Notas para el curso de Introducción a la Estadística,

dictado por Juan Kalemkerian [1]
en la Facultad de Ciencias, el segundo semestre de 2008.

[1]

Los errores que puedan contener son total responsabilidad de quien las transcribe
Por sugerencias y correcciones: acholaquidis@cmat.edu.uy

Índice general

Capítulo 1

Introducción

Este capítulo pretende introducir los conceptos de esperanza condicional, así como las nociones de convergencia de variables aletorias, que serán necesarios para los siguientes capítulos. Se asumira que el lector está familiarizado con los conceptos básicos de la probabilidad, correspondientes a un primer curso introductorio, no así los del análisis real.

1.1. Esperanza Condicional

Definición 1.1. Dado $\big(\Omega, \mathcal{A}, P\big)$ un espacio de probabilidad, y $X, Y : \Omega \to \mathbb{R}$ variables aleatorias, $E(X) < \infty$, definimos la Esperanza Condicional de X dado Y que anotaremos

$$E\big(X|Y\big),$$

como la función de Y, que denotaremos $E(X|Y) : \Omega \to \mathbb{R}$ que verifica

$$E\big(X\mathbb{I}_Y(B)\big) = E\big(E(X|Y)\mathbb{I}_Y(B)\big)$$

para todo B perteneciente a la sigma álgebra de Borel de $\mathbb{R}$, que anotaremos de aquí en mas como $\mathcal{B}(\mathbb{R})$.

Observación 1.2. *Veamos que $E(X|Y)$ está bien definida:*

existencia: *Que $E(X|Y)$ existe y es una variable aleatoria se sigue del Teorema de Radon-Nikodym.*

unicidad: *Supongamos que $\alpha(Y)$ y $\beta(Y)$ cumplen*

$$\begin{aligned} E\big(X\mathbb{I}_Y(B)\big) &= E\big(\alpha(Y)\mathbb{I}_Y(B)\big) \qquad \forall B \in \mathcal{B}(\mathbb{R}) \\ &= E\big(\beta(Y)\mathbb{I}_Y(B)\big). \end{aligned}$$

Consideremos $B = \{y \in \mathbb{R} : \alpha(y) > \beta(y)\}$, sabemos que $0 = E\big((\alpha(Y) - \beta(Y))\mathbb{I}_Y(B)\big)$, como $\big(\alpha(Y) - \beta(Y)\big)\mathbb{I}_Y(B) \geq 0$ y su esperanza es 0 entonces $\big(\alpha(Y) - \beta(Y)\big)\mathbb{I}_Y(B) = 0 \quad c.s.$, pero por otro lado $\big(\alpha(Y) - \beta(Y)\big)\mathbb{I}_Y(B) > 0$ $\forall\omega$ tal que $Y(\omega) \in B$, por lo tanto $P(Y \in B) = 0$ esto es $P\big(\alpha(Y) > \beta(Y)\big) = 0$. De forma totalmente análoga, tomando $\hat{B} = \{y \in \mathbb{R} : \beta(y) > \alpha(y)\}$, obtenemos que $P(Y \in \hat{B}) = P\big(\beta(Y) > \alpha(Y)\big) = 0, \quad c.s.$, de donde se sigue que $\alpha(Y) = \beta(Y) \quad c.s.$

Proposición 1.3. *Veamos ahora algunas propiedades de las esperanza condicional, X, Y, Z serán variables aleatorias a valores reales y a, b números reales.*

1) Linealidad: $E(aX + bY|Z) = aE(X|Z) + bE(Y|Z)$.

2) Si $X \geq 0 \quad c.s.$ entonces $E(X|Y) \geq 0 \quad c.s.$.

3) Si $X \leq Z$ entonces $E(X|Y) \leq E(Z|Y)$.

4) $E(X|X) = X$.

5) $E(a|Y) = a$.

6) $E(X|Y) = E(X)$ si X e Y son independientes.

7) $E(Xg(Y)|Y) = g(Y)E(X|Y)$.

8) $E\big(E(X|Y)\big) = E(X)$.

Demostración.

