Ingeniería Matemática
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE
MA3403-3 Probabilidades y Estadística, Primavera 2012

Roberto Cortez
Francisco Castro
Alfredo Torrico

Pauta Control 3

P1. Sean $X_1, \dots, X_n$ los resultados de las duraciones de las 16 baterías probadas.

(a) Trabajamos con el estadístico

$$T = \frac{\overline{X} - \mu}{\sqrt{s^2/n}},$$

el cual sabemos que se distribuye como una t_{n-1}, es decir, como una t_{15}. Imponemos que T esté en un intervalo simétrico con probabilidad $1-\alpha$ con $\alpha = 0{,}05$:

$$1 - \alpha = \mathbb{P}(T \in [-c, c]) = 1 - 2\mathbb{P}(T > c), \quad \text{es decir} \quad \mathbb{P}(T > c) = \alpha/2 = 0{,}025.$$

Mirando la tabla de la distribución t-student, obtenemos $c = 2{,}131$. Despejando μ en la inclusión $T \in [-c, c]$, se tiene que

$$\mu \in \left[\overline{X} - c\sqrt{s^2/n}, \overline{X} + c\sqrt{s^2/n}\right] = \left[7 - 2{,}131\frac{\sqrt{0{,}9}}{4}, 7 + 2{,}131\frac{\sqrt{0{,}9}}{4}\right].$$

(b) Trabajamos con el estadístico

$$S = (n-1)\frac{s^2}{\sigma^2},$$

el cual sabemos que se distribuye como una χ^2_{n-1}, es decir, como una χ^2_{15}. Imponemos entonces:

$$1 - \alpha = \mathbb{P}(S \in [c, d]),$$

donde c es tal que $\mathbb{P}(S > c) = 1 - \alpha/2$ y d es tal que $\mathbb{P}(S > d) = \alpha/2$, con $\alpha/2 = 0{,}025$. Mirando la tabla de la distribución chi-cuadrado obtenemos $c = 6{,}26$ y $d = 27{,}5$. Despejando σ^2 en la inclusión $S \in [c, d]$, obtenemos:

$$\sigma^2 \in \left[\frac{(n-1)s^2}{d}, \frac{(n-1)s^2}{c}\right] = \left[\frac{15 \cdot 0{,}9}{27{,}5}, \frac{15 \cdot 0{,}9}{6{,}26}\right].$$

(c) Trabajamos con el estadístico

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}},$$

el cual sabemos que se distribuye como una normal estándar. Imponemos:

$$1 - \alpha = \mathbb{P}(Z \in [-c, c]) = 1 - 2\mathbb{P}(Z > c), \quad \text{es decir} \quad \mathbb{P}(Z > c) = \alpha/2 = 0{,}025.$$

Mirando la tabla de una normal estándar obtenemos $c = 1{,}96$. Por lo tanto, despejando μ en la inclusión $Z \in [-c, c]$, se obtiene

$$\mu \in \left[\overline{X} - c\frac{\sigma}{\sqrt{n}}, \overline{X} + c\frac{\sigma}{\sqrt{n}}\right] = \left[7 - \frac{1{,}96}{4}, 7 + \frac{1{,}96}{4}\right].$$

(d) El nivel de confianza es el mismo si mantenemos el mismo $c = 1{,}96$. Por otro lado, el largo del intervalo anterior es la diferencia de los extremos, es decir

$$2\frac{1{,}96}{\sqrt{n}} = 2\frac{1{,}96}{4} = \frac{1{,}96}{2}.$$

Queremos encontrar un n^* tal que que este intervalo se reduzca un 20 %, para lo cual imponemos entonces que el largo del nuevo intervalo sea un 80 % veces la cantidad anterior, es decir:

$$2\frac{1{,}96}{\sqrt{n^*}} = 0{,}8\frac{1{,}96}{2}, \quad \text{es decir} \quad \sqrt{n^*} = 5.$$

Por lo tanto $n^* = 25$, lo que significa que deben probarse 9 baterías adicionales para reducir el largo del intervalo un 20 %.

P2. (a) Recordemos que la densidad f de una variable Gamma(θ, λ) es

$$f(x) = f(x; \theta, \lambda) = \frac{\lambda^\theta e^{-\lambda x} x^{\theta-1}}{\Gamma(\theta)} \mathbf{1}_{[0,\infty)}(x).$$

