Universidad de la República | **Probabilidad y Estadística**
Facultad de Ingeniería - IMERL | **Cursos 2018**

Solución del Examen – 30 de Enero de 2019

Nº de examen	Cédula	Apellido y nombre	Salón

Múltiple opción (Total: 70 puntos)

En cada pregunta hay una sola opción correcta.

Respuesta correcta: 7 puntos   Respuesta incorrecta: -3 puntos   No responde: 0 punto

Respuestas

Ej. 1	Ej. 2 2)	Ej. 3 1)	Ej. 3 2)	Ej. 4	Ej. 5 1)	Ej. 5 2)	Ej 6 1)	Ej. 6 2)	Ej. 7

Ejercicio 1.
Sea X una variable aleatoria discreta con la siguiente función de probabilidad puntual:

x	0	1	2	3
$\mathbb{P}(\mathbf{X}=x)$	$2\theta/3$	$\theta/3$	$2(1-\theta)/3$	$(1-\theta)/3$

en donde $0 \leq \theta \leq 1$ es un parámetro. Se obtuvieron las siguientes 10 observaciones independientes:

3	0	2	1	3	2	1	0	2	1

Entonces una estimación de θ por el método de máxima verosimilitud es:

(A) 1/2   (B) 3/8   (C) 5/16

Solución Ej.1
La función de verosimilitud está dada por:

$$L(\theta) = \prod_{i=1}^{10} P_\theta(X_i) = \frac{1-\theta}{3}\frac{2\theta}{3}\frac{2(1-\theta)}{3}\frac{\theta}{3}\frac{(1-\theta)}{3}\frac{2(1-\theta)}{3}\frac{\theta}{3}\frac{2\theta}{3}\frac{2(1-\theta)}{3}\frac{\theta}{3} = \frac{2^5}{3^{10}}(1-\theta)^5\theta^5$$

Tomando logaritmo, obtenemos que:

$$\log L(\theta) = \log(\frac{2^5}{3^{10}}) + 5\log(1-\theta) + 5\log(\theta)$$

Buscamos θ^* que maximice la verosimilitud, es decir un cero de la derivada. Derivando entonces respecto a θ resulta que:

$$-\frac{5}{1-\theta} + \frac{5}{\theta} = 0 \quad \text{de donde} \quad \theta^* = \frac{5}{10} = \frac{1}{2}$$

Observar que por el signo de la derivada θ^* es efectivamente un máximo.

Ejercicio 2.
Se consultó a 6 estudiantes de la facultad cuántas veces tuvo que reiniciar su computadora en el último mes. De los 6 estudiantes consultados, 4 tienen computadoras marca A y 2 tienen computadoras marca B. Los estudiantes con marca A tuvieron que reiniciar 1, 2, 2 y 8 veces; aquellos con marca B tuvieron que hacerlo 2 y 3 veces.
Se elige al azar un estudiante entre los 6. Sea X la variable aleatoria Bernoulli que indica la marca de su computadora asumiendo $X = 0$ si la marca es A. Sea Y la variable aleatoria que indica la cantidad de veces que el estudiante reinició su computadora.

1. Completar el siguiente cuadro correspondiente a la distribución conjunta de X e Y

	1	2	3	8
0				
1				

2. El valor esperado de Y^2 es aproximadamente:

(A) 11/4 (B) 3 (C) 2

Solución Ejercicio 2

1. La distribución conjunta de X e Y está dada por la siguiente tabla:

X \ Y	**1**	**2**	**3**	**8**
0	1/6	2/6	0	1/6
1	0	1/6	1/6	0

2. La función de probabilidad puntual de Y es: $p_Y(1) = p_Y(3) = p_Y(8) = 1/6$ y $p_Y(2) = 1/2$. Por lo tanto,

$$\mathbb{E}(Y^2) = 1^2 \times \frac{1}{6} + 2^2 \times \frac{1}{2} + 3^2 \times \frac{1}{6} + 8^2 \times \frac{1}{6} = 14,3$$

Ejercicio 3.
Se considera la variable aleatoria X con función de densidad $f_X(x) = \dfrac{K}{1+x^2}$ con $x \in \mathbb{R}$.

