Clase Auxiliar #10 Probabilidad y Estadística Prim '08 FC, VR

Intervalos de confianza y Funciones Pivote

Definición a) Sea $(X_1 \ldots X_n)$ MAS de $X \sim f(x|\theta)$. Se dice que el intervalo $I = [a, b] \subset \mathbb{R}$ es intervalo de confianza para "θ" de nivel de confianza "$1-\alpha$" si

$$\mathbb{P}_{\vec{X}}(\theta \in I) = \mathbb{P}_{\vec{X}}(a \leq \theta \leq b) = 1-\alpha.$$

Nota: este intervalo es aleatorio.

b) Una función $Q(\vec{X}, \theta)$ tal que su DISTRIBUCIÓN no depende de θ, se llama función pivote para θ.

Nota: Este tipo de funciones permite encontrar intervalos de confianza. (se recomienda mirar "estadísticos suficientes").

Problema 1 Sea $(X_1 \ldots X_n)$ MAS de una población $Exp(\lambda)$.

a) Probar que $Q(\vec{X}, \lambda) = \lambda \sum_{i=1}^{n} X_i$ es función pivote para λ.

sol Recordamos que si $X_1 \ldots X_n \sim Exp(\lambda)$, iid, entonces $\sum_{i=1}^{n} X_i \sim Gamma(n, \lambda)$.

Calculamos la fgm. de $Q(\vec{X}, \lambda)$:

$$\phi_{Q(\vec{X},\lambda)}(s) = \mathbb{E}\left(e^{sQ(\vec{X},\lambda)}\right) = \mathbb{E}\left(e^{s\lambda\sum_{i=1}^{n} X_i}\right)$$

$$= \phi_Y(s\lambda) \qquad \text{con } Y \sim Gamma(n, \lambda)$$

$$\left(\phi_Y(t) = \frac{1}{\left(1-\frac{t}{\lambda}\right)^n}, \quad t < \lambda\right)$$

$$= \frac{1}{\left(1-\frac{s\lambda}{\lambda}\right)^n}, \quad s\lambda < \lambda$$

$$= \frac{1}{(1-s)^n}, \quad s < 1$$

Notamos que $\phi_{Q(\vec{x},\lambda)}$ no depende de λ, por lo que $Q(\vec{x},\lambda)$ es función pivote para λ. Además, notamos que $Q(\vec{x},\lambda) \sim \text{Gamma}(n,1)$.

b) Encuentre un intervalo de confianza de nivel $1-\alpha$ para λ.

Sol: se busca un intervalo $I=[a,b]$ tal que

$$\mathbb{P}(a \leq \lambda \leq b) = 1-\alpha.$$

Notamos que $\mathbb{P}(a \leq \lambda \leq b) = \mathbb{P}\left(a\sum_{i=1}^{n} X_i \leq \lambda\sum_{i=1}^{n} X_i \leq b\sum_{i=1}^{n} X_i\right)$

$$= \mathbb{P}\left(a\sum_{i=1}^{n} X_i \leq Q(\vec{x},\lambda) \leq b\sum_{i=1}^{n} X_i\right)$$

Llamando $a^* = a\sum X_i$, $b^* = b\sum X_i$,

$$= \mathbb{P}(a^* \leq Q(\vec{x},\lambda) \leq b^*)$$

$$= F_{Q(\vec{x},\lambda)}(b^*) - F_{Q(\vec{x},\lambda)}(a^*) = 1-\alpha$$

con $F_{Q(\vec{x},\lambda)}$ función de distribución de $Q(\vec{x},\lambda)$.

Si tomamos $a^*=0$ (entonces $\boxed{a=0}$), buscamos $b^*>0$ tal que

$$F_{Q(\vec{x},\lambda)}(b^*) = 1-\alpha$$

$$\Rightarrow b^* = F^{-1}_{Q(\vec{x},\lambda)}(1-\alpha) \quad \text{(la función inversa)}$$

$$\Rightarrow b^* = b\,\overline{X}_n = F^{-1}_{Q(\vec{x},\lambda)}(1-\alpha)$$

$$\Rightarrow \boxed{b = \frac{F^{-1}_{Q(\vec{x},\lambda)}(1-\alpha)}{\sum X_i}}$$

Así, $I=[a,b]=[0,b]$ es tal que $\mathbb{P}\left(0 \leq \lambda \leq \dfrac{F^{-1}_{Q(\vec{x},\lambda)}(1-\alpha)}{\sum X_i}\right) = 1-\alpha$. //

Problema 2 Considere $(X_1, \dots, X_n)$ MAS de una población $Unif(0,\theta)$.

a) Pruebe que $T = \frac{\max\{X_1, \dots, X_n\}}{\theta}$ es función pivote para θ.

