir si se puede afirmar que estos datos provienen de una v.a. $\mathbf{N}(\mu = 4, \sigma^2 = 1)$. Decidir a nivel $\alpha = 0{,}05$.
*Sugerencia: complete la siguiente tabla para determinar cuál es el estadístico.*

| $i$ | $x_i^*$ | $F_0(x_i^*)$ | $\lvert\frac{i}{n} - F_0(x_i^*)\rvert$ | $\lvert\frac{i-1}{n} - F_0(x_i^*)\rvert$ |
|---|---|---|---|---|
| 1 | 2.0 | | | |
| 2 | 2.2 | | | |
| 3 | 2.7 | | | |
| 4 | 3.0 | | | |
| 5 | 3.1 | | | |
| 6 | 3.8 | | | |
| 7 | 4.7 | | | |
| 8 | 4.9 | | | |
| 9 | 5.2 | | | |
| 10 | 5.3 | | | |

*c*) Se considera otra muestra iid de otra variable aleatoria $Y$:

| $i$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $y_i^*$ | 4.5 | 7.9 | 8.5 | 9.4 | 11.3 |

Realice un *Test de Kolmogorov Smirnov para dos muestras* para decidir si es posible afirmar que estas dos muestras provienen de la misma distribución. Decidir a nivel $\alpha = 0{,}05$.

*Sugerencia: Complete las siguientes tablas para determinar cuál es el estadístico en este caso.*

| $i$ | $x_i^*$ | $F_m^Y(x_i^*)$ | $\lvert\frac{i}{n} - F_m^Y(x_i^*)\rvert$ |
|---|---|---|---|
| 1 | 2.0 | | |
| 2 | 2.2 | | |
| 3 | 2.7 | | |
| 4 | 3.0 | | |
| 5 | 3.1 | | |
| 6 | 3.8 | | |
| 7 | 4.7 | | |
| 8 | 4.9 | | |
| 9 | 5.2 | | |
| 10 | 5.3 | | |

| $j$ | $y_j^*$ | $F_n^X(y_j^*)$ | $\lvert\frac{j}{m} - F_n^X(y_j^*)\rvert$ |
|---|---|---|---|
| 1 | 4.5 | | |
| 2 | 7.9 | | |
| 3 | 8.5 | | |
| 4 | 9.4 | | |
| 5 | 11.3 | | |

**Ejercicio 4 (10 puntos)** Cinco niños de $2, 3, 5, 7$ y $8$ años de edad pesan, respectivamente, $14, 20, 32, 42$ y $44$ kilos.

*a*) Halle la ecuación de la recta de regresión de la edad sobre el peso.

*b*) ¿Cuál sería el peso aproximado de un niño de seis años?

SOLUCIÓN

**Ejercicio 1** Sean $a > 0$, $X \sim U[-a, a]$. y $Y = X^2 - a^2$.

*a)*

$F_Y(y) = \begin{cases} F_X(\sqrt{y+a^2}) - F_X(-\sqrt{y+a^2}) \text{ si } y \geq -a^2 \\ 0 \text{ si } y < a^2 \end{cases}$

Como $F_X(x) = \begin{cases} 0 \text{ si } x < -a \\ \frac{x+a}{2a} \text{ si } -a \leq x \leq a \\ 1 \text{ si } x > a \end{cases}$, observar que:

- $\sqrt{y+a^2} < a \iff y < 0$
- $-\sqrt{y+a^2} > -a \iff y < 0$

Entonces:

- Si $-a^2 < y < 0$, $F_X(\sqrt{y+a^2}) = \frac{\sqrt{y+a^2}+a}{2a}$ y $F_X(-\sqrt{y+a^2}) = \frac{-\sqrt{y+a^2}+a}{2a}$
- Si $y > 0$, entonces $\sqrt{y+a^2} > a$ y $-\sqrt{y+a^2} < -a$, por lo que $F_X(\sqrt{y+a^2}) = 1$ y $F_X(-\sqrt{y+a^2}) = 0$

Finalmente, $F_Y(y) = \begin{cases} 0 \text{ si } y < -a^2 \\ \frac{\sqrt{y+a^2}}{a} \text{ si } -a^2 < y < 0 \\ 1 \text{ si } y \geq 0 \end{cases}$ y $f_Y(y) = \begin{cases} 0 \text{ si } y < -a^2 \\ \frac{1}{2a\sqrt{y+a^2}} \text{ si } -a^2 < y < 0 \\ 0 \text{ si } y \geq 0 \end{cases}$

*b)* $\mathbb{E}(Y) = \mathbb{E}(X^2 - a^2) = \frac{a^2}{3} - a^2 = -\frac{2}{3}a^2$.

*c)* Sea $Y_1, \ldots, Y_n$ una muestra aleatoria simple con la misma distibución que la v.a. $Y$.

1) Puesto que $\mathbb{E}(Y) = -\frac{2}{3}a^2$, planteamos:

$$\bar{Y}_n = -\frac{2}{3}a^2 \iff \hat{a} = \sqrt{\frac{-3}{2}\bar{Y}_n}.$$

2) $L(a|Y_1, ..., Y_n) = \begin{cases} \prod_{i=1}^n \frac{1}{2a\sqrt{Y_i+a^2}} \text{ si } -a^2 \leq Y_i < 0 \, \forall i \\ 0 \text{ en otro caso} \end{cases}$

Luego, para que $L(a)$ sea positiva es necesario que $-a^2 \leq Y_i < 0 \, \forall i$, es decir que $-a^2 \leq \min\{Y_1, ..., Y_n\} = Y_1^*$ y $a \geq \sqrt{-Y_1^*}$. Además, si $a \geq \sqrt{-Y_1^*}$,

$$L(a) = \frac{1}{(2a)^n} \sqrt{\prod_{i=1}^n \frac{1}{Y_i + a^2}}$$

y esta última expresión se maximiza cuando $a$ sea lo más chico posible, es decir que $\hat{a}_{MV} = \sqrt{-Y_1^*}$.