1. El valor de K es:

(A) $1/\pi$ (B) $2/\pi$ (C) 1

2. Si $Y = \arctan(X)$ entonces:

(A) $Y \sim \mathbf{U}[0,1]$ (B) $Y \sim \mathbf{U}\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ (C) $Y \sim \mathbf{U}\left[0, \frac{\pi}{2}\right]$

Solución Ejercicio 3

1. Para que la función f que es positiva y continua sea densidad debe integrar 1 en su recorrido, es decir:

$$1 = \int_{-\infty}^{+\infty} \frac{k}{1+x^2} dx = k \times \arctan(x)|_{-\infty}^{\infty} = k \times \left(\frac{\pi}{2} - \frac{-\pi}{2}\right) = k\pi \Rightarrow k = \frac{1}{\pi}.$$

2. $Y = \arctan(X)$, por lo tanto el recorrido de Y es $[-\pi/2, \pi/2]$. La función de distribución de Y es:

$$F_Y(y) \quad = \quad \mathbb{P}(\arctan(Y) \leqslant y) = \begin{cases} 1 & y > \pi/2 \\ \int\limits_{-\infty}^{y} \frac{1}{\pi(1+x^2)} dx & -\pi/2 \leqslant y \leqslant \pi/2 \\ 0 & y \leqslant 0 \end{cases} = \begin{cases} 1 & y > \pi/2 \\ \frac{1}{\pi}(\arctan(\tan(y)) + \dfrac{\pi}{2}) & -\pi/2 \leqslant y \leqslant \pi/2 \\ 0 & y \leqslant 0 \end{cases}$$

$$= \begin{cases} 1 & y > \pi/2 \\ \dfrac{2y+\pi}{2\pi} & -\pi/2 \leqslant y \leqslant \pi/2 \Rightarrow \\ 0 & y \leqslant 0 \end{cases}$$

De donde resulta que Y es Uniforme en el intervalo $[-\pi/2, \pi/2]$.

Ejercicio 4.
Se consideran dos variables normales independientes X e Y tales que:

- $\mathbb{E}(2X + Y + 1) = 10$
- La densidad de Y es simétrica respecto de $x = 4$

Entonces $\mathbb{P}(X \geq 2{,}5, Y \leq 4) =$

(A) 0,25 (B) 0 (C) mayor que 0,5

Solución Ejercicio 4
Al ser independientes resulta que:

$$\mathbb{P}(X \geq 2{,}5, Y \leq 4) = \mathbb{P}(X \geq 2{,}5)\mathbb{P}(Y \leq 4)$$

De los datos que tenemos resulta que:

- al ser Y una variable aleatoria normal simétrica respecto a 4, tenemos que $\mathbb{E}(Y) = 4$ y $\mathbb{P}(Y \leq 4) = 0,5$.
- $10 = \mathbb{E}(2X + Y + 1) = 2\mathbb{E}(X) + \mathbb{E}(Y) + 1 = 2\mathbb{E}(X) + 4 + 1$, de donde $\mathbb{E}(X) = 2,5$ y $\mathbb{P}(X \geq 2{,}5) = 0,5$.

Por lo tanto $\mathbb{P}(X \geq 2{,}5, Y \leq 4) = 0,5 \times 0,5 = 0,25$.

Ejercicio 5.
Una sala de espectáculos tiene capacidad para 150 espectadores sentados. Para el estreno de su nuevo espectáculo decide vender 160 entradas de modo de asegurarse que la sala esté llena el día del estreno. Los registros indican que 1 de cada 10 personas que compran entrada no concurre el día del espectáculo. Se asume además que los espectadores actuán de manera independiente unos de otros.