Sol Calculamos la función de distribución de T.

$$\mathbb{P}(T \leq t) = \mathbb{P}\left(\frac{\max\{X_1 \dots X_n\}}{\theta} \leq t\right)$$

$$= \mathbb{P}(\max\{X_1 \dots X_n\} \leq \theta t)$$

$$= \mathbb{P}(X_1 \leq \theta t)^n \quad ; \rightsquigarrow \; 0 \leq \theta t \leq \theta$$

$$= \left(\frac{\theta t}{\theta}\right)^n \quad ; \; 0 \leq t \leq 1$$

$$= t^n \quad ; \; t \in [0,1]$$

Entonces T es función pivote para θ.

b) Encontrar el intervalo de confianza de nivel $1-\alpha$ para θ, más corto.

Sol Buscamos $I = [a,b]$ tal que

$$\mathbb{P}(a \leq \theta \leq b) = 1-\alpha$$

$$\mathbb{P}(a \leq \theta \leq b) = \mathbb{P}\left(\frac{1}{b} \leq \frac{1}{\theta} \leq \frac{1}{a}\right)$$

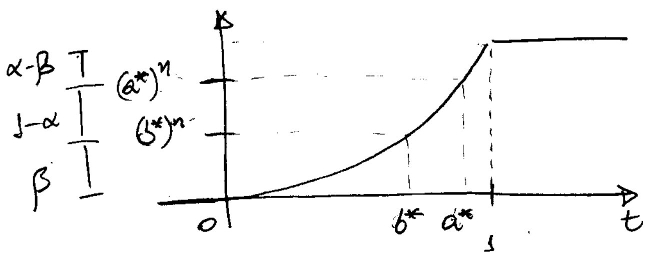
$$= \mathbb{P}\left(\frac{X_{(n)}}{b} \leq T \leq \frac{X_{(n)}}{a}\right) \quad \text{con } X_{(n)} = \max\{X_1, \dots, X_n\}.$$

Llamando $a^* = \frac{X_{(n)}}{a}$, $b^* = \frac{X_{(n)}}{b}$, buscamos a^*, b^* tales que

$$\mathbb{P}(b^* \leq T \leq a^*) = 1-\alpha = F_T(a^*) - F_T(b^*)$$

$$\Rightarrow \quad 1-\alpha = (a^*)^n - (b^*)^n$$

¿Cómo encontrar el intervalo de largo mínimo?



Primero, supongamos que a la derecha de a^* haya probabilidad $\alpha-\beta$, (y a la izquierda de b^* haya entonces probabilidad β)

$$\Rightarrow \quad F_T(b^*) - F_T(0) = (b^*)^n = \beta$$

$$\Rightarrow \boxed{b^* = \sqrt[n]{\beta}} \Rightarrow \boxed{b = \frac{X_{(n)}}{\sqrt[n]{\beta}}}$$

$$\text{y} \quad F_T(1) - F_T(a^*) = 1-(a^*)^n = \alpha-\beta$$

$$\Rightarrow \boxed{a^* = \sqrt[n]{1-\alpha+\beta}} \Rightarrow \boxed{a = \frac{X_{(n)}}{\sqrt[n]{1-\alpha+\beta}}}$$

Para encontrar el intervalo $[a,b]$ de largo mínimo se debe resolver un problema de optimización. El problema es:

$$(P)\begin{cases} \min \quad b-a = X_{(n)} \left(\dfrac{1}{\sqrt[n]{\beta}} - \dfrac{1}{\sqrt[n]{1-\alpha+\beta}} \right) \\ \text{s.a.} \quad \beta \in [0,\alpha] \end{cases}$$

El largo mínimo se obtiene cuando $\beta = \alpha$, entonces $b^* = \sqrt[n]{\alpha}$ y $a^* = 1$

$$\Rightarrow \boxed{a = 1} \quad \text{y} \quad \boxed{b = \frac{1}{\sqrt[n]{\alpha}}}$$

y el intervalo queda $I = \left[X_{(n)} ; \dfrac{X_{(n)}}{\sqrt[n]{\alpha}} \right]$

Problema 3 Considere dos poblaciones normales: $X \sim N(\mu_x, \sigma^2)$ e $Y \sim N(\mu_y, \sigma^2)$; y dos m.a.s. $(X_1 \ldots X_n)$ de X e $(Y_1, \ldots, Y_m)$ de Y (indep.), con σ^2 conocido.

a) Encuentre un intervalo de confianza de nivel $(1-\alpha)$ para $(\mu_x - \mu_y)$.

Sol Sabemos que $\overline{X_n} \sim N(\mu_x, \sigma^2/n)$ e $\overline{Y_m} \sim N(\mu_y, \sigma^2/m)$.

