

IDEAS C. AUXILIAR #4

Los temas a desarrollar están contenidos en las secciones 2.11-2.13 del apunte, relacionadas con los intervalos de confianza, en especial, con el método del pivote.

1. Sea $\mathbf{x}$ generado por un vector $\mathbf{X} = (X_1, \dots, X_n)$ de variables aleatorias X_i iid, con ley uniforme en el intervalo $[\theta, \theta + 1]$, donde $\theta \in \Theta \subseteq (0, \infty)$ es un parámetro desconocido.

La densidad del vector $\mathbf{X}$ está dada por

$$f_{\theta}(\mathbf{x}) = \prod_{i=1}^n \mathbf{1}_{\{\theta \leq x_i \leq \theta+1\}} = \mathbf{1}_{\{\theta \leq \min\{x_i\} \leq \max\{x_i\} \leq \theta+1\}}.$$

- a) Compruebe que las siguientes funciones son pivotes para $g(\theta) = \theta$: $T_1(\mathbf{x}, \theta) = \sum_{i=1}^n x_i - n\theta$, $T_2(\mathbf{x}, \theta) = x_{1:n} - \theta = \min\{x_1, \dots, x_n\} - \theta$ y $T_3(\mathbf{x}, \theta) = x_{n:n} - \theta = \max\{x_1, \dots, x_n\} - \theta$.

Sol: Basta comprobar que las v.a. $X_i - \theta$ son iid uniformes en $[0, 1]$. En primer lugar escribimos la función de distribución F_{θ} de X_i .

$$F_{\theta}(x) = \mathbb{P}_{\theta}(X_i \leq x) = \begin{cases} 0 & \text{si } x < \theta \\ x - \theta & \text{si } \theta \leq x \leq \theta + 1 \\ 1 & \text{si } x > \theta + 1. \end{cases}$$

Por lo tanto, la distribución de $X_i - \theta$ está dada por

$$\mathbb{P}_{\theta}(X_i - \theta \leq y) = \mathbb{P}_{\theta}(X_i \leq y + \theta) = \begin{cases} 0 & \text{si } y < 0 \\ y & \text{si } 0 \leq y \leq 1 \\ 1 & \text{si } y > 1. \end{cases}$$

Es decir, $X_i - \theta$ es uniforme en $[0, 1]$. Esto implica que $T_1(\mathbf{X}) = \sum_{i=1}^n (X_i - \theta)$ se distribuye como la suma de n v.a. iid $U[0, 1]$ (su ley no depende de θ) y, además, es obviamente función decreciente de $g(\theta) = \theta$.

Por otra parte, $T_2(\mathbf{X}, \theta) = X_{1:n} - \theta = \min\{X_1 - \theta, \dots, X_n - \theta\}$ se distribuye como mínimo de v.a. iid $U[0, 1]$ y también es función decreciente de θ .

Finalmente, $T_3(\mathbf{X}, \theta) = X_{n:n} - \theta = \max\{X_1 - \theta, \dots, X_n - \theta\}$ se distribuye como máximo de v.a. iid $U[0, 1]$ y es función decreciente de θ .

Por lo anteriormente expuesto, las funciones $T_i, i = 1, 2, 3$ son pivotes porque cumplen con las dos condiciones de la definición.

- b) Calcule intervalos de confianza para θ usando los pivotes del apartado anterior. Deje sus resultados de la forma “más explícita posible”.

Sol: Comenzamos con T_1 , reconociendo que no hay una expresión cerrada para la distribución de la suma de $U[0, 1]$ iid, aunque (si n es “grande”) podríamos usar una aproximación gaussiana, gracias al TCL.

