FEBRERO-2017.pdf

AzaharaFS

Inferencia Estadística

3º Grado en Matemáticas

Facultad de Ciencias
Universidad de Granada

FEBRERO 2017

INFERENCIA ESTADÍSTICA.

1.- a) Teoría

b) Sean $(X_1,\dots,X_7)$, $(Y_1,\dots,Y_{11})$ muestras independientes de poblaciones $N(3,\sigma^2)$ y $N(5,\sigma^2)$, respectivamente con medias muestrales $\bar{X}$ e $\bar{Y}$.

Calcular K tal que $P(V<K)=0'89$ siendo

$$V=\frac{\bar{X}-\bar{Y}+2}{\sqrt{\sum_{i=1}^{7}(X_i-\bar{X})+\sum_{i=1}^{11}(Y_i-\bar{Y})^2}} \qquad n_1=7 \quad n_2=11$$

Sabemos que $\bar{X}\to N(3,\frac{\sigma^2}{n_1})$, $\bar{Y}\to N(5,\frac{\sigma^2}{n_2})$. Por el Lema de Fisher tenemos que los estadísticos $\bar{X}$ e $\bar{Y}$ son independientes, además $\bar{X}-\bar{Y}\rightsquigarrow N((3-5),\sigma^2(\frac{1}{7}+\frac{1}{11}))=N(-2,\sigma^2(\frac{18}{77}))$

Tipificando obtenemos que $\dfrac{\bar{X}-\bar{Y}-(-2)}{\sigma\sqrt{18/77}}\rightsquigarrow N(0,1)$

Por otro lado, sabemos que $\dfrac{(n_1-1)S_1^2}{\sigma^2}\rightsquigarrow \chi^2(n_1-1)$ y $\dfrac{(n_2-1)S_2^2}{\sigma^2}\rightsquigarrow\chi^2(n_2-1)$. Como por el Lema de Fisher S_1^2 y S_2^2 son independientes, podemos usar la reproductividad de la χ^2.

$X_1,\dots,X_7$ indep. y $X_i\to\chi^2(n_i)\Rightarrow\sum_{i=1}^{n}X_i\to\chi^2(\sum_{i=1}^{n}n_i)$

Luego $\dfrac{\sum_{i=1}^{7}(X_i-\bar{X})^2}{\sigma^2}+\dfrac{\sum_{i=1}^{11}(Y_i-\bar{Y})^2}{\sigma^2}=\dfrac{1}{\sigma^2}\left(\sum_{i=1}^{7}(X_i-\bar{X})^2+\sum_{i=1}^{11}(Y_i-\bar{Y})^2\right)\rightsquigarrow\chi^2(16)$
$\qquad\qquad \chi^2(n_1+n_2-2)$

Sabemos que si A y B son independientes, $A\to N(0,1)$, $B\rightsquigarrow\chi^2(n)\Rightarrow T=\dfrac{A}{\sqrt{B/n}}\to t(n)$.

Así,
$$\frac{\dfrac{\bar{X}-\bar{Y}+2}{\sigma\sqrt{18/77}}}{\sqrt{\dfrac{\dfrac{\sum_{i=1}^{7}(X_i-\bar{X})+\sum_{i=1}^{11}(Y_i-\bar{Y})}{\sigma^2}}{16}}} = \frac{\bar{X}-\bar{Y}+2}{\sqrt{\sum_{i=1}^{7}(X_i-\bar{X})^2+\sum_{i=1}^{11}(Y_i-\bar{Y})^2}}\cdot\frac{4}{\sqrt{\frac{18}{77}}} \rightarrow t(16)$$

Por tanto, $P(V<K) = P\left(\frac{4V}{\sqrt{18/77}} < \frac{4K}{\sqrt{18/77}}\right) = 1-P\left(\frac{4V}{\sqrt{18/77}} \geq \frac{4K}{\sqrt{18/77}}\right) = 0'99$

$P\left(\frac{4V}{\sqrt{18/77}} \geq \frac{4K}{\sqrt{18/77}}\right) = 0'01 \Rightarrow \frac{4K}{\sqrt{18/77}} = 2'5835 \Rightarrow K = 0'312276$

