Ingeniería Matemática
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE
MA3403-4 Probabilidades y Estadística
Otoño 2016

Roberto Cortez
Cristóbal Parraguez
Bryan Sagredo

Pauta Examen

P1. *a)* Usted posee 10 pares de calcetines blancos y 20 pares azules, de los cuales extrae 16 calcetines al azar y los aparea de acuerdo a su color. Sea X la variable que denota la cantidad de *pares* blancos obtenidos.

1) (1,0 pto.) Indique el rango de X.

Solución. Es claro que $R_X = \{0, \dots, 8\}$, pues la mínima cantidad de pares blancos obtenidos es 0 (caso en que se extrae 0 ó 1 calcetín blanco) y lo máximo es 8 (caso en que los 16 calcetines extraídos son blancos), siendo posibles todos los casos intermedios.

2) (2,0 ptos.) Calcule la distribución de X.

Solución. Trabajamos en el espacio equiprobable de los $\binom{60}{16}$ subconjuntos de tamaño 16 del total de $60 = 2 \times 10 + 2 \times 20$ calcetines (distinguiendo cada calcetín). Dado $k \in \{0, \dots, 8\}$, para que ocurra que $X = k$ se necesita extraer exactamente $2k$ o bien $2k+1$ calcetines blancos del total de 20, y el resto deben ser escogidos entre los 40 azules. La cantidad de formas que esto ocurre es

$$\binom{20}{2k}\binom{40}{16-2k} + \binom{20}{2k+1}\binom{40}{16-2k-1},$$

entendiendo que en el caso $k = 8$ el segundo término es 0. Luego, la distribución de X viene dada por

$$p_X(k) = \mathbb{P}(X = k) = \frac{\binom{20}{2k}\binom{40}{16-2k} + \binom{20}{2k+1}\binom{40}{16-2k-1}}{\binom{60}{16}}.$$

b) Su pókemon favorito aparecerá en las cercanías de la Facultad en un instante distribuido uniformemente entre las 9:00 y las 18:00, pero usted tiene un examen entre las 12:00 y las 15:00 y no podrá atraparlo mientras lo rinde. Si aparece en cualquier instante antes de las 12:00, la probabilidad que usted lo atrape es $1/2$, pero si aparece después de las 15:00, la probabilidad es de $1/8$.

1) (1,0 pto.) Calcule la probabilidad de atraparlo.

Solución. Sea X la v.a. del instante en que aparece el pokémon, y sea A el evento en que usted lo atrapa. Usando probabilidades totales, tenemos:

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A \mid X \in [9,12])\mathbb{P}(X \in [9,12]) + \mathbb{P}(A \mid X \in [12,15])\mathbb{P}(X \in [12,15]) \\ &\quad + \mathbb{P}(A \mid X \in [15,18])\mathbb{P}(X \in [15,18]) \\ &= \frac{1}{2} \times \frac{3}{9} + 0 \times \frac{3}{9} + \frac{1}{8} \times \frac{3}{9} = \frac{5}{24}.\end{aligned}$$

2) (2,0 ptos.) Suponiendo que usted logró atrapar al pokémon, calcule la probabilidad de que éste haya aparecido después de las 11:00.

Solución. Por la regla de probabilidades totales, tenemos:

$$\begin{aligned}\mathbb{P}(X > 11, A) &= \mathbb{P}(X > 11, A \mid X \in [9,11])\mathbb{P}(X \in [9,11]) + \mathbb{P}(X > 11, A \mid X \in [11,12])\mathbb{P}(X \in [11,12]) \\ &\quad + \mathbb{P}(X > 11, A \mid X \in [12,15])\mathbb{P}(X \in [12,15]) + \mathbb{P}(X > 11, A \mid X \in [15,18])\mathbb{P}(X \in [15,18]) \\ &= 0 \times \frac{2}{9} + \frac{1}{2} \times \frac{1}{9} + 0 \times \frac{3}{9} + \frac{1}{8} \times \frac{3}{9} = \frac{7}{72},\end{aligned}$$

donde hemos usado el hecho que el evento $X > 11$ es incompatible con $X \in [9,11]$ (intersección vacía), y es redundante con el evento $X \in [11,12]$, con $X \in [12,15]$ y con $X \in [15,18]$. Luego, por definición de probabilidad condicional, la probabilidad buscada es:

$$\mathbb{P}(X > 11 \mid A) = \frac{\mathbb{P}(X > 11, A)}{\mathbb{P}(A)} = \frac{7/72}{5/24} = \frac{7}{15}.$$

P2. *a)* Sea X una variable aleatoria con densidad $f_X(x) = \theta^{-2} x e^{-x/\theta} \mathbb{1}_{[0,\infty)}(x)$, donde $\theta > 0$ es un parámetro desconocido.

1) (1,5 ptos.) Muestre que $\mathbb{E}(X) = 2\theta$ y $\text{var}(X) = 2\theta^2$.