1. El número esperado de personas que concurren el día del estreno es:

   (A) 150 (B) 144 (C) 145

2. Sea A el evento "el día del estreno hay al menos una persona parada". La probabilidad de A es aproximadamente:

   (A) 0,1 (B) 0,032 (C) 0,313

Solución Ejercicio 5

1. Sea W la variable aleatoria que cuenta el número de personas que asisten al espectáculo, resulta entonces que $W \sim \text{Bin}(160, \frac{9}{10})$, de donde $\mathbb{E}(W) = 160\frac{9}{10} = 144$

2. Para que hayan al menos una persona parada deben asistir al menos 151 personas, eso es $W \geqslant 151$. Utilizando la aproximación Normal a la Binomial, es decir $W \approx \mathcal{N}(\mu, \sigma^2)$ con $\mu = 144$ y $\sigma^2 = 160\frac{9}{10}\frac{1}{10} = 14,4$, resulta que:

$$\mathbb{P}(W \geqslant 151) \approx 1 - \Phi\left(\frac{151 - 144}{\sqrt{14,4}}\right) = 1 - 0{,}9671 = 0{,}0329.$$

Ejercicio 6.
Se asume que los puntajes obtenidos por los estudiantes en un examen (normalizados entre 0 y 1) se distribuyen según la siguiente densidad:

$$f(x) = \begin{cases} ax & 0 \leq x \leq 1/2 \\ a(1-x) & 1/2 \leq x \leq 1 \\ 0 & \text{en otro caso} \end{cases}$$

El umbral de aprobación es igual a 0,55.

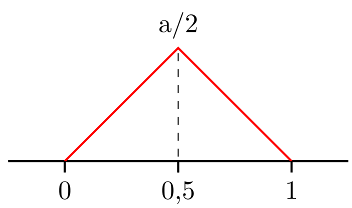
1. La probabilidad de que un estudiante apruebe el examen es:

   (A) $0,605$ (B) $0,2025$ (C) $0,405$

2. El primer cuartil de la distribución de puntajes es:

   (A) $0,25$ (B) $0,35$ (C) $0,125$

Solución Ejercicio 6
Graficando la densidad de los puntajes, observamos que el gráfico es un triángulo isósceles con base en $[0, 1]$ y altura $\frac{a}{2}$. Entonces, para que el área sea 1 debe ser $a = 4$.



1. Equivale a hallar el área del triángulo con base $[0{,}55, 1]$, esto es $\frac{0{,}45 \cdot 1{,}8}{2} = 0{,}405$.

2. El primer cuartil q_1 es tal que el área bajo al densidad del intervalo $[0, q_1]$ es igual a 0.25. Buscamos entonces q_1 tal que $\frac{q_1 * 4q_1}{2} = 0,25$, de donde $q_1 = 0{,}354$ (solo sirve la solución positiva).

Ejercicio 7.
Una embotelladora vierte refresco en sus botellas según una distribución normal con media $500\,mL$ y varianza $100\,mL^2$. El llenado de cada botella se hace de manera independiente. Si se llenan cuatro botellas, entonces con probabilidad 0.95 el total de refresco vertido en las cuatro botellas pertenece al intervalo:

(A) [1967 ; 2033] (B) [1980,4 ; 2019,6] (C) [1960,8 ; 2039,2]

Solución Ejercicio 7

Sea S_4 la variable aleatoria que representa el total de refresco vertido en las cuatro botellas. La distribución de S_4 es Normal con parámetros $\mu = 4 \times 500 = 2000mL)$ y $\sigma^2 = 4 \times 100 = 400ml^2$.

El intervalo buscado está dado entonces por $[\mu - \epsilon, \mu + \epsilon]$ con $\epsilon = z_{0,025} \times \sigma = 1,96\sqrt{400} = 1,96 \times 20 = 39,2$. Es decir $[19860,8; 2039,2]$.

Desarrollo (Total: 30 puntos)

Justificar todos los cálculos y argumentos utilizados en las respuestas.