Buscamos $z_{\alpha,1}, z_{\alpha,2}$ tales que

$$\mathbb{P}_{\theta} \left(z_{\alpha,1} \leq \sum_{i=1}^n (X_i - \theta) \leq z_{\alpha,2} \right) = F_n(z_{\alpha,2}) - F_n(z_{\alpha,1}) \geq 1 - \alpha,$$

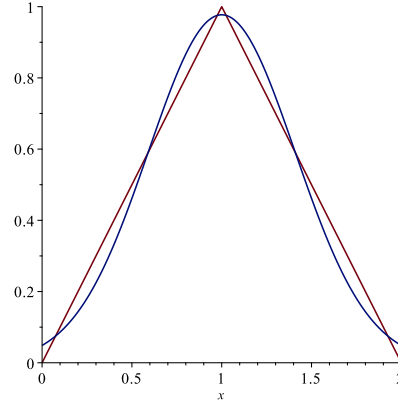


Figura 1: Densidad de suma de dos v.a. iid $U[0, 1]$ y la normal asociada.

donde F_n es la función de distribución de la suma de v.a. iid $U[0, 1]$ y $\alpha \in (0, 1)$. La función F_n es continua y estrictamente creciente en $[0, n]$ de manera que (usando el mismo argumento que en el primer ejemplo del apunte, con la normal) nos aseguramos que los z existen. Entonces podemos despejar θ de la desigualdad para encontrar los extremos del intervalo.

$$z_{\alpha,1} \leq \sum_{i=1}^n (X_i - \theta) \leq z_{\alpha,2} \Leftrightarrow \bar{X} - \frac{z_{\alpha,2}}{n} \leq \theta \leq \bar{X} - \frac{z_{\alpha,1}}{n}.$$

Es decir, el intervalo es

$$I(\mathbf{x}) = \left[\bar{x} - \frac{z_{\alpha,2}}{n}, \bar{x} - \frac{z_{\alpha,1}}{n} \right].$$

Respecto a la longitud del intervalo tenemos $l(\mathbf{x}) = \frac{1}{n}(z_{\alpha,2} - z_{\alpha,1})$ y podríamos plantear el problema de minimización de largo

$$\text{mín } z_{\alpha,2} - z_{\alpha,1} \quad \text{s.a.} \quad \int_{z_{\alpha,1}}^{z_{\alpha,2}} f_n(x) dx = 1 - \alpha,$$

donde $f_n(x) = \frac{dF_n}{dx}(x)$ es la densidad asociada a F_n , para la cual existe una fórmula “cerrada”, dada¹ por

$$f_n(x) = \frac{1}{(n-1)!(b-a)^n} \sum_{i=0}^{\lfloor \frac{x-na}{b-a} \rfloor} (-1)^i \binom{n}{i} (x-na-i(b-a))^{n-1} \mathbf{1}_{\{na \leq x \leq nb\}},$$

para v.a. uniformes en $[a, b]$. En el caso particular que nos interesa ($a = 0, b = 1$) la fórmula se simplifica y obtenemos

$$f_n(x) = \frac{1}{(n-1)!} \sum_{i=0}^{\lfloor x \rfloor} (-1)^i \binom{n}{i} (x-i)^{n-1} \mathbf{1}_{\{0 \leq x \leq n\}}.$$

Es interesante notar que la densidad f_n es simétrica con respecto a $n/2$, es decir, que $f_n(x) =$

¹Ver A. Rényi (1970) Probability Theory. North-Holland, Amsterdam, y F. Killmann and E. von Collani, A Note on the Convolution of the Uniform and Related Distributions and Their Use in Quality Control. Economic Quality Control, Vol 16 (2001), No. 1, 17 – 41.

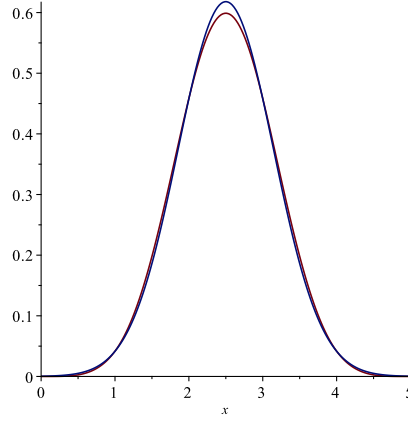


Figura 2: Densidad de suma de 5 v.a. iid $U[0, 1]$ y la normal asociada.

$f_n(n - x)$, para todo $x \in [0, n]$. Esto se puede comprobar viendo que la función característica de la suma centrada (restando $n/2$) es real.