Para $x_1, \dots, x_n \geq 0$, el logaritmo de la verosimilitud es entonces

$$\begin{aligned}\log L = \log L(x_1, \dots, x_n; \theta, \lambda) &= \log \prod_{i=1}^n \frac{\lambda^\theta e^{-\lambda x_i} x_i^{\theta-1}}{\Gamma(\theta)} \\ &= \log\left(\frac{\lambda^{n\theta}}{\Gamma(\theta)^n} e^{-\lambda \sum_{i=1}^n x_i} \left[\prod_{i=1}^n x_i\right]^{\theta-1}\right) \\ &= n\theta \log\lambda - n\log\Gamma(\theta) - n\bar{x}\lambda + n(\theta-1)\bar{y},\end{aligned}$$

donde hemos llamado $y_i = \log x_i$. Para maximizar, hacemos gradiente (con respecto a θ y λ) igual a 0. Primero con respecto a λ:

$$0 = \frac{\partial \log L}{\partial \lambda} = \frac{n\theta}{\lambda} - n\bar{x}, \quad \text{es decir,} \quad \lambda = \frac{\theta}{\bar{x}}.$$

Utilizando esto en la ecuación que se obtiene al derivar con respecto a θ, tenemos:

$$\begin{aligned}0 = \frac{\partial \log L}{\partial \theta} &= n\log\lambda - \frac{n\Gamma'(\theta)}{\Gamma(\theta)} + n\bar{y} \\ &= n\log\theta - n\log\bar{x} - \frac{n\Gamma'(\theta)}{\Gamma(\theta)} + n\bar{y},\end{aligned}$$

de donde se obtiene $\Gamma'(\theta)/\Gamma(\theta) - \log\theta = \bar{y} - \log\bar{x}$. Reemplazando $x_1, \dots, x_n$ por la muestra $X_1, \dots, X_n$ (con los correspondientes $Y_i = \log X_i$), obtenemos que los estimadores de máxima verosimilitud de θ y λ satisfacen lo deseado, es decir,

$$\frac{\Gamma'(\hat{\theta}_{\text{mv}})}{\Gamma(\hat{\theta}_{\text{mv}})} - \log(\hat{\theta}_{\text{mv}}) = \overline{Y} - \log\overline{X} \quad \text{y} \quad \hat{\lambda}_{\text{mv}} = \frac{\hat{\theta}_{\text{mv}}}{\overline{X}}.$$

(b) Si $X \sim$ Gamma(θ, λ), sabemos que

$$\mathbb{E}(X) = \frac{\theta}{\lambda} \quad \text{y} \quad \mathbb{E}(X^2) = \text{var}(X) + \mathbb{E}(X)^2 = \frac{\theta}{\lambda^2} + \frac{\theta^2}{\lambda^2} = \frac{\theta(1+\theta)}{\lambda^2}.$$

Para encontrar los estimadores del método de los momentos, imponemos que los momentos muestrales m_1 y m_2 coincidan con los momentos reales de la variable. Para m_1, tenemos entonces:

$$m_1 = \frac{\theta}{\lambda}, \quad \text{es decir,} \quad \lambda = \frac{\theta}{m_1}.$$

Utilizando esto en la ecuación que se obtiene para m_2, tenemos:

$$m_2 = \frac{\theta(1+\theta)}{\lambda^2} = \frac{\theta(1+\theta)}{\theta^2/m_1^2} = \frac{m_1^2(1+\theta)}{\theta},$$

y despejando θ, obtenemos

$$\theta = \left(\frac{m_2}{m_1^2} - 1\right)^{-1} = \frac{m_1^2}{m_2 - m_1^2}.$$

Recordando que $\lambda = \theta/m_1$, obtenemos finalmente los estimadores del método de los momentos:

$$\hat{\theta}_{\text{mom}} = \frac{m_1^2}{m_2 - m_1^2} \quad \text{y} \quad \hat{\lambda}_{\text{mom}} = \frac{m_1}{m_2 - m_1^2}.$$

(c) Por LFGN, sabemos que $m_1 = \overline{X} \to \mathbb{E}(X_1) = \theta/\lambda$ casi seguramente cuando $n \to \infty$. Del mismo modo, definiendo $Y_i = X_i^2$, por LFGN se tiene que $m_2 = \overline{Y} \to \mathbb{E}(Y_1) = \theta(1+\theta)/\lambda^2$ casi seguramente cuando $n \to \infty$. Por estas convergencias, se tiene entonces que

$$\hat{\theta}_{\text{mom}} = \frac{m_1^2}{m_2 - m_1^2} \to \frac{(\theta/\lambda)^2}{\theta(1+\theta)/\lambda^2 - (\theta/\lambda)^2} = \frac{(\theta/\lambda)^2}{\theta/\lambda^2} = \theta$$

y

$$\hat{\lambda}_{\text{mom}} = \frac{m_1}{m_2 - m_1^2} \to \frac{\theta/\lambda}{\theta(1+\theta)/\lambda^2 - (\theta/\lambda)^2} = \frac{\theta/\lambda}{\theta/\lambda^2} = \lambda$$

casi seguramente cuando $n \to \infty$, como deseábamos demostrar.