Solución. Calculemos:

$$\mathbb{E}(X) = \frac{1}{\theta^2}\int_0^\infty x^2 e^{-x/\theta} dx = \frac{1}{\theta^2}\left(-x^2\theta e^{-x/\theta} + \int_0^\infty 2x\theta e^{-x/\theta} dx\right) = 2\theta \int_0^\infty \frac{1}{\theta^2} x e^{-x/\theta} dx,$$

donde la última integral vale 1 pues el integrando es la densidad de X. Se obtiene entonces $\mathbb{E}(X) = 2\theta$. Calculemos $\mathbb{E}(X^2)$:

$$\mathbb{E}(X^2) = \frac{1}{\theta^2}\int_0^\infty x^3 e^{-x/\theta} dx = \frac{1}{\theta^2}\left(-x^3\theta e^{-x/\theta} + \int_0^\infty 3x^2\theta e^{-x/\theta} dx\right) = 3\theta \int_0^\infty \frac{1}{\theta^2} x^2 e^{-x/\theta} dx,$$

donde la última integral corresponde a la esperanza de X. Tenemos entonces que $\mathbb{E}(X^2) = (3\theta)(2\theta) = 6\theta^2$. Luego, la varianza es:

$$\text{var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = 6\theta^2 - (2\theta)^2 = 2\theta^2.$$

Alternativamente: al inspeccionar la forma de f_X se puede reconocer la densidad de una variable gamma(η, λ) para $\eta = 2$ y $\lambda = 1/\theta$. Como la esperanza y varianza de una variable gamma(η, λ) son η/λ y η/λ^2 respectivamente, es directo que $\mathbb{E}(X) = 2\theta$ y $\text{var}(X) = 2\theta^2$.

2) (1,5 ptos.) Considere una m.a.s. $X_1, \ldots, X_n$ proveniente de la distribución de X. Calcule el estimador de máxima verosimilitud de θ, muestre que es insesgado, y que converge casi seguramente a θ cuando $n \to \infty$.

Solución. Si $x_1, \ldots, x_n$ son los valores que toma la muestra (necesariamente todos no negativos), la verosimilitud es

$$L(x_1, \ldots, x_n; \theta) = \prod_{i=1}^n f_X(x_i) = \prod_{i=1}^n x_i \frac{1}{\theta^2} e^{-x_i/\theta} = \theta^{-2n} e^{-\sum x_i/\theta} \prod_{i=1}^n x_i.$$

Para encontrar el estimador de máxima verosimilitud, derivamos con respecto a θ e igualamos a cero. Es decir:

$$0 = -2n\theta^{-2n-1} e^{-\sum x_i/\theta} \prod_{i=1}^n x_i + \theta^{-2n} e^{-\sum x_i/\theta} \prod_{i=1}^n x_i \frac{1}{\theta^2} \sum x_i.$$

Simplificando la exponencial y el producto, y multiplicando por θ^{2n+2}, sigue que $0 = -2n\theta + \sum x_i$. Despejando θ, obtenemos que el estimador de máxima verosimilitud es $\hat{\theta} = \bar{X}/2$. Como $\bar{X}$ es un estimador insesgado de la media, es directo que $\mathbb{E}(\hat{\theta}) = \mathbb{E}(\bar{X})/2 = \mathbb{E}(X)/2 = 2\theta/2 = \theta$, es decir, $\hat{\theta}$ es insesgado. Finalmente: por la LFGN sabemos que $\bar{X}$ converge casi seguramente a $\mathbb{E}(X) = 2\theta$ cuando $n \to \infty$, por lo tanto $\hat{\theta} = \bar{X}/2$ converge casi seguramente a $2\theta/2 = \theta$.

b) (3,0 ptos.) Se requiere medir, con la mayor precisión posible, el diámetro d (en metros) de un asteroide que orbita el sistema solar. Se dispone de un instrumento de medición adecuado, pero cada vez que se mide, no se obtiene exactamente d, sino que una variable aleatoria cuya distribución es desconocida, pero que puede asumirse que tiene esperanza d y varianza de 10000. Usted planea tomar varias mediciones independientes y estimar d usando el promedio. ¿Cuántas mediciones necesita tomar para que, con probabilidad (aproximada) de 95,44 %, el error de su estimación sea inferior a 5 metros?

Solución. Sean $X_1, \ldots, X_n$ las mediciones a tomar, las cuales son variables i.i.d. con media d y varianza $\sigma^2 = 10000$, y n es la incógnita. Imponemos la probabilidad deseada y armamos el estadístico apropiado para usar el TCL:

$$95{,}44\,\% = \mathbb{P}(|\bar{X} - d| < 5) = \mathbb{P}\left(\frac{|\bar{X} - d|}{\sigma/\sqrt{n}} < \frac{5}{100/\sqrt{n}}\right) \approx \mathbb{P}(|Z| < \sqrt{n}/20) = 1 - 2\mathbb{P}(Z \geq \sqrt{n}/20),$$

donde en el penúltimo paso hemos utilizado el TCL para aproximar la probabilidad mediante una variable $Z \sim \mathcal{N}(0,1)$. Lo anterior equivale a $\mathbb{P}(Z \geq \sqrt{n}/20) \approx 2{,}28\,\%$, y de la tabla normal se deduce que $\sqrt{n}/20 \approx 2$. Es decir, $n \approx 1600$.