Recordemos que la f.c. de una suma de uniformes $[0, 1]$ iid está dada por

$$\int_0^n e^{itx} f_n(x) dx = \left(\frac{e^{it} - 1}{it} \right)^n$$

Por otra parte, la f.c. de la suma centrada es

$$e^{-\frac{itn}{2}} \left(\frac{e^{it} - 1}{it} \right)^n = \left(\frac{e^{\frac{it}{2}} - e^{-\frac{it}{2}}}{2i(t/2)} \right)^n = \left(\frac{\sin(t/2)}{t/2} \right)^n \in \mathbb{R}.$$

Ahora podemos planear la búsqueda del intervalo de confianza más corto, buscando (por método de Lagrange) puntos críticos de $g(x, y) = y - x - \lambda(F_n(y) - F_n(x) - (1 - \alpha))$. Esto conduce a las ecuaciones

$$-1 + \lambda f_n(x) = 0, \quad 1 - \lambda f_n(y) = 0, \quad , F_n(y) - F_n(x) - (1 - \alpha) = 0.$$

Lo anterior implica $f_n(x) = f_n(y)$, lo cual, debido a la simetría de f_n , se reduce a (suponemos $\alpha < 1/2$)

$$z_{\alpha,1} = F_n^{-1}(\alpha/2) < \frac{n}{2}, z_{\alpha,2} = n - z_{\alpha,1} = F_n^{-1}(1 - \alpha/2) > \frac{n}{2}.$$

Entonces, el intervalo más corto de nivel $100(1 - \alpha)\%$ se escribe como

$$I(\mathbf{x}) = \left[\bar{x} - 1 + \frac{F_n^{-1}(\alpha/2)}{n}, \bar{x} - \frac{F_n^{-1}(\alpha/2)}{n} \right].$$

Ahora, para resolver encontrar intervalos concretos (numéricos) hay que resolver numéricamente la ecuación $F_n(z_{\alpha,1}) = \alpha/2$. Por ejemplo, para $n = 5, \alpha = 0,05$ encontramos $F_5^{-1}(0,025) \sim 1,247$, lo que lleva al intervalo

$$I(\mathbf{x}) = \left[\bar{x} - 1 + \frac{1,247}{5}, \bar{x} - \frac{1,247}{5} \right] = [\bar{x} - 0,751, \bar{x} - 0,249]$$

cuyo largo es $0,751 - 0,249 = 0,502$. Simulando 5 datos $U[10, 11]$ entrega $\bar{x} = 10,34$ y llegamos al intervalo

$$[9,59, 10,09],$$

que contiene al verdadero $\theta = 10$.

Los intervalos para los otros pivotes se obtienen de manera completamente análoga pero teniendo presente que ya no hay simetría de la densidad del estadístico usado.

Para T_2 resulta

$$z_{\alpha,1} \leq X_{1:n} - \theta \leq z_{\alpha,2} \Leftrightarrow X_{1:n} - z_{\alpha,2} \leq \theta \leq X_{1:n} - z_{\alpha,1}$$

y para T_3 se obtiene

$$z_{\alpha,1} \leq X_{n:n} - \theta \leq z_{\alpha,2} \Leftrightarrow X_{n:n} - z_{\alpha,2} \leq \theta \leq X_{n:n} - z_{\alpha,1}.$$

En cada caso los $z_{\alpha,i}$ se calculan usando la distribución F_n del mínimo o máximo de v.a. iid $U[0, 1]$. Estas distribuciones son muy sencillas. Por ejemplo, para el máximo es $F_n(x) = x^n$, $x \in [0, 1]$ y debemos resolver $z_{\alpha,2}^n - z_{\alpha,1}^n = 1 - \alpha$.

Si interesa obtener el intervalo más corto resolvemos el problema de optimización

$$\min z_{\alpha,2} - z_{\alpha,1} \quad \text{s.a.} \quad z_{\alpha,2}^n - z_{\alpha,1}^n = 1 - \alpha,$$

que se puede reformular como $\min_x \sqrt[n]{x^n + 1 - \alpha} - x$, con $0 \leq x \leq \sqrt[n]{\alpha}$. Esta función es decreciente y alcanza su mínimo en $x = \sqrt[n]{\alpha}$. Es decir, los valores óptimos para el intervalo son $z_{\alpha,1} = \sqrt[n]{\alpha}$, $z_{\alpha,2} = 1$.

