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 Control 3

P1. *a)* (3,0 ptos.) Se sabe que el valor esperado del puntaje que obtiene un alumno en el examen final de un ramo es de 75 y la varianza es 25. ¿Cuántos alumnos tienen que dar el examen para asegurar que, con probabilidad de al menos un 99 %, el promedio de notas esté entre 70 y 80? Obtenga un resultado sin utilizar el teorema central del límite, y otro utilizándolo.

Solución. Sean $X_1, \dots, X_n$ los resultados de n alumnos, y sea $\bar{X}$ su promedio. Tenemos:

$$\mathbb{P}(\bar{X} \in (70, 80)) = \mathbb{P}(|\bar{X} - 75| < 5) = 1 - \mathbb{P}(|\bar{X} - \mu| > 5),$$

donde μ también es la esperanza de $\bar{X}$. Además, $\text{var}(\bar{X}) = \sigma^2/n$. Luego, aplicando Chebyshev obtenemos:

$$\mathbb{P}(\bar{X} \in (70, 80)) \geq 1 - \frac{\text{var}(X)}{5^2} = 1 - \frac{25}{25n} = 1 - \frac{1}{n}.$$

Por lo tanto, si se desea que esta probabilidad sea al menos de un 99 %, basta imponer que $1 - 1/n = 0{,}99$, es decir, $n = 100$. Ahora, obtengamos un resultado utilizando el teorema central del límite:

$$\mathbb{P}(\bar{X} \in (70, 80)) = \mathbb{P}(|\bar{X} - 75| < 5) = \mathbb{P}\left(\left| \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right| < \frac{5}{\sigma/\sqrt{n}} \right) \approx \mathbb{P}(|Z| < \sqrt{n}),$$

donde hemos aproximado la distribución de $(\bar{X} - \mu)/(\sigma/\sqrt{n})$ por una normal estándar Z, de acuerdo al teorema central del límite. Imponiendo que lo anterior es igual a 0,99 y despejando, debe tenerse que

$$\mathbb{P}(Z \geq \sqrt{n}) = 0{,}005,$$

por lo cual $\sqrt{n}$ debe ser similar a 2,58, de acuerdo a la tabla de la normal estándar. Es decir, $n \approx 6{,}65$, luego $n = 7$.

b) (3,0 ptos.) Un productor afirma que al menos el 20 % del público prefiere su producto. Se toma una muestra de 100 personas para verificar su afirmación. Con $\alpha = 0{,}05$, ¿cuál es la mínima cantidad de personas que prefieren el producto de manera que no haya suficiente evidencia para rechazar la afirmación del productor? Si solo 10 personas prefieren el producto, ¿cuál es el resultado del test?

Solución. Denotemos X_i a la variable aleatoria que vale 1 si la i-ésima persona de la muestra prefiere el producto, 0 si no, para $i = 1, \dots, 100$. Cada X_i es una variable Bernoulli con parámetro p desconocido. Definamos las hipótesis

$$H_0 : p = 20\,\%$$
$$H_1 : p < 20\,\%.$$

Notemos que $\mathbb{E}(X_1) = p$ y $\text{var} X_1 = p(1-p)$. Por lo tanto, bajo H_0, sabemos que el estadístico $Z = (\bar{X} - 0{,}2)/\sqrt{0{,}2 \cdot (1 - 0{,}2)/100}$ tiene distribución aproximadamente normal estándar. Dada la forma de las hipótesis, buscamos una región de rechazo del tipo $Z \in (-\infty, c]$, para una cierta constante c por determinar. Imponiendo que la probabilidad del error sea $\alpha = 0{,}05$, tenemos que:

$$0{,}05 = \mathbb{P}(Z \in (-\infty, c] \mid H_0) \approx \mathbb{P}(\mathcal{N}(0,1) \leq c) = \mathbb{P}(\mathcal{N}(0,1) \geq -c).$$

Mirando la tabla de la normal, se obtiene que $-c = 1{,}64$ (también sirve 1,65). Entonces, la hipótesis del productor no se rechaza cuando $Z > c$, es decir:

$$c < Z = \frac{\bar{X} - 0{,}2}{\sqrt{0{,}2 \cdot (1 - 0{,}2)/100}} = 10 \cdot \frac{\bar{X} - 1/5}{\sqrt{1/5 \cdot 4/5}} = 25 \cdot (\bar{X} - 1/5) \iff \bar{X} > \frac{5 - 1{,}64}{25} = \frac{13{,}44}{100}.$$

Notemos que $\bar{X}$ corresponde a $A/100$, donde A es la cantidad de personas que prefieren el producto. Por lo tanto, la mínima cantidad de estas personas tal que no se rechaza la hipótesis del productor es 14. Si 10 personas prefieren el producto, eso es menos que la cantidad mínima recién calculada, por lo tanto se rechaza H_0 y se falla en favor de H_1.

