Ingeniería Matemática
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE
MA3403-3 Probabilidades y Estadística
Primavera 2015

Roberto Cortez
Roberto Bobadilla
Dafne Pincheira

Pauta Examen

P1. *a)* De un grupo de 5 mujeres y 7 hombres debe formarse un comité de 2 mujeres y 3 hombres. Determine cuántos posibles comités hay si:

1) (1,0 pto.) no hay restricciones.

Solución. Hay $\binom{5}{2} = 10$ formas de escoger a las 2 mujeres, y $\binom{7}{3} = 35$ formas de escoger a los 3 hombres. Luego, la cantidad buscada es $\binom{5}{2}\binom{7}{3} = 350$.

2) (1,0 pto.) dos hombres están peleados y no pueden ser seleccionados ambos.

Solución. La cantidad de formas en que van ambos es $\binom{5}{2}\binom{5}{1} = 50$. Luego, la cantidad de formas en que no van ambos simultáneamente es $350 - 50 = 300$.

3) (1,0 pto.) hay un hombre y una mujer que son pareja y sólo aceptarán ser parte del comité si se seleccionan a ambos.

Solución. Las formas en que van ambos son $\binom{4}{1}\binom{6}{2} = 4 \times 15 = 60$. Las formas en que no va ninguno son $\binom{4}{2}\binom{6}{3} = 6 \times 20 = 120$. Luego, la cantidad buscada es $60 + 120 = 180$.

b) (3,0 ptos.) Dos bebés *gemelos* provienen del mismo óvulo, por lo cual tienen siempre el mismo sexo, mientras que dos *mellizos* provienen de óvulos distintos y por lo tanto pueden tener distinto sexo con 50 % de probabilidad. Se sabe que en un 64 % de los partos dobles los bebés son del mismo sexo. ¿Qué porcentaje de los partos dobles son de gemelos?

Solución. Sean los eventos

$$G: \text{bebés gemelos}$$
$$S: \text{son del mismo sexo.}$$

Sabemos que $\mathbb{P}(S \mid G^c) = 1/2$, pero $\mathbb{P}(S \mid G) = 1$. Usando la regla de probabilidades totales, tenemos:

$$64\,\% = \mathbb{P}(S) = \mathbb{P}(S \mid G)\mathbb{P}(G) + \mathbb{P}(S \mid G^c)\mathbb{P}(G^c) = \mathbb{P}(G) + (1/2) \times (1 - \mathbb{P}(G))$$

Despejando se obtiene $\mathbb{P}(G) = 28\,\%$.

P2. El tiempo de espera de un bus del transporte público se modela como una variable aleatoria X con densidad $f(x) = \theta^{-1} e^{-x/\theta} \mathbb{1}_{[0,\infty)}(x)$, donde θ es un parámetro desconocido.

a) (1,2 ptos.) Argumente o demuestre que $\mathbb{E}(X) = \theta$ y $\text{var}(X) = \theta^2$.

Solución. Es evidente que $X \sim \exp(1/\theta)$. Como la esperanza y varianza de una variable $\exp(\lambda)$ son $1/\lambda$ y $1/\lambda^2$, respectivamente, se deduce directamente que $\mathbb{E}(X) = \theta$ y $\text{var}(X) = \theta^2$. *Observación:* también puede resolverse calculando las integrales asociadas a la esperanza y varianza.

b) (1,2 ptos.) Muestre que $e^{-X/\theta}$ tiene distribución unif$(0,1)$.

Solución. Sea $Y = e^{-X/\theta}$. Calculemos F_Y. Como X toma valores en $(0,\infty)$, es claro que Y toma valores en $(0,1)$. Para $y \in (0,1)$, tenemos:

$$F_Y(y) = \mathbb{P}(e^{-X/\theta} \le y) = \mathbb{P}(X \ge -\theta \log y) = e^{-(-\theta \log y)/\theta} = y,$$

donde en el último paso hemos usado el hecho que $\mathbb{P}(X \ge x) = e^{-x/\theta}$, lo cual es directo del hecho que $X \sim \exp(1/\theta)$. Derivando, se obtiene $f_Y(y) = 1$ para $y \in (0,1)$, mientras que para $y \notin (0,1)$ es directo que $f_Y(y) = 0$. Esto muestra que $Y \sim \text{unif}(0,1)$.

c) (1,2 ptos.) Se dispone de una m.a.s $X_1, \dots, X_n$ de la distribución de X. Calcule el estimador de máxima verosimilitud de θ.