**Ejercicio 2** Si $X \sim N(\mu, \sigma^2)$, entonces $\varphi_X(t) = \mathbb{E}(e^{tX}) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$. Además, sabemos que si $X_1, X_2, ..., X_n$ son independientes, entonces:

$$\begin{aligned}\varphi_{X_1+X_2+...+X_n}(t) &= \varphi_{X_1}(t)\varphi_{X_2}(t)...\varphi_{X_n}(t)\\ &= e^{\mu_1 t + \frac{\sigma_1^2 t^2}{2}} \times e^{\mu_2 t + \frac{\sigma_2^2 t^2}{2}} \times ... \times e^{\mu_n t + \frac{\sigma_n^2 t^2}{2}}\\ &= e^{(\mu_1+\mu_2+...+\mu_n)t + \frac{(\sigma_1^2+\sigma_2^2+...+\sigma_n^2)t^2}{2}}\end{aligned}$$

y esta última secorresponde a la función generatriz de una v.a. :

$$N(\mu_1 + \mu_2 + ... + \mu_n, \sigma_1^2 + \sigma_2^2 + ... + \sigma_n^2)$$

Por unicidad de la función generatriz, se deduce que $X_1 + X_2 + ... + X_n \sim N(\mu_1 + \mu_2 + ... + \mu_n, \sigma_1^2 + \sigma_2^2 + ... + \sigma_n^2)$.

**Ejercicio 3** Se considera una muestra proveniente de una variable aleatoria:

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $x_i^*$ | 2.0 | 2.2 | 2.7 | 3.0 | 3.1 | 3.8 | 4.7 | 4.9 | 5.2 | 5.3 |

*a*) Puesto que asumimos que estos datos provienen de normales (con $\sigma$ desconocido), un intervalo de confianza a nivel $1-\alpha$ para $\mu$ es:

$$I_{1-\alpha} = [\bar{x}_n - k_\alpha, \bar{x}_n + k_\alpha] \qquad \text{con } k_\alpha = \frac{s_n}{\sqrt{n}} t_{1-\frac{\alpha}{2}}(n-1)$$

Como $\bar{x}_n = 3{,}69$, $s_n = 1{,}26$ y $t_{0{,}975}(9) = 2{,}262$, se tiene que $k = 0{,}9$ y $I_{0{,}95} = [2{,}79,\ 4{,}59]$.

*b*) El valor más grande de $|\frac{i}{n} - F_0(x_i^*)|$ es aproximadamente 0,32 y el valor más grande de $|\frac{i-1}{n} - F_0(x_i^*)|$ es aproximadamente 0,22, luego $d_n = 0{,}32$. Mirando la tabla KS para una muestra, resulta que:

$$p - valor = \mathbb{P}(D_n > 0{,}32 | H_0) > 0,32$$

Entonces NO se rechaza $H_0$ a nivel $\alpha = 0{,}05$.

*c*) El valor más grande de la primera tabla para $|\frac{i}{n} - F_m^Y(x_i^*)|$ es $d_1 = 0{,}8$ y el más grande para la segunda tabla es $d_2 = 0{,}6$. Luego, $d_{n,m} = 0{,}8$ y $nmd_{n,m} = 40$. Entonces:

$$p - valor = \mathbb{P}(nmD_{n,m} \geq 40) = 0{,}019 < 0{,}05$$

por lo que se rechaza $H_0$.

**Ejercicio 4** *a*) Sean:

| $i$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $x_i$ | 2 | 3 | 5 | 7 | 8 |
| $y_i$ | 14 | 20 | 32 | 42 | 44 |

La ecuación de la recta de regresión (la que minimiza mínimos cuadrados) es:

$$y = \hat{\beta}_1 x + \hat{\beta}_0 \qquad \text{con} \qquad \hat{\beta}_1 = \frac{\sum(x_i - \bar{x}_n)(y_i - \bar{y}_n)}{\sum(x_i - \bar{x}_n)^2} \qquad \hat{\beta}_0 = \bar{y}_n - \hat{\beta}_1 \bar{x}_n$$

En este caso:

- $\bar{x}_n = 5$
- $\bar{y}_n = 30{,}4$
- $\sum(x_i - \bar{x}_n)(y_i - \bar{y}_n) = 134$
- $\sum(x_i - \bar{x}_n)^2 = 26$
- $\hat{\beta}_1 = 5{,}2$
- $\hat{\beta}_0 = 4{,}4$

Finalmente, la ecuación de la recta de regresión es: $y = 5{,}2x + 4{,}4$

*Observación:* Lo razonable sería considerar $y = peso$ en función de $x = edad$. De todas maneras, se considerará como correcto aquellos que los consideren al revs.

*b*) Según la recta de regresión, el peso de un niño de 6 años sería:

$$y = 5{,}2 \times 6 + 4{,}4 = 35{,}6 \quad Kg$$