1) Por la unicidad, basta demostrar que, para todo $B \in \mathcal{B}(\mathbb{R})$

$$E\big((aX + bY)\mathbb{I}_B(Z)\big) = E\big((aE(X|Z) + bE(Y|Z))\mathbb{I}_B(Z)\big),$$

usando la linealidad de la esperanza el último término es

$$aE\big(E(X|Z)\mathbb{I}_B(Z)\big) + bE\big(E(Y|Z)\mathbb{I}_B(Z)\big),$$

que, por definición de esperanza condicional, es igual a

$$aE\big(X\mathbb{I}_B(Z)\big) + bE\big(Y\mathbb{I}_B(Z)\big).$$

2) La demostración necesita de conceptos del análisis real.

3) Es consecuencia inmediata de 2).

4) Es consecuencia inmediata de la unicidad.

6) Queremos ver que $E\big(X\mathbb{I}_B(Y)\big) = E\big(E(X)\mathbb{I}_B(Y)\big)$, $E\big(X\mathbb{I}_B(Y)\big) = E(X)E(\mathbb{I}_B(Y))$ por la independencia, y $E\big(E(X)\mathbb{I}_B(Y)\big) = E(X)E(\mathbb{I}_B(Y))$ dado que una constante es independiente de cualquier variable.

5) Es una consecuencia inmediata de 6).

7) La demostración necesita de conceptos del análisis real.

8) Basta tomar en la definición $B = \mathbb{R}$.

$\square$

Proposición 1.4. Desigualdad de Jensen: *Sea $\varphi : \mathbb{R} \to \mathbb{R}$ convexa, entonces*

$$\begin{aligned} \varphi\big(E(X)\big) &\leq E\big(\varphi(X)\big) \\ \varphi\big(E(X|Y)\big) &\leq E\big(\varphi(X)|Y\big) \end{aligned}$$

La primera desigualdad es estricta si φ es convexa en sentido estricto y X no es constante. Recordemos que φ es convexa si y solo si $\varphi(\lambda p + (1-\lambda)q) \leq \lambda\varphi(p) + (1-\lambda)\varphi(q)\ \forall p, q,\ \forall \lambda \in [0,1]$, y que si φ es C^2, φ es convexa si y solo si $\varphi''(x) \geq 0\ \forall x$.

Notación: Sea $\alpha(Y) = E\big(X|Y\big)$, denotamos $E\big(X|Y = y\big) = \alpha(y) \in \mathbb{R}$. Dicho de otra manera:

$$E\big(X|Y = y\big) = E\big(X|Y\big)(\omega)$$

donde ω es tal que $Y(\omega) = y$.

Definición 1.5. Distribución Condicional: Dadas X, Y v.a., definimos

$$F_{X|Y=y}(x) := P(X \leq x|Y = y) := E(\mathbb{I}_{(-\infty,x]}(X)|Y = y).$$

Definición 1.6. Probabilidad Condicional: Dadas X, Y v.a., definimos

$$P(X \in [a,b]|Y) := E(\mathbb{I}_{[a,b]}(X)|Y).$$

Ejemplo 1.7. Veamos por separado, primero el caso en que las variables son discretas, y luego el caso continuo.

Caso Discreto: Sea (X, Y) vector aleatorio bidimensional tal que $Rec(X, Y) = \{(x_n, y_m) : n, m \in \mathbb{N}\}$, definimos la probabilidad condicional en el sentido usual, como

$$P_{X|Y=y}(x) = P(X = x|Y = y) = \frac{P_{X,Y}(x,y)}{P_Y(y)} \quad \forall x \in Rec(X), \forall y \in Rec(Y),$$

entonces

$$E(X|Y) = \sum_{x \in Rec(X)} x P_{X|Y}(x),$$

donde $P_{X|Y}(x)$ es la variable aleatoria definida en $\omega \in \Omega$ como $P_{X|Y}(x)(\omega) = P_{X|Y=Y(\omega)}(x)$.

Demostración. Sea

$$\alpha(y) = \sum_{x \in Rec(X)} x P_{X|Y=y}(x) = \frac{1}{P_Y(y)} \sum_{x \in Rec(X)} x P_{X,Y}(x,y)$$

queremos demostrar que

$$E\big(\alpha(Y)\mathbb{I}_B(Y)\big) = E\big(X\mathbb{I}_B(Y)\big), \forall B \in \mathcal{B}(\mathbb{R})$$

En efecto

$$\begin{aligned} E\big(\alpha(Y)\mathbb{I}_B(Y)\big) &= \sum_{y \in Rec(Y)} \alpha(y)\mathbb{I}_B(y)P_Y(y) \\ &= \sum_{y \in Rec(Y)} \sum_{x \in Rec(X)} x\mathbb{I}_B(y)P_{X,Y}(x,y) \\ &= E\big(X\mathbb{I}_B(Y)\big). \end{aligned}$$

□

Caso Continuo: Sea (X, Y) absolutamente continuo, entonces

$$E(X|Y) = \int_{-\infty}^{+\infty} x \frac{f_{X,Y}(x,Y)}{f_Y(Y)} dx,$$

donde $f_{X,Y}(x, Y)$ es la variable aleatoria definida en ω como $f_{X,Y}(x, Y)(\omega) = f_{X,Y}(x, Y(\omega))$.