P2. La luminosidad de una estrella, medida por un instrumento adecuado, se modela como una variable aleatoria normal con ambos parámetros desconocidos. Al tomar una muestra de 25 datos usted obtiene un promedio de 204,5 y una desviación estándar muestral de 7,2.

a) (1,5 ptos.) Encuentre un intervalo de confianza para la media al nivel 90 %.

Solución. Denotemos $X_1, \ldots, X_n$ a la muestra, con distribución común $\mathcal{N}(\mu, \sigma^2)$ con ambos parámetros desconocidos. Usamos el estadístico $T = (\bar{X} - \mu)/(S/\sqrt{n})$, el cual posee distribución $t_{n-1} = t_{24}$. Imponemos el nivel deseado a T, usando un intervalo simétrico:

$$90\,\% = \mathbb{P}(-c \leq T \leq c) \;\Rightarrow\; \mathbb{P}(T > c) = 5\,\%,$$

y de la tabla de la t-student se obtiene $c = 1{,}711$. De acá obtenemos el intervalo para μ:

$$\begin{aligned} 90\,\% = \mathbb{P}(-c \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq c) &= \mathbb{P}(\bar{X} - \frac{cS}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{cS}{\sqrt{n}}) \\ &= \mathbb{P}(204{,}5 - \frac{1{,}711 \times 7{,}2}{5} \leq \mu \leq 204{,}5 + \frac{1{,}711 \times 7{,}2}{5}). \end{aligned}$$

b) (1,5 ptos.) Encuentre un intervalo de confianza para la varianza al nivel 90 %.

Solución. Trabajamos con el estadístico $U = (n-1)S^2/\sigma^2$, el cual posee distribución $\chi^2_{n-1} = \chi^2_{24}$. Imponemos el nivel deseado U usando un intervalo con simetría de probabilidad:

$$90\,\% = \mathbb{P}(c \leq U \leq d) \;\Rightarrow\; \mathbb{P}(U > c) = 95\,\%, \quad \mathbb{P}(U > d) = 5\,\%.$$

De una tabla se obtiene $c = 13{,}85$ y $d = 36{,}4$. El intervalo para σ^2 queda:

$$\begin{aligned} 90\,\% = \mathbb{P}(c < \frac{(n-1)S^2}{\sigma^2} < d) &= \mathbb{P}(\frac{(n-1)S^2}{d} < \sigma^2 < \frac{(n-1)S^2}{c}) \\ &= \mathbb{P}(\frac{24 \times 7{,}2^2}{36{,}4} < \sigma^2 < \frac{24 \times 7{,}2^2}{13{,}85}). \end{aligned}$$

c) (1,0 ptos.) Usted se da cuenta que las especificaciones técnicas del instrumento de medición indican que la varianza real de las mediciones es de 49,0. Encuentre un nuevo intervalo de confianza para la media al nivel 90 %.

Solución. Ahora sabemos que $\sigma^2 = 49{,}0$. Usamos el estadístico $Z = (\bar{X} - \mu)/(\sigma/\sqrt{n})$, el cual tiene distribución normal estándar. Como antes, imponemos el nivel deseado a Z usando un intervalo simétrico:

$$90\,\% = \mathbb{P}(-c \leq Z \leq c) \;\Rightarrow\; \mathbb{P}(Z > c) = 5\,\%,$$

y de la tabla de la normal se obtiene $c = 1{,}64$ (también sirve 1,65). De acá obtenemos el intervalo para μ:

$$\begin{aligned} 90\,\% = \mathbb{P}(-c \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq c) &= \mathbb{P}(\bar{X} - \frac{c\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{c\sigma}{\sqrt{n}}) \\ &= \mathbb{P}(204{,}5 - \frac{1{,}64 \times 7}{5} \leq \mu \leq 204{,}5 + \frac{1{,}64 \times 7}{5}). \end{aligned}$$

d) (1,0 ptos.) Usted no está satisfecho con el resultado anterior, y desea obtener un nuevo intervalo cuyo largo sea de a lo más 2,3, manteniendo el nivel de confianza. ¿Cuántas mediciones adicionales necesita?