El ejercicio 10 corresponde al 1er semestre 2018, el ejercicio 9 al 2do semestre 2018.

Se deberá realizar uno y sólo uno de estos ejercicios a elección.

Ejercicio 8.

Sea $X_1, \ldots, X_n$ una muestra aleatoria simple con distribución normal $\mathcal{N}(\mu, \sigma^2)$ con μ desconocido y σ conocido. Sea $\alpha \in (0,1)$, probar que el intervalo $I_n(\alpha)$ siendo

$$I_n(\alpha) = \left[\overline{X}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}; \overline{X}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

es un intervalo de confianza para μ al $(1-\alpha) \times 100\,\%$. Se asume como conocido el siguiente resultado: si $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ e $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ independientes, entonces $X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.

Solución Ejercicio 8

Basta observar que $\overline{X}_n \sim \mathcal{N}(\mu_1, \sigma_1^2)$ con $\mu_1 = \mu$ y $\sigma_1^2 = \frac{\sigma^2}{n}$. Resulta entonces que:

$$\mathbb{P}(\mu \in I_n(\alpha)) = \mathbb{P}\left(-z_{\alpha/2} \leq \frac{\overline{X}_n - \mu_1}{\sigma_1} \leq z_{\alpha/2}\right) = \Phi(z_{\alpha/2}) - \Phi(-z_{\alpha/2}) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha$$

Ejercicio 9.

1. Se consideran los siguientes datos:

3,2	0,3	1,95	0,5	0,25	0,9	2,85	1,9	0,85	1,4

   ¿Se puede suponer que esta muestra proviene de una distribución normal?

2. Se consideran los siguientes datos:

2,5	2,0	4,35	3,05	3,9	3,6

   ¿Se puede suponer que estos datos provienen de la misma distribución que los datos anteriores?

Solución Ejercicio 9

1. Realizaremos el test de ajuste a distribuciones normales de D'Agostino.

$$\begin{cases} H_0 : \text{Los datos son Normales} \\ H_1 : \text{ no } H_0 \end{cases}$$

El estadístico es $D = \frac{T}{n^2 \cdot \sigma_x}$ con $T = \sum_{i=1}^{n}(i - \frac{n+1}{2})X_{(i)}$ y $\sigma_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}$.

Para la muestra dada, resulta que $T = 27{,}65$, $\sigma_x = 0{,}9893$ y $D = 0,2794$.

Tabla del test, para $n = 10$:

α	Intervalo		¿H_0 =es normal?
0.2	0.2632	0.2835	No rechaza H_0
0.1	0.2573	0.2843	No rechaza H_0
0.05	0.2513	0.2849	No rechaza H_0
0.02	0.2436	0.2855	No rechaza H_0
0.01	0.2379	0.2857	No rechaza H_0

En este caso No se rechaza H_0 para ninguno de los valores de α. En particular para $\alpha = 0{,}1$.

2. Realizaremos el test de Kolmogorov-Smirnov de comparación de dos muestras:

$$\begin{cases} H_0 : F_Y = F_X \\ H_1 : \text{ no } H_0 \end{cases}$$

El estadístico del test es $D_{n,m} = \sup_{x \in \mathbb{R}} |F_m^Y(x) - F_n^X(x)|$

