Ingeniería Matemática
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE
MA3403-5 Probabilidades y Estadística
Otoño 2013

Roberto Cortez
Francisco Castro
Alfredo Torrico

Pauta control 3

P1. (a) Numeremos los papeles adheridos de 1 a n, y sean $X_1, \dots, X_n$ variables aleatorias dadas por

$$X_i = \begin{cases} 1 & \text{si el } i\text{-ésimo papel es rojo} \\ 0 & \text{si es blanco.} \end{cases}$$

Dado que la cantidad total de papeles que había en la fiesta era grande, es razonable suponer que los X_i son independientes y con ley Bernoulli de parámetro p igual a la probabilidad de que un papel sea rojo. Como los papeles rojos provienen de la caja cúbica, se tiene que $p = C/(C+E)$, siendo C el volumen del cubo y E el volumen de la esfera. Además, notemos que $r = \sum_{i=1}^n X_i = n\bar{X}$, y que $b = n - r = n(1-\bar{X})$, con lo cual tenemos:

$$\frac{6b}{r} = 6\frac{n(1-\bar{X})}{n\bar{X}} = 6\left(\frac{1}{\bar{X}} - 1\right) \to 6\left(\frac{1}{p} - 1\right),$$

donde en el último paso hemos utilizado que $\bar{X}$ converge casi seguramente a $\mathbb{E}(X_1) = p$ cuando $n \to \infty$, gracias a la ley fuerte de los grandes números. Reemplazando el valor de $p = C/(C+E)$ y anotando R al radio de la esfera (luego el lado del cubo es $2R$), tenemos:

$$\frac{6b}{r} \to 6\left(\frac{C+E}{C} - 1\right) = 6\frac{E}{C} = 6\frac{\frac{4}{3}\pi R^3}{(2R)^3} = \pi.$$

Es decir, suponiendo que n es relativamente grande (digamos, mayor que 30), se tiene que $6b/r$ debería parecerse a π, como deseábamos mostrar.

(b) Los X_i siguen una $\mathcal{N}(\mu, \sigma^2)$, con ambos parámetros desconocidos. Planteamos las hipótesis

$$H_0 : \mu = 7$$
$$H_1 : \mu < 7.$$

Trabajamos con el estadístico

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}},$$

el cual se distribuye como una t-student con $n - 1 = 24$ grados de libertad. Sabiendo que $\sum X_i = 172{,}508$ y $s = \sqrt{0{,}04} = 0{,}2 = 1/5$, obtenemos que el valor observado del estadístico T bajo la hipótesis H_0 es:

$$T_{\text{obs}} = \frac{\frac{\sum X_i}{25} - 7}{\frac{1/5}{\sqrt{25}}} = \frac{\frac{\sum X_i}{25} - 7}{1/25} = \sum X_i - 25 \times 7 = 172{,}508 - 175 = -2{,}492.$$

El p-valor del test corresponde a la probabilidad, bajo H_0, de obtener un valor al menos tan extremo como el de la muestra. Usando la simetría de la distribución t-student y mirando una tabla, obtenemos:

$$p\text{-valor} = \mathbb{P}(T \le T_{\text{obs}} \mid H_0) = \mathbb{P}(t_{24} \le -2{,}492) = \mathbb{P}(t_{24} \ge 2{,}492) = 0{,}01$$

Es decir, el p-valor es de $1\,\%$. Para $\alpha = 5\,\%$, esto significa que debemos rechazar H_0, con lo cual se concluye que la máquina efectivamente está produciendo piezas con grosor inferior a 7 cm.

P2. (a) 1) Calculemos el estimador de máxima verosimilitud $\hat{\theta}_{\text{mv}}$. La función de verosimilitud es

$$L(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} (\theta+1) x_i^{\theta} \mathbf{1}_{[0,1]}(x_i) = (\theta+1)^n \left[\prod_{i=1}^{n} x_i \mathbf{1}_{[0,1]}(x_i)\right]^{\theta}.$$

Como cada x_i correponderá a una realización de la variable aleatoria, se tendrá que x_i siempre pertenece al intervalo $[0,1]$, de modo que las indicatrices anteriores son siempre 1. Tomando logaritmo de L y derivando con respecto a θ e igualando a 0 para maximizar, tenemos:

$$\begin{aligned} 0 &= \frac{d \log L}{d\theta} \\ &= \frac{d}{d\theta}\left[n \log(1+\theta) + \theta \sum_{i=1}^{n} \log x_i\right] \\ &= \frac{n}{1+\theta} + \sum_{i=1}^{n} \log x_i. \end{aligned}$$

Despejando θ y reemplazando los X_i, obtenemos $\hat{\theta}_{\text{mv}} = -(1 + n/\sum_{i=1}^{n} \log X_i)$. Veamos el estimador de los momentos $\hat{\theta}_{\text{mom}}$, para lo cual necesitamos calcular la esperanza de distribución común de los X_i:

$$\mathbb{E}(X_1) = \int_{-\infty}^{\infty} x f(x) dx = (\theta+1) \int_0^1 x^{\theta+1} dx = \frac{\theta+1}{\theta+2}.$$

Imponiendo la ecuación $\bar{X} = \mathbb{E}(X_1) = (\theta+1)/(\theta+2)$ y despejando θ, obtenemos $\hat{\theta}_{\text{mom}} = (1-2\bar{X})/(\bar{X}-1)$.