Solución. Sea m el tamaño de la muestra que incluye a las mediciones previas y las por tomar. Por lo hecho en el ítem anterior sabemos que el largo del intervalo es $2c\sigma/\sqrt{m}$, con el mismo $c = 1{,}64$. Imponemos que el largo sea el deseado:

$$2{,}3 = \frac{2c\sigma}{\sqrt{m}} \;\Rightarrow\; m = \left(\frac{2c\sigma}{2{,}3}\right)^2 = \left(\frac{2 \times 1{,}64 \times 7}{2{,}3}\right)^2 = \left(\frac{22{,}96}{2{,}3}\right)^2 \approx 100.$$

Luego, se necesitan $100 - 25 = 75$ mediciones adicionales.

e) (1,0 ptos.) Usted se da cuenta que no dispone de tiempo para tomar esas mediciones adicionales, y decide trabajar con los 25 datos iniciales. ¿Qué nivel de confianza entrega un intervalo con el largo deseado de 2,3?

Solución. Ahora es c el que debe calcularse en función del largo, usando $n = 25$ datos:

$$2{,}3 = \frac{2c\sigma}{\sqrt{n}} \;\Rightarrow\; c = \frac{2{,}3\sqrt{n}}{2\sigma} = \frac{2{,}3 \times 5}{2 \times 7} = 0{,}8214.$$

De la tabla normal se obtiene $\mathbb{P}(\mathcal{N}(0,1) > c) = 0{,}2061 = \alpha/2$. Luego, $\alpha = 0{,}4122$, y entonces el nivel buscado es $1 - \alpha = 58{,}78\,\%$.

P3. Dadas las constantes $\beta > 0$ conocida y $r > 0$ desconocida, sea X variable aleatoria con densidad

$$f_X(x) = \frac{\beta}{r}\left(\frac{x}{r}\right)^{-(\beta+1)} e^{-(x/r)^{-\beta}} \mathbb{1}_{[0,\infty)}(x).$$

a) (1,2 ptos.) Muestre que $F_X(x) = e^{-(x/r)^{-\beta}} \mathbb{1}_{[0,\infty)}(x)$. Muestre que $X^{-\beta} \sim \exp(r^\beta)$.

Solución. Es directo ver que $e^{-(x/r)^{-\beta}}$ es la primitiva de $\frac{\beta}{r}\left(\frac{x}{r}\right)^{-(\beta+1)} e^{-(x/r)^{-\beta}}$. Sabiendo que $F_X(x) = \int_{-\infty}^{x} f_X(y)dy$, se concluye que $F_X(x) = e^{-(x/r)^{-\beta}} \mathbb{1}_{[0,\infty)}(x)$. Vemos lo otro:

sea $Y = X^{-\beta}$, y usando el cálculo previo, tenemos para $y > 0$:

$$\begin{aligned}\mathbb{P}(Y > y) = \mathbb{P}(X^{-\beta} > y) = \mathbb{P}(X < y^{-1/\beta}) &= F_X(y^{-1/\beta}) \\ &= e^{-(y^{-1/\beta}/r)^{-\beta}} \\ &= e^{-r^\beta y}.\end{aligned}$$

Luego, $F_Y(y) = 1 - \mathbb{P}(Y > y) = 1 - e^{-r^\beta y}$, lo cual significa que $Y \sim \exp(r^\beta)$.

b) (1,2 ptos.) En lo que sigue, sea $X_1, \ldots, X_n$ una m.a.s. proveniente de la distribución de X. Muestre que el estimador de máxima verosimilitud de r es $\hat{r} = (\frac{1}{n}\sum_{i=1}^n X_i^{-\beta})^{-1/\beta}$.

Solución. La función de verosimilitud es:

$$\begin{aligned}L = L(x_1, \ldots, x_n; r) = \prod_{i=1}^n f_X(x_i; r) &= \prod_{i=1}^n \frac{\beta}{r}\left(\frac{x_i}{r}\right)^{-(\beta+1)} e^{-(x_i/r)^{-\beta}} \\ &= \beta^n r^{n\beta} \prod_{i=1}^n x_i^{-(\beta+1)} e^{-(x_i/r)^{-\beta}}.\end{aligned}$$

Luego:

$$\log L = n\log\beta + n\beta\log r - \sum_{i=1}^n [(\beta+1)\log x_i + r^\beta x_i^{-\beta}].$$