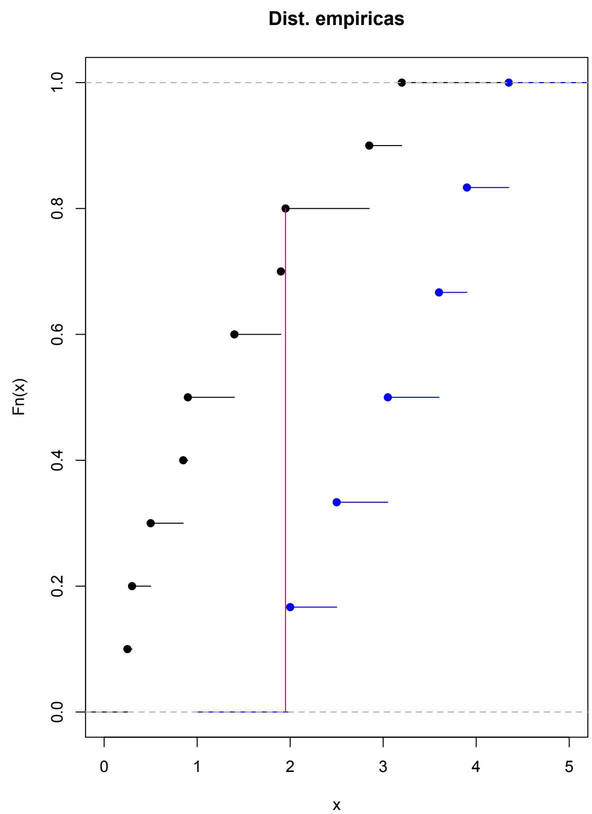
y	0.25	0.30	0.50	0.85	0.90	1.40	1.90	1.95	2.85	3.20
F_Y	1/10	1/5	3/10	2/5	1/2	3/5	7/10	4/5	9/10	1
F_X	0	0	0	0	0	0	0	0	1/3	1/2
$\lim_{x \to y^-} F_X(x)$	0	0	0	0	0	0	0	0	1/3	1/2
$\|F_Y - F_X\|$	1/10	1/5	3/10	2/5	1/2	3/5	7/10	4/5	17/30	1/2
$\|F_Y - \lim_{x \to y^-} F_X(x)\|$	1/10	1/5	3/10	2/5	1/2	3/5	7/10	4/5	17/30	1/2

x	2.00	2.50	3.05	3.60	3.90	4.35
F_Y	4/5	4/5	9/10	1	1	1
F_X	1/6	1/3	1/2	2/3	5/6	1
$\lim_{y \to x^-} F_Y(y)$	1/6	1/3	1/2	2/3	5/6	1
$\|F_Y - F_X\|$	19/30	7/15	2/5	1/3	1/6	0
$\|F_Y - \lim_{y \to x^-} F_Y(y)\|$	19/30	7/15	2/5	1/3	1/6	0

La máxima diferencia es entonces $D_{n,m} = 4/5 = 0{,}8$ y de la tabla resulta que el p-valor es 0,009.

Por lo tanto para un umbral de $\alpha = 0{,}05$ (o $\alpha = 0{,}1$), se rechaza H_0, es decir que no hay evidencia suficiente que indique que los datos provienen de la misma distribución.

Otra forma de calcular el estadśitico $D_{n,m}$ es observar la gráfica de ambas distribuciones emprícas:



Ejercicio 10.

Se desea estudiar el efecto de un nuevo químico sobre la tensión disruptiva (en kV) de un material. Para esto se divide al azar un lote de 16 muestras del material en dos grupos de 8, unas reciben el tratamiento con el nuevo químico (Tratamiento) y las otras se usan de control. Los resultados se muestran en el diagrama de tallos (espalda con espalda) que está a la derecha.

Tratamiento		Control
00	53	
0000	54	0
0	55	0
0	56	000
	57	00
	58	0

Considere la hipótesis nula

$$H_0 : \text{El tratamiento no tiene efecto sobre la tensión disruptiva del material}$$

y el estadístico $X =$ (suma de respuestas en tratamiento)-(suma de respuestas en control). Se asume que el tratamiento no aumenta la tensión disruptiva del material.

1. Usando la Figura 1, calcular el p-valor $\text{pval}\,(X_{\text{obs}})$. Decidir si rechazar o no H_0 para el valor umbral $p_u = 0{,}05$.
2. Se considera la región de rechazo $I_r(c) = (-\infty, c]$. Usando la Figura 1, hallar el mayor valor de c para el cual la probabilidad de error de tipo I es menor o igual que $\alpha = 0{,}05$. Decidir si rechazar o no H_0.