2) Por la ley fuerte de los grandes números, sabemos que $\bar{X}$ converge casi seguramente a $\mathbb{E}(X_1) = (\theta+1)/(\theta+2)$ cuando $n \to \infty$. Por lo tanto:

$$\hat{\theta}_{\text{mom}} = \frac{1-2\bar{X}}{\bar{X}-1} \to \frac{1 - 2\frac{\theta+1}{\theta+2}}{\frac{\theta+1}{\theta+2} - 1} = \frac{\theta+2-2(\theta+1)}{\theta+1-(\theta+2)} = \theta,$$

lo cual prueba la convergencia deseada para el estimador de los momentos. Para probar la convergencia del estimador de máxima verosimilitud, definamos $Y_i = \log X_i$ y calculemos $\mathbb{E}(Y_1)$:

$$\mathbb{E}(Y_1) = \int_0^1 \log(x)(1+\theta) x^{\theta} dx = \left[\log(x) x^{\theta+1}\Big|_0^1 - \int_0^1 \frac{x^{\theta+1}}{x} dx\right] = -\frac{1}{\theta+1}.$$

Además, notemos que $\hat{\theta}_{\text{mv}} = -(1 + 1/\bar{Y})$. Por la ley fuerte de los grandes números sabemos que $\bar{Y}$ converge casi seguramente a $\mathbb{E}(Y_1) = -1/(\theta+1)$ cuando $n \to \infty$, con lo cual se concluye lo buscado:

$$\hat{\theta}_{\text{mv}} \to -\left(1 + \frac{1}{-\frac{1}{\theta+1}}\right) = -(1-(\theta+1)) = \theta.$$

(b) 1) Utilizamos el estadístico

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}},$$

el cual sabemos que tiene distribución t-student con $n-1=24$ grados de libertad. Para encontrar el intervalo buscado, imponemos que T esté en un intervalo simétrico:

$$95\,\% = \mathbb{P}(T \in [-c,c]) = 1 - \mathbb{P}(|T| > c) = 1 - 2\mathbb{P}(T > c),$$

es decir, $\mathbb{P}(T > c) = 2{,}5\,\%$. De una tabla de una t-student, obtenemos $c = 2{,}064$. Despejando μ en la inclusión $T \in [-c,c]$, tenemos:

$$\begin{aligned} & -c \leq \frac{\bar{X}-\mu}{s/\sqrt{n}} \leq c \\ \Leftrightarrow \quad & \bar{X} - \frac{cs}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{cs}{\sqrt{n}} \end{aligned}$$

Notando que $s^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2 = \frac{2400}{24} = 100$, se tiene que $cs/\sqrt{n} = 2{,}064 \times 10/\sqrt{25} = 4{,}128$. Con esto, el intervalo buscado es:

$$\left[\bar{X} - \frac{cs}{\sqrt{n}}, \bar{X} + \frac{cs}{\sqrt{n}}\right] = [65 - 4{,}128, \quad 65 + 4{,}128] = [60{,}872, \quad 69{,}128]$$

2) Utilizamos el estadístico

$$V = (n-1)\frac{s^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2,$$

el cual sabemos que posee distribución chi-cuadrado con $n-1=24$ grados de libertad. Imponemos $95\,\% = \mathbb{P}(V \in [a,b])$, donde a y b satisfacen simetría de probabilidad, es decir $\mathbb{P}(V < a) = \mathbb{P}(V > b)$. Por lo tanto $\mathbb{P}(V \geq a) = 97{,}5\,\%$ y $\mathbb{P}(V > b) = 2{,}5\,\%$, y de la tabla se obtiene $a = 12{,}4$ y $b = 39{,}4$. Despejando σ^2 en la inclusión $V \in [a,b]$, tenemos:

$$\begin{aligned} & a \leq (n-1)\frac{s^2}{\sigma^2} \leq b \\ \Leftrightarrow \quad & \frac{(n-1)s^2}{b} \leq \sigma^2 \leq \frac{(n-1)s^2}{a}. \end{aligned}$$

Notando que $(n-1)s^2 = \sum(X_i - \bar{X})^2 = 2400$, el intervalo de confianza para σ^2 es

$$\left[\frac{2400}{39{,}4}, \quad \frac{2400}{12{,}4}\right] \approx [60, \quad 200].$$

3) El largo del intervalo obtenido para σ^2 es la diferencia de sus extremos, es decir, $(n-1)s^2[\frac{1}{a} - \frac{1}{b}]$. Luego, el valor esperado del largo es

$$\mathbb{E}\left((n-1)s^2\left[\frac{1}{a} - \frac{1}{b}\right]\right) = (n-1)\left[\frac{1}{a} - \frac{1}{b}\right]\mathbb{E}(s^2) = 24 \times \left[\frac{1}{12{,}4} - \frac{1}{39{,}4}\right]\sigma^2,$$

donde en el último paso hemos utilizado el hecho que s^2 es un estimador insesgado de σ^2.