Para maximizar lo anterior con respecto a r derivamos e igualamos a 0:

$$0 = \frac{\partial \log L}{\partial r} = \frac{n\beta}{r} - \beta r^{\beta-1}\sum_{i=1}^n x_i^{-\beta},$$

es decir, $0 = n - r^\beta \sum_{i=1}^n x_i^{-\beta}$. Despejando r y reemplazando los x_i por la muestra, obtenemos que el estimador de máxima verosimilitud es $\hat{r} = (\frac{1}{n}\sum_{i=1}^n X_i^{-\beta})^{-1/\beta}$.

c) (1,2 ptos.) Muestre que $\hat{r} \to r$ casi seguramente cuando $n \to \infty$. *Indicación:* defina $Y_i := X_i^{-\beta}$ y observe que $\hat{r}^{-\beta} = \bar{Y}$; luego utilice la parte (a) en conjunto con un teorema adecuado.

Solución. Definimos $Y_i = X_i^{-\beta}$. Por la parte (a) sabemos que $Y_1, \ldots, Y_n$ es una m.a.s. con distribución común $\exp(r^\beta)$. Por la LFGN, se deduce que $1/\hat{r}^\beta = \bar{Y}$ converge casi seguramente a $\mu := \mathbb{E}(Y_1) = \mathbb{E}(\exp(r^\beta)) = 1/r^\beta$ cuando $n \to \infty$. Tomando $(\cdot)^{-1/\beta}$, esto equivale a que $\hat{r} \to r$.

d) (1,2 ptos.) Muestre que el estadístico $Z := (\hat{r}^{-\beta} - r^{-\beta})/(r^{-\beta}/\sqrt{n})$ posee distribución aproximada $\mathcal{N}(0,1)$ cuando n es grande. *Indicación:* recuerde que si $Y \sim \exp(\lambda)$ entonces $\sigma^2 := \mathrm{var}(Y) = 1/\lambda^2$.

Solución. Escribamos Z en términos de la muestra $Y_1, \ldots, Y_n$: sabemos que $\hat{r}^{-\beta} = \bar{Y}$, y como cada Y_i es una variable $\exp(r^\beta)$, también tenemos que $\mu = \mathbb{E}(Y_1) = r^{-\beta}$ y que $\sigma^2 = \mathrm{var}(Y_1) = r^{-2\beta}$. Luego,

$$Z = \frac{\hat{r}^{-\beta} - r^{-\beta}}{r^{-\beta}/\sqrt{n}} = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}},$$

es decir, Z es el estadístico que aparece en el TCL asociado a la muestra $Y_1, \ldots, Y_n$. Para n grande, el TCL asegura que Z posee distribución aproximada $\mathcal{N}(0,1)$.

e) (1,2 ptos.) Utilice el estadístico Z para obtener un intervalo de confianza genérico para r al nivel $1-\alpha$.

Solución. Imponemos el nivel deseado al estadístico Z, el cual posee distribución aproximada $\mathcal{N}(0,1)$, usando un intervalo simétrico respecto al origen:

$$1-\alpha = \mathbb{P}(-c \le Z \le c) \approx \mathbb{P}(-c \le \mathcal{N}(0,1) \le c) = 1 - 2\mathbb{P}(\mathcal{N}(0,1) > c),$$

donde hemos usado la simetría de la distribución $\mathcal{N}(0,1)$. Despejando, se obtiene $\mathbb{P}(\mathcal{N}(0,1) > c) \approx \alpha/2$, y el valor de c se despeja de una tabla o de un software estadístico (en cátedra utilizamos la notación $c \approx z_{\alpha/2}$, pero acá lo dejaremos como c simplemente). Con esto, obtenemos el intervalo buscado para r:

$$\begin{aligned} 1-\alpha \approx \mathbb{P}\left(-c \le \frac{\hat{r}^{-\beta} - r^{-\beta}}{r^{-\beta}/\sqrt{n}} \le c\right) &= \mathbb{P}\left(-\frac{c}{\sqrt{n}} \le \frac{r^{\beta}}{\hat{r}^{\beta}} - 1 \le \frac{c}{\sqrt{n}}\right) \\ &= \mathbb{P}\left(\hat{r}\left(1 - \frac{c}{\sqrt{n}}\right)^{1/\beta} \le r \le \hat{r}\left(1 + \frac{c}{\sqrt{n}}\right)^{1/\beta}\right). \end{aligned}$$