2.a) Teoría

b) Sea $X_1,\dots,X_n$ una m.a.s. de una v.a. simple X con función de densidad
$$f_\theta(x) = \frac{2x}{\theta^2-1} \qquad 1<x<\theta$$

i) Encontrar un estadístico suficiente y completo, probando que lo es.

ii) Encontrar, si existe, el UMVUE para θ.

iii) Especificar la función de verosimilitud y encontrar el estimador máximo verosímil de θ justificando su obtención. ¿Es insesgado ese estimador? (justificar respuesta).

i) Ya que $\chi_\theta = (1,\theta)$ depende de θ, la familia no es exponencial uniparamétrica, luego para buscar un estadístico suficiente vamos a utilizar el Teorema de Factorización de Neyman-Fisher. Un estadístico $T(X_1,\dots,X_n)$ es suficiente si y solo, para cualquier valor de θ se tiene
$$f_\theta^n(x_1,\dots,x_n) = h(x_1,\dots,x_n)\,g(T(x_1,\dots,x_n)) \quad \forall (x_1,\dots,x_n)\in\chi^n$$
donde $h(x_1,\dots,x_n)$ es indep. de θ y g depende de la muestra solo a través del estadístico.

Así, $f_\theta^n(x_1,\dots,x_n) = \dfrac{2^n \prod_{i=1}^n x_i}{(\theta^2-1)^n} \; I_{(-\infty,\theta)}(\underbrace{X_{(n)}}_{\max X_i}) \cdot I_{(1,+\infty)}(\underbrace{X_{(1)}}_{\min X_i})$

Como $\mathcal{X} = \bigcup_\theta \mathcal{X}_\theta = \bigcup_\theta (1,\theta) = (1,+\infty)$, $I_{(1,+\infty)}(X_{(1)}) = 1 \quad \forall x \in \mathcal{X}$,

por tanto

$$f_\theta^n(x_1,\dots,x_n) = \frac{2^n \prod_{i=1}^n x_i}{(\theta^2-1)^n} I_{(-\infty,\theta)}(X_{(n)})$$ y tomando

$h(x_1,\dots,x_n) = 2^n \prod_{i=1}^n x_i$ y $g(T(x_1,\dots,x_n)) = \dfrac{1}{(\theta^2-1)^n} \cdot I_{(-\infty,\theta)}(X_{(n)})$

concluimos que $X_{(n)}$ es un estadístico suficiente.

Para ver que es completo, tengo que ver que si tengo cualquier función medible g tal que

$E_\theta(g(T)) = 0 \quad \forall \theta \in \Theta = (1,+\infty)$, entonces $P_\theta(g(T)=0) = 1 \quad \forall \theta \in \Theta$

Así, vamos a calcular primero la función de densidad del estadístico:

$F_\theta(x) = \displaystyle\int_1^x \frac{2t}{\theta^2-1}\,dt = \frac{x^2-1}{\theta^2-1}$ con $1<x<\theta \Rightarrow F_{X_{(n)}}(x) = (F(x))^n = \left(\dfrac{x^2-1}{\theta^2-1}\right)^n$

Así, $f_{X_{(n)}}(x) = n\left(\dfrac{x^2-1}{\theta^2-1}\right)^{n-1} \cdot \dfrac{2x}{(\theta^2-1)} = \dfrac{2xn(x^2-1)^{n-1}}{(\theta^2-1)^n}$

$$E_\theta(g(T)) = \int_1^\theta g(t) f_\theta(t)\,dt = \int_1^\theta g(t)\frac{2nt(t^2-1)^{n-1}}{(\theta^2-1)^n}\,dt =$$

$$= \frac{2n}{(\theta^2-1)^n}\int_1^\theta g(t)t(t^2-1)^{n-1}\,dt = 0 \Leftrightarrow \int_1^\theta g(t)t(t^2-1)^{n-1}\,dt = 0$$

Por el TFC sabemos que $g(t)t(t^2-1)^{n-1}$ tiene primitiva G cumpliendo que $G(\theta)-G(1) = 0$. Derivando respecto de θ, tenemos que

$G'(\theta)-G'(1)=0 \Rightarrow g(\theta)\theta(\theta^2-1)^{n-1}=0 \Leftrightarrow g(\theta)=0 \quad \forall\theta\in\Theta$

Ya que $\theta(\theta^2-1)^{n-1}\neq 0$ por ser $\theta>1$.