Solución Ejercicio 10

1. Se considera

   $$H_0 : \text{El tratamiento no tiene efecto sobre la tensión disruptiva del material}$$

   y el estadístico $X =$ (suma de respuestas en tratamiento)-(suma de respuestas en control).

A partir del diagrama de tallos se obtiene que $X_{obs} = 177 - 193 = -16$. Al asumir que el tratamiento no aumenta la tensión disruptiva, las opciones son que reduzca o no tenga efecto, por lo tanto nos planteamos el test a una cola de donde $pval(X_{obs}) = P(X \leq X_{obs}) = P(X \leq -16)$. De la figura asociada (distribución de aleatorización de X) se obtiene que

$$pval(X_{obs}) = P(X \leq -16) = \frac{44+9+2}{C_8^{16}} = \frac{55}{12870} = 0{,}00427 < p_u = 0{,}05$$

Por lo tanto, hay evidencia para rechazar H_0, es decir para afirmar que el químico tiene efecto sobre la tensión disruptiva.

2. Consideramos la región de rechazo $I_r(c) = (-\infty, c]$ y buscamos el máximo c tal que:

$$\alpha = P(\text{error tipo I}) = P(X \in I_r(c)|H_o) = P(X \leq c|H_0) \leq 0{,}05$$

De la Figura se obtiene que:

- si $c = -12$ entonces $\alpha = \frac{412}{12870} = 0{,}032$.
- si $c = -10$ entonces $\alpha = \frac{871}{12870} = 0{,}0646$.

Por lo tanto $c = -12$, es decir la región de rechazo es $I_r = (-\infty, -12]$. Dado que $X_{obs} = -16 \in I_r$, entonces se rechaza H_0.

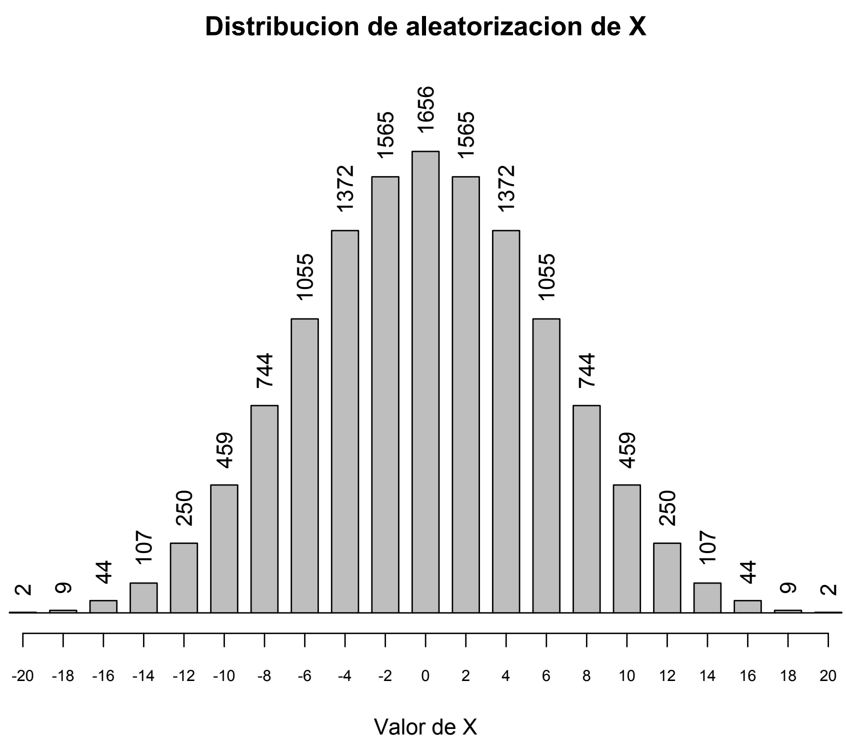


Figura 1: Distribución de aleatorización de X para usar en el Ejercicio 10.