Demostración.

$$\begin{aligned} E\left(\int_{-\infty}^{+\infty} x \frac{f_{X,Y}(x,Y)}{f_Y(Y)} dx \mathbb{I}_B(Y)\right) &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} x \frac{f_{X,Y}(x,y)}{f_Y(y)} \mathbb{I}_B(y) dx\right) f_Y(y) dy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x\mathbb{I}_B(y) f_{X,Y}(x,y) dx dy \\ &= E(X\mathbb{I}_B(Y)). \end{aligned}$$

Luego, la tesis es consecuencia de la unicidad. □

Proposición 1.8. Fórmula de la distribución conjunta: Dadas X, Y v.a. se tiene que

$$F_{X,Y}(x,y) = \int_{-\infty}^{y} F_{X|Y=t}(x) dF_Y(t)$$

Demostración.

$$\begin{aligned} F_{X,Y}(x,y) &= P(X \leq x, Y \leq y) = E\big(\mathbb{I}_{(-\infty,x]}(X)\mathbb{I}_{(-\infty,y]}(Y)\big) \\ &= E\Big(E\big(\mathbb{I}_{(-\infty,x]}(X)\mathbb{I}_{(-\infty,y]}(Y)\big)\big|Y\Big) \\ &= \int_{-\infty}^{+\infty} \mathbb{I}_{(-\infty,y]}(t) F_{X|Y=t}(x) dF_Y(t) \\ &= \int_{-\infty}^{y} F_{X|Y=t}(x) dF_Y(t) \end{aligned}$$

□

Definición 1.9. Distribución condicionada a un conjunto: Dada X v.a. y $A \in \mathcal{A}$ con $P(A) \neq 0$ definimos

$$F_{X|A} = P(X \leq x|A) = \frac{P(X \leq x \cap A)}{P(A)}$$

Definición 1.10. Esperanza condicionada a un conjunto:

$$E(X|A) = \int_{-\infty}^{+\infty} x dF_{X|A}(x) \quad A \in \mathcal{A}, P(A) \neq 0$$

1.2. Nociones de convergencia de variables aletorias

Definición 1.11. Convergencia en probabilidad y casi segura: Dado $\big(\Omega, \mathcal{A}, P\big)$ espacio de probabilidad, $\{X_n\}$ una sucesión de v.a. y X una v.a. decimos que

1) X_n converge a X en probabilidad, y anotamos $X_n \xrightarrow{P} X$ si

$$\lim_{n \to +\infty} P\big(|X_n - X| \leq \varepsilon\big) = 1$$

2) X_n converge a X casi seguramente, y anotamos $X_n \xrightarrow{c.s.} X$ si

$$P\left(\lim_{n \to +\infty} X_n = X\right) = 1$$

Definición 1.12. Convergencia en distribución: Sean X_n v.a. en $\big(\Omega_n, \mathcal{A}_n, P_n\big)$ y X v.a. en $\big(\Omega, \mathcal{A}, P\big)$, decimos que X_n converge en distribución a X y anotamos

$$X_n \xrightarrow{d} X \quad si \quad \lim_{n \to +\infty} F_{X_n}(x) = F_X(x) \quad \forall x \text{ punto de continuidad de } F_X$$

Proposición 1.13. Relación entre convergencias: Si $\{X_n\}$ y X son v.a. sobre $\big(\Omega, \mathcal{A}, P\big)$ entonces

$$X_n \xrightarrow{c.s.} X \Rightarrow X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X.$$

Observación 1.14. *Todos los recíprocos de la proposición anterior son falsos.*

Teorema 1.15. Ley Fuerte de los grandes números: Sean $\{X_n\}$ v.a. sobre $\big(\Omega, \mathcal{A}, P\big)$ y X_n independientes idénticamentes distribuidas (i.i.d.) en L^1 y $\mu = E(X)$ entonces

$$\overline{X_n} = \frac{X_1 + \cdots + X_n}{n} \xrightarrow{c.s.} \mu$$

Teorema 1.16. Teorema Central del Límite: Sean $\{X_n\}$ definidas en $\big(\Omega, \mathcal{A}, P\big)$ v.a. i.i.d. en L^2 entonces

$$\frac{\overline{X_n} - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1).$$

Donde $N(0,1)$ denota la distribución normal con esperanza 0 y varianza 1.