Hemos demostrado que $\{t<\theta\}\subseteq\{t/g(t)=0\}$, y tomando probabilidades tenemos que

$$1\geq P(g(T)=0)\geq P(T<\theta)=1$$

$$\Rightarrow \boxed{P(g(T)=0)=1}$$

Por tanto $T=X_{(n)}$ es completo.

ii) UMVUE?

Sabemos que el UMVUE se puede calcular mediante una función de un estadístico suficiente y completo, luego sea $h(T)$ dicha función, imponemos la insesgadez.

$$E_\theta(h(T))=\int_1^\theta h(t)f_{X_{(n)}}(t)\,dt = \underbrace{\frac{2n}{(\theta^2-1)^n}}_{\text{apartado anterior}}\int_1^\theta h(t)t(t^2-1)^{n-1}dt=\theta \Leftrightarrow$$

$$\Leftrightarrow \int_1^\theta h(t)t(t^2-1)^{n-1}dt=\frac{\theta(\theta^2-1)^n}{2n}$$ Aplicando de nuevo el TFC

Existe H primitiva de $h(t)t(t^2-1)^{n-1}$ tal que

$$H(\theta)-H(1)=\frac{\theta(\theta^2-1)^n}{2n}$$ Derivamos respecto de θ

$$\Rightarrow h(\theta)\theta(\theta^2-1)^{n-1}=\frac{(\theta^2-1)^n+2\theta^2n(\theta^2-1)^{n-1}}{2n}$$

Luego $h(T)=\frac{T^2-1}{2nT}+T=\frac{T}{2n}-\frac{1}{2nT}+T$ es candidato a UMVUE.

- Veamos que es estimador de θ, es decir, que toma valores en $\Theta=(1,+\infty)$. Como $\frac{T^2-1}{2nT}>0$ por ser $T>1$

$\Rightarrow h(T)=\frac{T^2-1}{2nT}+T\in(1,+\infty)$, luego sí es estimador.

• Veamos por último si tiene momento de segundo orden finito.

$E_\theta(h(T)^2) = \int_1^\theta h(t)^2 f_{X_{(n)}}(t)\,dt < +\infty$ ya que tenemos la integral en un recinto finito de un polinomio.

$\Rightarrow h(T) = \frac{T}{2n} - \frac{1}{2nT} + T$ es el UMVUE.

iii) Función de verosimilitud:

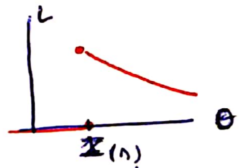
$$L_{x_1,\dots,x_n}(\theta) = \begin{cases} 0 & \text{si } X_{(1)} \le 1 \text{ o } X_{(n)} \ge \theta \\ \dfrac{2^n \prod_{i=1}^n x_i}{(\theta^2-1)^n} & \text{si } 1 < X_{(1)} \le X_{(n)} < \theta \end{cases}$$

(Siempre se va a cumplir que $X_{(1)} > 1$)

Para encontrar el EMV, vamos a buscar el máximo de la función de verosimilitud. Como en este caso la función no es derivable, ya que los intervalos dependen del parámetro, dibujamos.

Definimos $g_{x_1,\dots,x_n}(\theta) = \frac{2^n \prod_{i=1}^n x_i}{(\theta^2-1)^n}$ y estudiamos su monotonía. $g_{x_1,\dots,x_n}(\theta) = \underbrace{\frac{1}{(\theta^2-1)^n}}_{\text{Decreciente en } (1,+\infty)} \cdot \underbrace{2^n \prod_{i=1}^n x_i}_{\text{No depende de } \theta}$ es decreciente en θ

• Si $\theta < X_{(n)}$, $L = 0$



Por tanto, la función $L_{x_1,\dots,x_n}(\theta)$ es decreciente y alcanza su máximo cuanto menor es θ, es decir, cuando $\theta = X_{(n)}$.
↑ pero siempre mayor que $X_{(n)}$

Así, $\hat\theta = X_{(n)}$ es EMV de θ.