P3. *a*) (3,0 ptos.) Al discutir los efectos del cambio climático en su ciudad natal, usted argumenta que los inviernos son notoriamente más fríos que hace dos décadas, mientras que una amiga suya plantea que eso no es posible porque los efectos del cambio climático son a más largo plazo. Para zanjar la discusión, se accede a los datos de la temperatura más fría del año entre 1989 y 1996, y lo mismo entre 2009 y 2016, obteniendo promedios de 2,5 y 1,2, y desviaciones estándar de 1,0 y 1,0, respectivamente. Suponiendo muestras normales con igual varianza, plantee las hipótesis correspondientes, calcule el p-valor del test, e indique su conclusión para un nivel de 5 %.

Solución. Sean $X_1, \ldots, X_8$ los datos entre 1989 y 1996, e $Y_1, \ldots, Y_8$ los datos entre 2009 y 2016, los cuales se modelan como muestras $\mathcal{N}(\mu, \sigma^2)$ y $\mathcal{N}(\nu, \sigma^2)$, respectivamente. Planteamos las hipótesis

$$H_0 : \mu = \nu,$$
$$H_1 : \mu > \nu.$$

Para calcular el p-valor, calculemos primero el valor observado del estadístico correspondiente:

$$S^2_{\text{obs}} = \frac{(n-1)S^2_{X,\text{obs}} + (m-1)S^2_{Y,\text{obs}}}{n+m-2} = \frac{7 \times 1 + 7 \times 1}{14} = 1,$$
$$T_{\text{obs}} = \frac{\bar{X}_{\text{obs}} - \bar{Y}_{\text{obs}}}{S_{\text{obs}}\sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{2{,}5 - 1{,}2}{1 \times \sqrt{\frac{1}{4}}} = 2{,}6.$$

Sabiendo que bajo H_0 el estadístico T tiene distribución t_{n+m-2}, calculamos el p-valor usando la tabla t-student:

$$p\text{-valor} = \mathbb{P}(T > T_{\text{obs}} \mid H_0) = \mathbb{P}(t_{14} > 2{,}6) \approx \mathbb{P}(t_{14} > 2{,}624) = 1\,\%.$$

Como p-valor $\approx 1\,\% < 5\,\% = \alpha$, corresponde rechazar H_0. Es decir, hay suficiente evidencia para afirmar que los inviernos son más fríos que hace dos décadas.

b) (3,0 ptos.) Una autopista posee 4 pistas, y se desea investigar si los conductores tienen preferencia por alguna de ellas. Se observó la pista por la que transitaron 1000 automóviles, cuyos resultados se resumen en la siguiente tabla. Calcule el p-valor (o una aproximación) del test correspondiente, e indique su conclusión para $\alpha = 5\,\%$.

Pista	1	2	3	4
Cantidad observada	265	220	260	255

Solución. La hipótesis nula es que no hay preferencia de pistas. Es decir, $p_i = p_i^0$ para $p_i^0 = 1/4 = 250/1000$, $\forall i = 1,2,3,4$, donde p_i es la probabilidad de que un conductor escoja la pista i. La hipótesis alternativa es que algún p_i es distinto de $1/4$. Para calcular el p-valor, necesitamos el valor observado del estadístico correspondiente. Anotando $n_1 = 265$, $n_2 = 220$, $n_3 = 260$ y $n_4 = 255$, tenemos:

$$\Delta_{\text{obs}} = n \sum_{j=1}^{4} (\hat{p}_{j,\text{obs}} - p_j^0)^2 / p_j^0 = \frac{1}{250} \sum_{j=1}^{4} (n_j - 250)^2 = \frac{1}{250}[15^2 + (-30)^2 + 10^2 + 5^2] = 5.$$

Sabemos que bajo H_0, el estadístico Δ tiene distribución aproximada χ_{4-1}, lo cual permite calcular el p-valor:

$$p\text{-valor} = \mathbb{P}(\Delta > \Delta_{\text{obs}} \mid H_0) \approx \mathbb{P}(\chi_3 > 5) \geq \mathbb{P}(\chi_3 > 6{,}25) = 10\,\%.$$

(También puede aproximarse por $\mathbb{P}(\chi_3 > 4{,}11) = 25\,\%$ o con un valor intermedio entre 10 % y 25 %, lo cual no cambia la siguiente conclusión). Como p-valor $\geq 10\,\% > 5\,\% = \alpha$, no se rechaza H_0. Es decir, no hay suficiente evidencia en los datos para afirmar que algunas pistas son preferidas sobre otras.