El intervalo se escribe

$$I(\mathbf{x}) = [x_{n:n} - 1, x_{n:n} - \sqrt[n]{\alpha}].$$

Por ejemplo, con $n = 5$, $\alpha = 0,05$ y los mismos datos del intervalo anterior, tenemos $x_{n:n} = 10,8$, $\sqrt[5]{0,05} = 0,549$, llegando a

$$I(\mathbf{x}) = [10,8 - 1, 10,8 - 0,549] = [9,8, 10,25].$$

Otra posibilidad (subóptima) es definir $z_{\alpha,1} = F_n^{-1}(\alpha/2) = \sqrt[n]{\alpha/2} = \sqrt[5]{0,025} = 0,478$ y $z_{\alpha,2} = F_n^{-1}(1 - \alpha/2) = \sqrt[n]{1 - \alpha/2} = \sqrt[5]{0,975} = 0,995$. Entonces el intervalo queda como

$$I(\mathbf{x}) = [x_{n:n} - \sqrt[n]{1 - \alpha/2}, x_{n:n} - \sqrt[n]{\alpha/2}] = [10,8 - 0,995, 10,8 - 0,478] = [9,805, 10,32],$$

siendo la longitud igual a 0,515, un poco más largo que el óptimo, que tiene longitud $1 - \sqrt[n]{\alpha} = 0,45$.

2. Sean $X_1, X_2, \dots, X_n$ una MAS con distribución de Pareto. Es decir,

$$F_\theta(x) = \mathbb{P}_\theta(X_1 \leq x) = 1 - (\theta/x)^c, \quad x \geq \theta,$$

donde $\theta > 0$ es un parámetro desconocido y $c > 0$ es una constante conocida. Sea $Q(\mathbf{X}, \theta) = Z_n/\theta$, con $Z_n = \min(X_1, X_2, \dots, X_n)$.

a) Muestre que la distribución de Z_n es $F_{n,\theta}(z) = 1 - (\theta/z)^{nc}$, $z \geq \theta$.

Sol:

$$1 - F_{n,\theta}(z) = \mathbb{P}_\theta(Z_n > z) = \mathbb{P}_\theta(X_1 > z)^n = \left(\frac{\theta}{z}\right)^{nc}.$$

b) Verifique que Q sirve como función pivote para θ .

Sol: Notamos primero que $Q = \min(X_1, \dots, X_n)/\theta = \min(X_1/\theta, \dots, X_n/\theta)$ y que

$$\mathbb{P}_\theta(X_1/\theta > y) = \mathbb{P}_\theta(X_1 > \theta y) = (\theta/(\theta y))^c = y^{-c}, \quad y \geq 1.$$

Por lo tanto, la distribución de Q no depende de θ . Además, Q es función decreciente de θ , por lo tanto califica como función pivote.

- c) Construya explícitamente un intervalo de confianza de nivel $100(1 - \alpha)\%$ para θ , basado en Q . Calcule t_{α_1} y t_{α_2} de manera que las probabilidades a la izquierda y la derecha sean $\alpha/2$.

Sol: Buscamos $t_{\alpha_1}, t_{\alpha_2}$ tales que

$$\mathbb{P}_\theta(t_{\alpha_1} \leq \frac{Z_n}{\theta} \leq t_{\alpha_2}) = t_{\alpha_1}^{-nc} - t_{\alpha_2}^{-nc} = 1 - \alpha, \quad t_{\alpha_1}, t_{\alpha_2} \geq 1.$$

Para que a la derecha tengamos $\alpha/2$ hay que resolver

$$\mathbb{P}_\theta(\frac{Z_n}{\theta} > t_{\alpha_2}) = t_{\alpha_2}^{-nc} = \alpha/2 \Rightarrow t_{\alpha_2} = (\alpha/2)^{-\frac{1}{nc}}.$$

En el caso de t_{α_1} se plantea la ecuación

$$\mathbb{P}_\theta(\frac{Z_n}{\theta} > t_{\alpha_1}) = t_{\alpha_1}^{-nc} = 1 - \alpha/2 \Rightarrow t_{\alpha_1} = (1 - \alpha/2)^{-\frac{1}{nc}}.$$