Como ya demostramos que $h(T) = \frac{T^2-1}{2nT} + T$ es UMVUE para θ, con $T = X_{(n)}$. Como el UMVUE para θ es único, entonces se tiene que $\hat{\theta} = X_{(n)}$ no es UMVUE para θ pues, en general $h(X_{(n)}) \neq X_{(n)}$. Por tanto, tampoco es insesgado, ya que si lo fuera, sería un estimador función del UMVUE para θ, contradiciendo así su unicidad.

3. a) ~~[illegible]~~ Teoría

b) Construir el test de Neyman-Pearson de tamaño α para contrastar $H_0: \theta = \theta_0$ frente a $H_1: \theta = \theta_1$ siendo $\theta_1 < \theta_0$, basándose en una muestra de tamaño n de una variable aleatoria con función de densidad

$$f_\theta(x) = \frac{1}{x \ln \theta} \quad 1 < x < \theta.$$

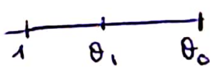
Calcular la potencia del test de tamaño α y representarla en función de α.

b) Sabemos que el test de Neyman-Pearson es de la forma

$$\varphi(X_1, \dots, X_n) = \begin{cases} 1 & \text{si } f_1^n(X_1, \dots, X_n) > k f_0^n(X_1, \dots, X_n) \\ \gamma & \text{si } f_1^n(X_1, \dots, X_n) = k f_0^n(X_1, \dots, X_n) \\ 0 & \text{si } f_1^n(X_1, \dots, X_n) < k f_0^n(X_1, \dots, X_n) \end{cases}$$

con $k \in \mathbb{R}^+ \cup \{0\}$ y $\gamma(X_1, \dots, X_n) = \gamma \in [0,1]$.

Tenemos que $\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta = \theta_1 \end{cases}$



$\mathcal{X}_0 = (1, \theta_0)$ $\mathcal{X}_1 = (1, \theta_1)$, como $\theta_1 < \theta_0 \Rightarrow \mathcal{X}_1 \subset \mathcal{X}_0 \Rightarrow \mathcal{X}^n = \mathcal{X}_0^n = (1, \theta_0)^n$

Así $f_{\theta_j}^n(x_1, \dots, x_n) = \frac{1}{(\ln \theta_j)^n \prod_{i=1}^n x_i} I_{(-\infty, \theta_j)}(X_{(n)}) \cdot I_{(1, +\infty)}(X_{(1)})$ $j = 0, 1$

por tanto $f_0^n(x_1, \dots, x_n) = \frac{1}{(\ln \theta_0)^n \prod_{i=1}^n x_i} \neq 0 \quad \forall (x_1, \dots, x_n) \in \mathcal{X}^n$

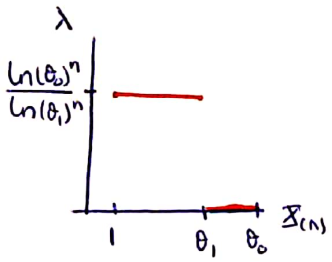
⇒ El test de Neyman-Pearson nos queda (dividiendo entre $f_0^n(x_1,\dots,x_n)$)

$$\varphi(x_1,\dots,x_n) = \begin{cases} 1 & \text{si } \lambda(x_1,\dots,x_n) > K \\ \gamma & \text{si } \lambda(x_1,\dots,x_n) = K \\ 0 & \text{si } \lambda(x_1,\dots,x_n) < K \end{cases}$$

donde $\lambda(x_1,\dots,x_n) = \dfrac{f_1^n(x_1,\dots,x_n)}{f_0^n(x_1,\dots,x_n)}$

$$\lambda(x_1,\dots,x_n) = \begin{cases} \dfrac{1/(\ln(\theta_1))^n \prod_{i=1}^n x_i}{1/(\ln(\theta_0))^n \prod_{i=1}^n x_i} = \dfrac{(\ln(\theta_0))^n}{(\ln(\theta_1))^n} & \text{si } 1 < X_{(n)} < \theta_1 \\ 0 & \text{si } \theta_1 \leq X_{(n)} < \theta_0 \end{cases}$$