Solución. La función de verosimilitud es

$$L = L(x_1, \dots, x_n; \theta) = \prod_{i=1}^{n} \theta^{-1} e^{-x_i/\theta} = \theta^{-n} e^{-n\bar{x}/\theta}.$$

A continuación maximizamos lo anterior con respecto a θ, para lo cual tomamos logaritmo, derivamos e igualamos a 0:

$$0 = \frac{\partial \log L}{\partial \theta} = \frac{\partial}{\partial \theta}(-n \log \theta - n\bar{x}/\theta) = -n/\theta + n\bar{x}/\theta^2,$$

de donde se obtiene $\theta = \bar{x}$. Es decir, el estimador de máxima verosimilitud de θ es $\bar{X}$.

d) (1,2 ptos.) Se utilizará $\bar{X}$ para estimar θ, y se desea que el error relativo (es decir, la diferencia absoluta entre la estimación y θ, dividido por θ) sea menor que 10 % con probabilidad de un 95 %. ¿Cuántas mediciones se necesitan? Aproxime con el TCL.

Solución. Sabemos que la esperanza y varianza de X son $\mu = \theta$ y $\sigma^2 = \theta^2$. Imponemos lo deseado y armamos el estadístico del TCL:

$$95\,\% = \mathbb{P}\left(\frac{|\bar{X} - \theta|}{\theta} \leq \frac{1}{10}\right) = \mathbb{P}\left(\frac{|\bar{X} - \mu|}{\sigma/\sqrt{n}} \leq \frac{\sqrt{n}}{10}\right) \approx \mathbb{P}\left(|\mathcal{N}(0,1)| \leq \frac{\sqrt{n}}{10}\right),$$

donde en el último paso hemos utilizado el TCL. Lo anterior equivale a que $\mathbb{P}(\mathcal{N}(0,1) > \sqrt{n}/10) \approx 2{,}5\,\%$, y de la tabla normal se obtiene que $\sqrt{n}/10 \approx 1{,}96$. Es decir, $n \approx 19{,}6^2 = 384{,}16 \approx 385$.

e) (1,2 ptos.) Un parámetro importante es p, correspondiente a la probabilidad que el tiempo de espera sea mayor a 10 minutos. En una muestra de tamaño 100, se observó que 10 de los buses tardaron más de 10 minutos. Entregue un intervalo de confianza para p al nivel 90 %.

Solución. Trabajamos con el estadístico

$$Z = \frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})/n}},$$

el cual sabemos que posee distribución aproximada $\mathcal{N}(0,1)$. Imponemos el nivel deseado a Z usando un intervalo simétrico, es decir,

$$90\,\% = \mathbb{P}(-c \leq Z \leq c) \approx \mathbb{P}(|\mathcal{N}(0,1)| \leq c),$$

lo cual equivale a que $\mathbb{P}(\mathcal{N}(0,1) > c) \approx 5\,\%$, y de la tabla se obtiene $c = 1{,}65$. A partir de lo anterior despejamos p y obtenemos el intervalo deseado:

$$90\,\% = \mathbb{P}\left(-c \leq \frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})/n}} \leq c\right) = \mathbb{P}\left(\hat{p} - \frac{c\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} \leq p \leq \hat{p} + \frac{c\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right).$$

Como el valor observado de $\hat{p}$ es $10/100$, se tiene que $c\sqrt{\hat{p}(1-\hat{p})}/\sqrt{n} = 1{,}65 \times \sqrt{9/100}/10 = 4{,}95/100$. Luego, el intervalo buscado es $[10/100 - 4{,}95/100,\ 10/100 + 4{,}95/100] = [5{,}05\,\%,\ 14{,}95\,\%]$.

P3. *a*) (3,0 ptos.) Los empleados en una fábrica deben asistir a una capacitación para aprender a armar los componentes de un determinado producto. Existen dos posibles capacitaciones, la Estándar y la Novedosa. Antes de escoger permanentemente una de ellas, se envía a 9 empleados a capacitarse en la primera y a otros 9 a capacitarse en la segunda, y luego se midió el tiempo (en segundos) que tardó cada empleado en armar el producto. Los datos relevantes de estas mediciones se resumen en la tabla. Suponiendo que las muestras provienen de distribuciones normales con la misma varianza, ¿hay suficiente evidencia para afirmar que las capacitaciones generan tiempos de armado promedio distintos? Use $\alpha = 0{,}05$.