Entonces $\overline{X_n} - \overline{Y_m} \sim N\left(\mu_x - \mu_y, \sigma^2\left(\frac{1}{n} + \frac{1}{m}\right)\right)$

$$\Rightarrow \quad Z = \frac{(\overline{X_n} - \overline{Y_m}) - (\mu_x - \mu_y)}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim N(0,1)$$

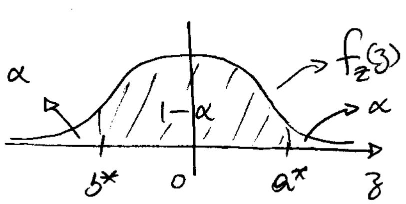
De lo anterior se concluye que Z es función pivote para $(\mu_x + \mu_y)$.

Entonces, buscamos $I = [a,b]$ tal que $\mathbb{P}(a \leq \mu_x + \mu_y \leq b) = 1-\alpha$

$$1-\alpha = \mathbb{P}(a \leq \mu_x - \mu_y \leq b) = \mathbb{P}\left(\underbrace{\frac{\overline{X_n} - \overline{Y_m} - b}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}}_{b^*} \leq Z \leq \underbrace{\frac{\overline{X_n} - \overline{Y_m} - a}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}}_{a^*}\right)$$

$$1-\alpha = \mathbb{P}(b^* \leq Z \leq a^*)$$

El intervalo de largo mínimo será un intervalo simétrico en torno a 0, por la forma de la densidad de $Z \sim N(0,1)$.



El área bajo la curva entre b^* y a^* es $1-\alpha$.
El intervalo más corto con tal área es el simétrico en torno a 0, o sea, $b^* = -a^*$.

Llamamos $z_\beta \in \mathbb{R}$ al número tal que $\mathbb{P}(Z \leq z_\beta) = \beta$.

Buscamos b^*, a^* tales que

$$\mathbb{P}(Z \le b^*) = \frac{\alpha}{2} \quad \Rightarrow \quad b^* = z_{\alpha/2}$$

$$\mathbb{P}(Z \le a^*) = 1 - \alpha/2 \quad \Rightarrow \quad a^* = \boxed{z_{1-\frac{\alpha}{2}} = -z_{\frac{\alpha}{2}}}$$

Los valores z_β se buscan en las tablas estadísticas.

En el caso $\alpha = 0{,}05$, $\quad z_{\alpha/2} = -1{,}96$; $\quad z_{1-\frac{\alpha}{2}} = 1{,}96$.

Llamamos $W = \overline{X}_n - \overline{Y}_m$, $\mu = \mu_X - \mu_Y$, $\tilde{\sigma}^2 = \sigma^2\left(\frac{1}{n} + \frac{1}{m}\right)$

$(W \sim N(\mu, \tilde{\sigma}^2))$.

Entonces, $a^* = \dfrac{W - a}{\tilde{\sigma}} \quad \Rightarrow \quad a = W - a^* \tilde{\sigma}$

$$\boxed{a = W - z_{1-\frac{\alpha}{2}} \tilde{\sigma}}$$

y $b^* = \dfrac{W - b}{\tilde{\sigma}} \quad \Rightarrow \quad b = W - b^* \tilde{\sigma} = W - z_{\alpha/2} \cdot \tilde{\sigma}$

$$\boxed{b = W + z_{1-\frac{\alpha}{2}} \tilde{\sigma}}$$

El intervalo de confianza de nivel $1-\alpha$ para $\mu = \mu_X - \mu_Y$ es

$$I = \left[(\overline{X}_n - \overline{Y}_m) - z_{1-\frac{\alpha}{2}}\, \sigma \sqrt{\frac{1}{n} + \frac{1}{m}} \; ; \; (\overline{X}_n - \overline{Y}_m) + z_{1-\frac{\alpha}{2}}\, \sigma \sqrt{\frac{1}{n} + \frac{1}{m}} \right]$$

//

5) Los cursos X (de 50 alumnos) e Y (de 38 alumnos) rinden cierta prueba. Las notas obtenidas siguen distribuciones Normales de medias μ_X y μ_Y respect, con idéntica distribución estandar (igual a 1,5). Los datos que se recopilan son $\overline{X}_n = 3,8$, $\overline{Y}_m = 4,1$. El coordinador de los cursos cree que las medias (teóricas) de los cursos son iguales. Con un nivel de confianza del 95%, puede ser posible lo que piensa el coordinador

Sol: Identificamos los datos: $n = 50$, $m = 38$; $\overline{X}_n = 3,8$; $\overline{Y}_m = 4,1$; $\sigma = 1,5$; $\alpha = 0,05$. El intervalo de confianza para $\mu_X - \mu_Y$ queda

$$I = [-1,1327 ; 0,1327]$$

Como este intervalo contiene al 0, y con un 95% de confianza la diferencia de medias puede estar en este intervalo, con un 95% de confianza puede ser posible lo que piensa el coordinador.