Como λ toma solo dos valores, le damos a K esos valores



$\boxed{K=0}$ λ no puede ser menor que K, luego el test nos queda

$$\varphi(x_1,\dots,x_n) = \begin{cases} 1 & \text{si } 1 < X_{(n)} < \theta_1 \\ \gamma & \text{si } \theta_1 \leq X_{(n)} < \theta_0 \end{cases}$$

Calculamos γ imponiendo el tamaño. ($\gamma \in [0,1]$)

• Tamaño: $E_{\theta_0}[\varphi_\alpha(X_1,\dots,X_n)]$

$$\alpha = 1 \cdot P_{\theta_0}(1 < X_{(n)} < \theta_1) + \gamma \cdot \underbrace{P_{\theta_0}(\theta_1 \leq X_{(n)} < \theta_0)}_{1 - P_{\theta_0}(1 < X_{(n)} < \theta_1)}$$

y como $P_{\theta_0}(1 < X_{(n)} < \theta_1) = (P_{\theta_0}(1 < X < \theta_1))^n = \left(\int_1^{\theta_1} \frac{1}{x \ln \theta_0} dx\right)^n = \left(\frac{\ln \theta_1}{\ln \theta_0}\right)^n$

$$\Rightarrow \alpha = 1\cdot\left(\frac{\ln\theta_1}{\ln\theta_0}\right)^n + \gamma\left(1-\left(\frac{\ln\theta_1}{\ln\theta_0}\right)^n\right) \Rightarrow \gamma = \frac{\alpha-\left(\frac{\ln\theta_1}{\ln\theta_0}\right)^n}{\underbrace{1-\left(\frac{\ln\theta_1}{\ln\theta_0}\right)^n}_{\overset{\wedge}{1}\ (\alpha\le 1)}} \in [0,1] \Leftrightarrow$$

$$\Leftrightarrow \alpha - \left(\frac{\ln\theta_1}{\ln\theta_0}\right)^n \geq 0 \Rightarrow \alpha \geq \left(\frac{\ln\theta_1}{\ln\theta_0}\right)^n$$

Así,

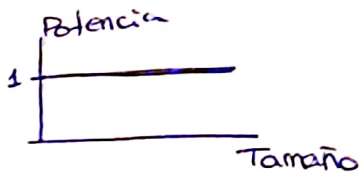
$$\text{TMP de tamaño } \alpha \geq \left(\frac{\ln\theta_1}{\ln\theta_0}\right)^n \longrightarrow \varphi_\alpha(X_1,\dots,X_n) = \begin{cases} 1 & \text{si } 1 < X_{(n)} < \theta_1 \\ \dfrac{\alpha-\left(\frac{\ln\theta_1}{\ln\theta_0}\right)^n}{1-\left(\frac{\ln\theta_1}{\ln\theta_0}\right)^n} & \text{si } \theta_1 \le X_{(n)} \overset{\wedge}{\theta_0} \end{cases}$$

- Potencia:

$$\beta_{\varphi_\alpha}(\theta_1) = E_{\theta_1}[\varphi_\alpha(X_1,\dots,X_n)] = \underbrace{P_{\theta_1}(1 < X_{(n)} < \theta_1)}_{1} + \gamma\cdot\underbrace{P_{\theta_1}(\theta_1 \le X_{(n)} < \theta_0)}_{0} = 1$$

$$\boxed{k = \left(\frac{\ln\theta_0}{\ln\theta_1}\right)^n}$$

Como λ no toma valores mayores que k, el test nos queda:



$$\varphi(X_1,\dots,X_n) = \begin{cases} \gamma & \text{si } 1 < X_{(n)} < \theta_1 \\ 0 & \text{si } \theta_1 \le X_{(n)} < \theta_0 \end{cases}$$