Observación 1.17. *Si n es 'grande' y fijo, $F_{\overline{X_n}}$ se aproxima por la distribución $N(\mu, \frac{\sigma^2}{n})$ donde $\mu = E(X_n)$ y $\sigma^2 = Var(X_n)$*

Capítulo 2

Muestreo aleatorio simple

2.1. Algunas definiciones previas

Definición 2.1. Muestra aleatoria simple (M.A.S.): $X_1, \ldots, X_n$ v.a. definidas en $\big(\Omega, \mathcal{A}, P\big)$ son una muestra aleatoria simple si son independientes idénticamente distribuidas (i.i.d.).

Definición 2.2. Media Muestral y Varianza Muestral: dada $X_1, \ldots, X_n$ una M.A.S. definimos

1) **Media Muestral:** $\overline{X_n} = \dfrac{X_1 + \ldots, X_n}{n}$.

2) **Varianza Muestral:** $S_n^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i - \overline{X_n})^2$

Observación 2.3. $S_n^2 = \dfrac{1}{n-1}\big(\displaystyle\sum_{i=1}^{n} X_i^2 - n\overline{X_n}^2\big)$

Observación 2.4. *Si* $X \in L^1$, $\overline{X_n} \xrightarrow{c.s.} \mu = E(X)$ *por L.F.G.N.*

Observación 2.5. *Si* $X \in L^2$, $S_n^2 \xrightarrow{c.s.} \sigma^2 = Var(X)$.

Demostración.

$$S_n^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X_n}^2\right) \overset{c.s.}{\to} E(X^2) - \mu^2 = \sigma^2,$$

donde hemos usado la L.F.G.N. para las variables $Y_n = X_n^2$. □

2.2. Muestreo en poblaciones normales

Definición 2.6. Distribución Gamma: Decimos que X tiene distribución $Gamma(\alpha, \lambda)$ si su densidad es

$$f_X(x) = \begin{cases} \dfrac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}.$$

donde $\Gamma(\alpha)$ es la función Γ que se define como

$$\Gamma(\alpha + 1) = \int_0^{+\infty} t^\alpha e^{-t} dt.$$

Observación 2.7. *Tres propiedades importantes de la distribución Gamma son:*

1) Si $X \backsim Gamma(\alpha, \lambda)$ entonces $E(X) = \alpha/\lambda$ y $Var(X) = \alpha/\lambda^2$.

2) Si $X \backsim Gamma(\alpha, \lambda)$ e $Y \backsim Gamma(\beta, \lambda)$ y son independientes entonces $X+Y \backsim Gamma(\alpha+\beta, \lambda)$.

3) Si $\alpha = 1$, $Gamma(1, \lambda) = Exp(\lambda)$.

Definición 2.8. Distribución Ji cuadrado con k grados de libertad: Decimos que $X \backsim \chi_k^2$ si $X \backsim Gamma(k/2, 1/2)$ es decir si

$$f_X(x) = \frac{1}{\Gamma(k/2)2^{k/2}} x^{k/2-1} e^{-x/2} \mathbb{I}_{(0,+\infty)}(x)$$

En la figura 2.1 se grafica f_X para diferentes valores de k.