Finalmente, el intervalo de confianza es

$$Z_n t_{\alpha_2}^{-1} \leq \theta \leq Z_n t_{\alpha_1}^{-1} \iff Z_n (\alpha/2)^{\frac{1}{nc}} \leq \theta \leq Z_n (1 - \alpha/2)^{\frac{1}{nc}}, \quad \alpha \in (0, 1).$$

3. De un lote muy grande de piezas se extrae aleatoriamente una muestra de n elementos. De esa muestra, k piezas resultan defectuosas. Llamando p a la probabilidad de que una pieza del lote sea defectuosa,

- a) Encuentre el intervalo de confianza de nivel $1 - \alpha$ para p , basado en la aproximación normal.

Sol: De acuerdo con la indicación, consideramos $X_1, \dots, X_n$ como MAS de una Bernoulli de parámetro $p \in [0, 1]$. Sea $\hat{p} = \frac{1}{n} \sum_{i=1}^n X_i$ la frecuencia relativa de artículos defectuosos en la muestra. Las X_i tienen todas esperanza $\mathbb{E}_p(X_i) = p$ y varianza $\text{Var}_p(X_i) = p(1-p)$. Entonces, por la aproximación normal (TCL) resulta que $\hat{p}$ es aproximadamente $N(p, p(1-p)/n)$. Por lo tanto

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

tiene distribución aproximada $N(0, 1)$. También, si reemplazamos p en el denominador por su estimador $\hat{p}$, resulta que

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

es aproximadamente $N(0, 1)$. Este es el teorema de Slutsky.

A continuación usamos esta última expresión como pivote aproximado. Calculamos $t_{\alpha/2}$ tal que

$$\mathbb{P}_p \left[-t_{\alpha/2} \leq \frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})/n}} \leq t_{\alpha/2} \right] = 1 - \alpha.$$

Dado que la distribución del pivote es aproximadamente normal, podemos escribir la expresión anterior en términos de la función de distribución Φ de la normal $N(0, 1)$. Es decir, $\Phi(t_{\alpha/2}) - \Phi(-t_{\alpha/2}) = 1 - \alpha$. Por otra parte, sabiendo que $\Phi(-x) = 1 - \Phi(x)$, la ecuación anterior se transforma en $2\Phi(t_{\alpha/2}) - 1 = 1 - \alpha$, o bien $\Phi(t_{\alpha/2}) = 1 - \alpha/2$. Habiendo calculado $t_{\alpha/2}$ podemos escribir el intervalo de confianza para p como

$$\left[\hat{p} - t_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + t_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right].$$

- b) Calcule explícitamente los intervalos de confianza para los valores siguientes: $n = 225$, $k = 15$ y niveles de confianza 0.9, 0.95 y 0.99.

Sol: Calculamos el intervalo de nivel 90 %. Notemos que $1 - \alpha = 0,9$ implica $\alpha/2 = 0,05$ y $1 - \alpha/2 = 0,95$. De la indicación tenemos que $\Phi(1,65) \approx 0,95$, de manera que $t_{\alpha/2} \approx 1,65$. Por otra parte, usando los valores de k y n tenemos que $\hat{p} = 15/225 = 1/15 \approx 0,067$, $\sqrt{\hat{p}(1 - \hat{p})/n} = \sqrt{14/225} \approx 0,0166$. Multiplicando por $t_{\alpha/2}$ obtenemos el intervalo de nivel 90 % dado por

$$[1/15 - 1,65 \times 0,0166, 1/15 + 1,65 \times 0,0166] = [0,0392, 0,0941].$$

- c) Use el intervalo de (b) para indicar si es razonable pensar que el verdadero valor del parámetro p es 0.1.

Sol: En vista del resultado anterior, dado que el intervalo no contiene a 0.1, diríamos que no es razonable pensar que el verdadero p sea 0.1.

Indicación: Note que $\Phi(1,29) \approx 0,9$, $\Phi(1,65) \approx 0,95$, $\Phi(2,33) \approx 0,99$. Considere que los elementos de la muestra se comportan como va Bernoulli independientes.