P3. (a) Como $T = n\bar{X}$, es claro que $\mathbb{E}(T) = n\mu = 8n$. Utilizando la desigualdad de Markov, tenemos:

$$\mathbb{P}(T \geq E) \leq \frac{\mathbb{E}(T)}{E} = \frac{8n}{112} = \frac{n}{14}$$

Si queremos que $1-\alpha \leq \mathbb{P}(T \geq E)$, lo anterior implica que $1-\alpha \leq n/14$, es decir, $n \geq 14 \times (1-\alpha) \geq 13$. Por lo tanto, necesariamente $n^* \geq 13$ (ó 14).

(b) Armaremos el estadístico $Z = (\bar{X} - \mu)/(\sigma/\sqrt{n})$, el cual sabemos que tiene distribución aproximadamente normal estándar, por el TCL. Imponiendo que $T \geq E$ ocurra con la probabilidad deseada, tenemos:

$$1 - \alpha = \mathbb{P}(n\bar{X} \geq E) = \mathbb{P}\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \geq \frac{\frac{E}{n} - \mu}{\sigma/\sqrt{n}}\right) \approx \mathbb{P}\left(\mathcal{N}(0,1) \geq \frac{\frac{E}{n} - \mu}{\sigma/\sqrt{n}}\right).$$

Pasando al complemento, obtenemos:

$$2{,}28\,\% = \alpha \approx \mathbb{P}\left(\mathcal{N}(0,1) < \frac{\frac{E}{n} - \mu}{\sigma/\sqrt{n}}\right) = \mathbb{P}\left(\mathcal{N}(0,1) > -\frac{\frac{E}{n} - \mu}{\sigma/\sqrt{n}}\right),$$

donde en el último paso utilizamos la simetría de la normal. Mirando una tabla, obtenemos que la cantidad a la derecha de la última desigualdad debe ser aproximadamente igual a $c = 2$. Es decir, buscamos n tal que

$$c = -\frac{\frac{E}{n} - \mu}{\sigma/\sqrt{n}},$$

lo cual es una ecuación cuadrática en la variable $x = \sqrt{n}$. Explícitamente, la ecuación es $\mu x^2 - \sigma c x - E = 0$, es decir $8x^2 - 4x - 112 = 0$, o equivalentemente: $2x^2 - x - 28 = 0$. Las soluciones son

$$x_{1,2} = \frac{1 \pm \sqrt{1 + 4 \times 2 \times 28}}{4} = \frac{1 \pm \sqrt{225}}{4}.$$

Como x es una raíz, sólo es válida la solución positiva, es decir, $x = 4$, con lo cual obtenemos $n^* = x^2 = 16$.

(c) Debemos imponer que $150\,\% \times T \leq S$ ocurra con probabilidad $1 - \alpha$. La idea es nuevamente armar un estadístico aproximadamente normal estándar. Sea $\gamma = 150\,\% = 3/2$.

$$1 - \alpha = \mathbb{P}(\gamma T \leq S) = \mathbb{P}\left(\sum_{i=1}^{m} \gamma X_i \leq \sum_{i=1}^{m} Y_i\right) = \mathbb{P}\left(\sum_{i=1}^{m} (\gamma X_i - Y_i) \leq 0\right)$$

Llamando $Z_i = \gamma X_i - Y_i$, lo anterior equivale a $1 - \alpha = \mathbb{P}(m\bar{Z} \leq 0)$. La esperanza y varianza de los Z_i corresponden a:

$$\lambda := \mathbb{E}(Z_1) = \gamma\mu - \nu = \frac{3}{2} \times 8 - 14 = -2$$
$$\rho^2 := \mathrm{var}(Z_1) = \gamma^2\sigma^2 + \tau^2 = \frac{9}{4} \times 4 + 16 = 25,$$

donde en el cálculo de la varianza hemos utilizado la independencia entre X_i e Y_i. Con esto, podemos armar el estadístico deseado en base a los Z_i:

$$1 - \alpha = \mathbb{P}\left(\frac{\bar{Z} - \lambda}{\rho/\sqrt{m}} \leq \frac{-\lambda}{\rho/\sqrt{m}}\right) \approx \mathbb{P}\left(\mathcal{N}(0,1) \leq \frac{-\lambda}{\rho/\sqrt{m}}\right) = 1 - \mathbb{P}\left(\mathcal{N}(0,1) > \frac{-\lambda}{\rho/\sqrt{m}}\right).$$

Es decir, $2{,}28\,\% \approx \mathbb{P}(\mathcal{N}(0,1) > -\lambda\sqrt{m}/\rho)$. Como antes, de la tabla obtenemos que $-\lambda\sqrt{m}/\rho$ debe ser aproximadamente igual a $c = 2$, es decir, $-\lambda\sqrt{m} = \rho c$. Despejando m y reemplazando los valores de las constantes, obtenemos el m^* buscado:

$$m^* = \left(\frac{\rho c}{-\lambda}\right)^2 = \left(\frac{5 \times 2}{2}\right)^2 = 25.$$