	Estándar	Novedosa
Tiempo promedio	390	310
$\sum(x_i - \bar{x})^2$	1950	1250

Solución. Llamemos μ y ν a las medias de los tiempos de las capacitaciones Estándar y Novedosa, respectivamente. Planteamos

$$H_0 : \mu = \nu$$
$$H_1 : \mu \neq \nu.$$

Bajo H_0, sabemos que el estadístico $T = (\bar{X} - \bar{Y})/(S\sqrt{\frac{1}{n} + \frac{1}{m}})$ tiene distribución $t_{n+m-2} = t_{16}$. Por la forma de H_1, rechazaremos cuando $|T| > c$, y para encontrar c imponemos el nivel deseado:

$$\alpha = 5\,\% = \mathbb{P}(|T| > c \mid H_0) = \mathbb{P}(|t_{16}| > c),$$

lo que equivale a que $\mathbb{P}(t_{16} > c) = 2{,}5\,\%$, y de la tabla se obtiene $c = 2{,}120$. Calculemos el valor observado de T en base a los datos:

$$S^2_{\text{obs}} = [(n-1)S^2_X + (m-1)S^2_Y]/(n+m-2) = [1950 + 1250]/16 = 200,$$

y entonces

$$T_{\text{obs}} = \frac{390 - 310}{\sqrt{200}\sqrt{\frac{1}{9} + \frac{1}{9}}} = \frac{80}{10\sqrt{2} \times \sqrt{2}/3} = 6.$$

Como $|T_{\text{obs}}| = 6 > 2{,}120 = c$, corresponde rechazar H_0. Es decir, las capacitaciones efectivamente tienen tiempo medio de armado distintos.

b) (3,0 ptos.) En el mercado de telefonía móvil existen 3 compañías: A, B y C. Un estudio de mercado encargado por la compañia A afirma que hay el doble de clientes A que clientes B, y que hay el doble de clientes B que clientes C. Usted, que trabaja para la competencia de A, desconfía de tal afirmación, por lo cual encarga realizar un nuevo estudio con 70 clientes, cuyos resultados se resumen en la tabla. Plantee las hipótesis del test de bondad de ajuste correspondiente, y calcule el p-valor (o una aproximación). ¿Es razonable rechazar la afirmación del estudio encargado por A?

Compañía	A	B	C
Cantidad de clientes	34	26	10

Solución. Si p_1, p_2 y p_3 son las probabilidades de que un cliente sea de la compañía A, B y C, respectivamente, la afirmación del estudio previo es que $p_1 = 2p_2$ y $p_2 = 2p_3$, y como $p_1 + p_2 + p_3 = 1$, al despejar sus valores se deduce que las hipótesis del test son

$$H_0 : p_1 = 4/7, \quad p_2 = 2/7, \quad p_3 = 1/7,$$
$$H_1 : p_1 \neq 4/7, \quad \text{o bien } p_2 \neq 2/7, \quad \text{o bien } p_3 \neq 1/7.$$

Para obtener el p-valor necesitamos calcular el valor observado del estadístico Δ:

$$\Delta_{\text{obs}} = 70\left[\frac{\left(\frac{34}{70} - \frac{40}{70}\right)^2}{\frac{40}{70}} + \frac{\left(\frac{26}{70} - \frac{20}{70}\right)^2}{\frac{20}{70}} + \frac{\left(\frac{10}{70} - \frac{10}{70}\right)^2}{\frac{10}{70}}\right] = \frac{36}{40} + \frac{36}{20} + \frac{0}{10} = \frac{108}{40} = 2{,}7.$$

Sabemos que bajo H_0, el estadístico Δ tiene distribución aproximada χ^2_{k-1}, donde $k = 3$ en este caso. Con esto y mirando la tabla podemos calcular el p-valor:

$$p\text{-valor} = \mathbb{P}(\Delta \geq \Delta_{\text{obs}} \mid H_0) \approx \mathbb{P}(\chi^2_2 \geq 2{,}7) \approx 0{,}25.$$

Es decir, el p-valor es de un 25 %, lo cual no es suficientemente pequeño para rechazar H_0, por lo cual no es razonable rechazar la afirmación del estudio encargado por A.