P3. (a) Llamemos $X_1, \ldots, X_n$, con $n = 200$, a las variables Bernoulli(p) que representan si cada persona renovó suscripción el último año, y definamos $\hat{p} = \overline{X}$. Sean

$$H_0 : p = p_0$$
$$H_1 : p < p_0,$$

con $p_0 = 60\,\%$. Trabajamos con el estadístico

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}},$$

el cual sabemos que, bajo H_0, tiene distribución aproximadamente normal estándar, por el TCL. Dada la forma de las hipótesis, buscamos una región de rechazo determinada por la desigualdad $Z \leq c$, donde c es una constante por encontrar. Imponiendo que la probabilidad del error tipo I sea igual a $\alpha = 2{,}5\,\%$, tenemos:

$$2{,}5\,\% = \mathbb{P}(Z \leq c \mid H_0) \approx \mathbb{P}(\mathcal{N}(0,1) \leq c) = \mathbb{P}(\mathcal{N}(0,1) \geq -c),$$

donde en el último paso hemos utilizado la simetría de la distribución normal estándar. Mirando la tabla normal, obtenemos $c = -1{,}96$. Esto quiere decir que la hipótesis nula se rechaza y se falla en favor del editor si y sólo si el valor observado del estadístico Z es menor o igual que c. Como el valor observado de $\hat{p}$ es $108/200 = 54\,\%$, tenemos:

$$Z_{\text{obs}} = \frac{54\,\% - 60\,\%}{\sqrt{60\,\% \times 40\,\%/200}} = \frac{-6\,\%}{\sqrt{1200/100^3}} = \frac{-6}{\sqrt{12}} = -\sqrt{3}.$$

Es decir, $Z_{\text{obs}} \approx -1{,}73 > -1{,}96$. Por lo tanto, no se rechaza la afirmación del dueño de la revista.

(b) Denotemos $X_1, \ldots, X_n$ los pesos de las cargas de los clientes, y sea $X = \sum_{i=1}^n X_i$ el peso total. Denotemos $C = 6000$ la capacidad del avión y $M = 100 = \mathbb{E}(X_1)$.

1) Utilizando la desigualdad de Markov, tenemos que

$$\mathbb{P}(X \geq C) \leq \frac{\mathbb{E}(X)}{C} = \frac{nM}{C}.$$

Luego, para asegurar que la probabilidad de sobrepasar la capacidad del avión sea a lo más de un $2{,}28\,\%$, basta pedir que se cumpla la desigualdad $nM/C \leq 2{,}28\,\%$. Es decir,

$$n \leq \frac{2{,}28C}{100M} = \frac{2{,}28 \times 60}{100} = 1{,}368.$$

Entonces, de acuerdo a este procedimiento, bajo el requisito que la probabilidad de sobrepase la carga del avión sea a lo más de $2{,}28\,\%$, sólo es posible llevar la carga de $n^* = 1$ cliente.

2) Llamemos $\sigma = 200$. Para utilizar el TCL, armamos el estadístico adecuado:

$$\mathbb{P}(X \geq C) = \mathbb{P}\left(\frac{\overline{X} - M}{\sigma/\sqrt{n}} \geq \frac{C/n - M}{\sigma/\sqrt{n}}\right) \approx \mathbb{P}\left(\mathcal{N}(0,1) \geq \frac{C/n - M}{\sigma/\sqrt{n}}\right),$$

donde en el último paso hemos utilizado el TCL sobre el estadístico a la izquierda de la desigualdad. Luego, para asegurar que $\mathbb{P}(X \geq C) \leq 2{,}28\,\%$, basta pedir que la cantidad $(C/n - M)/(\sigma/\sqrt{n})$ sea mayor que z, donde z satisface $\mathbb{P}(\mathcal{N}(0,1) \geq z) = 2{,}28\,\%$, es decir, $z = 2$ (ver tabla normal). Por lo tanto, imponemos que

$$\frac{C/n - M}{\sigma/\sqrt{n}} \geq z, \quad \text{es decir,} \quad C - Mn - \sigma z\sqrt{n} \geq 0.$$

Definiendo la variable $x = \sqrt{n}$, obtenemos $-Mx^2 - \sigma z x + C \geq 0$, lo cual es una inecuación cuadrática. Esta desigualdad se cumplirá para todo x que esté entre las raíces de la ecuación $-Mx^2 - \sigma z x + C = 0$, las cuales vienen dadas por

$$x_{1,2} = \frac{\sigma z \pm \sqrt{\sigma^2 z^2 + 4MC}}{-2M} = \frac{400 \pm \sqrt{160000 + 2400000}}{-200} = \frac{400 \pm 1600}{-200} = -10 \text{ y } 6.$$

Como nos interesa que $\sqrt{n}$ sea lo más grande posible, pedimos entonces que x sea lo más cercano a la raíz mayor, es decir, $x = 6$. Esto significa que puede llevarse la carga de hasta $n^* = 36$ clientes y se sigue cumpliendo la restricción deseada.