Calculamos γ en función del tamaño $(\gamma \in [0,1])$

- Tamaño: $E_{\theta_0}[\varphi_\alpha(X_1,\dots,X_n)]$

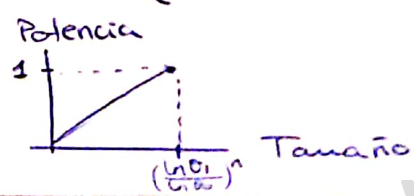
$$\alpha = 1\cdot P_{\theta_0}(1 < X_{(n)} < \theta_1) + 0\cdot P_{\theta_0}(\theta_1 \le X_{(n)} < \theta_0) = \left(\frac{\ln\theta_1}{\ln\theta_0}\right)^n \Rightarrow$$

$$\Rightarrow \gamma = \frac{\alpha}{\underbrace{\left(\frac{\ln\theta_1}{\ln\theta_0}\right)^n}_{\overset{\vee}{0}\ (\alpha\ge 0)}} \in [0,1] \Leftrightarrow \frac{\alpha}{\left(\frac{\ln\theta_1}{\ln\theta_0}\right)^n} \le 1 \Rightarrow \alpha \le \left(\frac{\ln\theta_1}{\ln\theta_0}\right)^n$$

$$\text{TMP de tamaño } \alpha < \left(\frac{\ln\theta_1}{\ln\theta_0}\right)^n \longrightarrow \varphi_\alpha(X_1,\dots,X_n) = \begin{cases} \dfrac{\alpha}{\left(\frac{\ln\theta_1}{\ln\theta_0}\right)^n} & \text{si } 1 < X_{(n)} < \theta_1 \\ 0 & \text{si } \theta_1 \le X_{(n)} < \theta_0 \end{cases}$$

- Potencia:

$$\beta_{\varphi_\alpha}(\theta_1) = E_{\theta_1}[\varphi_\alpha(X_1,\dots,X_n)] = \frac{\alpha}{\left(\frac{\ln\theta_1}{\ln\theta_0}\right)^n}$$



4. Teoría

5. La longitud en mm de los tornillos, medida en dos aparatos diferentes, da los siguientes resultados:

Aparato 1	103'17	103'19	103'17	103'20	103'22	104	104'01	...
Aparato 2	103'18	103'17	103'19	103'18	~~103'96~~ 103'22	~~103'22~~ 103'96	104	

...	104'03	103'5	103'27
	104'05	103'54	103'25

Suponiendo que las distribuciones de la longitud tienen la misma forma funcional, contrastar su igualdad mediante el test de los rangos signados. ¿Para qué niveles de significación se rechaza la hipótesis nula? Especificar las hipótesis necesarias para aplicar este test.

Para aplicar este test necesitaremos suponer que la distribución que siguen las longitudes es simétrica.

$X = A_1 - A_2$: Diferencia de la longitud de los tornillos del aparato 1 y del aparato 2.

$$\begin{cases} H_0: M_X = 0 \\ H_1: M_X \neq 0 \end{cases}$$

Observaciones de X:

$d_i = -0{,}01; 0{,}02; -0{,}02; 0{,}02; \cancel{0}; 0{,}04; 0{,}01; -0{,}02; -0{,}04; 0{,}02$ (eliminamos el 0 por ser $H_0: M_X = 0$)

$n = 9$

Nuestro estadístico de contraste será:

$T^+(X_1, \dots, X_n) =$ Suma de los rangos correspondientes a diferencias positivas.

$\lvert d_i \rvert$	0,01	0,01	0,02	0,02	0,02	0,02	0,02	0,04	0,04
Signo	+	−	+	+	+	−	−	+	−
Rango	1,5	1,5	5,4	5,4	5,4	5,4	5,4	8,5	8,5

$T^+_{exp} = 1{,}5 + 5{,}4 \cdot 3 + 8{,}5 = 26{,}2$

$p\text{-nivel} = P(T^+ \geq 26{,}2) = P(T^+ \geq 27) = 0'326$

Se acepta H_0 a niveles de significación inferiores a 0'326, y se rechaza a niveles superiores.

3.a) Sea $(X_1, \dots, X_n)$ una m.a.s. de $X \rightarrow N(\mu, \sigma^2)$ con ambos parámetros desconocidos. Deducir el test de razón de verosimilitud de tamaño α para contrastar $\begin{cases} H_1 : \mu \geq \mu_0 \\ H_2 : \mu < \mu_0 \end{cases}$

Escaneado con CamScanner