Observación 2.9. *Se puede demostrar que*

- $E(\chi_k^2) = k$
- $Var(\chi_k^2) = 2k$

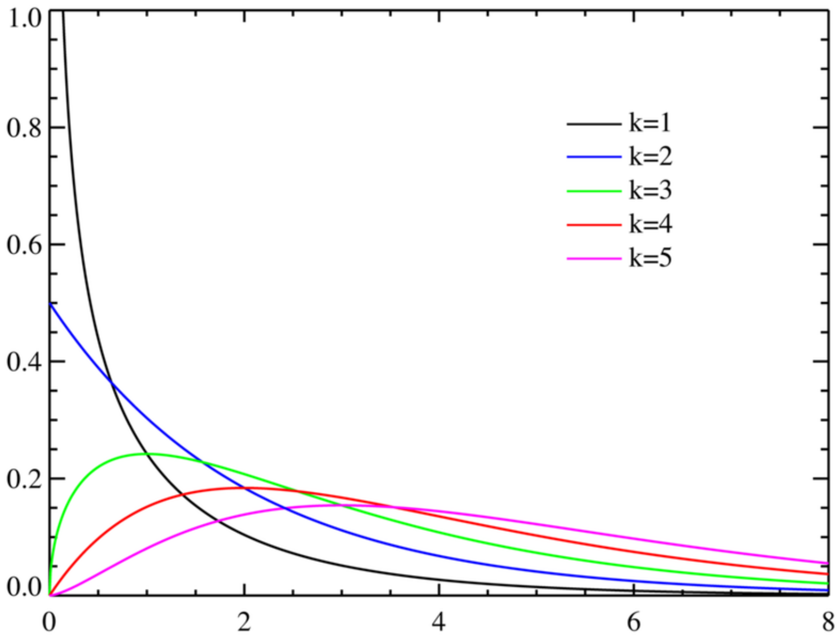

Figura 2.1: Gráfica de la distribución $\mathcal{X}^2$ para diferentes valores de k

Teorema 2.10. *Si $X_1, \ldots, X_n$ es una M.A.S. y $X \backsim N(0,1)$, entonces*

$$X_1^2 + \cdots + X_k^2 = \|(X_1, \ldots, X_k)\|^2 \backsim \chi_k^2$$

Demostración. Por la propiedad 2) de las distribuciones Gamma, es suficiente demostrar que $X_i^2 \backsim \chi_1^2$. Si $X \backsim N(0,1)$ entonces, tomando $t > 0$, $F_{X^2}(t) = P(X^2 \leq t) = P(|X| \leq \sqrt{t}) = P(-\sqrt{t} \leq X \leq \sqrt{t}) =$

$$\int_{-\sqrt{t}}^{\sqrt{t}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds = 2\int_0^{\sqrt{t}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-\frac{1}{2}u} \frac{1}{\sqrt{u}} du,$$

donde en la primera igualdad hemos usado que la función $e^{-\frac{1}{2}s^2}$ es par, y en la seguna hemos hecho el cambio de variable $u = s^2$, $2ds = 1/\sqrt{u}du$. Para concluir basta observar que

$$\frac{e^{-\frac{1}{2}u}}{\sqrt{2\pi}\sqrt{u}},$$

es la densidad de χ_1^2 pero esto se sigue de que $\Gamma(1/2) = \sqrt{\pi}$. $\square$

Definición 2.11. Distribución T-Student con k grados de libertad: Sean $X \backsim N(0,1)$ e $Y \backsim \chi_k^2$ independientes, la distribución de

$$T_k = \frac{X}{\sqrt{Y/k}},$$

se llama distribución $T - Student$ con k grados de libertad. Decimos que la variable T_k tiene distribución $T - Student$ no central, con parámetro de no centralidad $\mu > 0$ si

$$T_k = \frac{X+\mu}{\sqrt{Y/k}}$$

Observación 2.12. *Si $\mu = 0$ se verifica que*

- $E(T_k) = 0$
- $Var(T_k) = k/(k-2)$ *para* $k > 2$.

Teorema 2.13. *Sea $T \backsim T_k$, entonces la densidad es*

$$f_T(t) = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi}\Gamma(\frac{k}{2})} \frac{1}{\left(1 + \frac{t^2}{k}\right)^{\frac{k+1}{2}}}$$

Demostración. Tomemos el vector (X,Y), su densidad es

$$f_{X,Y}(x,y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \frac{y^{\frac{k}{2}-1}e^{-\frac{y}{2}}}{\Gamma(\frac{k}{2})2^{\frac{k}{2}}} \mathbb{I}_{(0,+\infty)}(y).$$

Sea $g : \mathbb{R} \times \mathbb{R}^+ \longrightarrow \mathbb{R} \times \mathbb{R}^+$ tal que

$$g(x,y) = \left(\frac{x}{\sqrt{y/k}}, y\right), \ g \text{ es difeomofismo y } g^{-1}(u,v) = \left(u\sqrt{v/k}, v\right),$$

tenemos entonces que $g(X,Y) = (U,V)$

$$f_{g(X,Y)}(u,v) = f_{U,V}(u,v) = f_{X,Y}\big(g^{-1}(u,v)\big) \frac{\mathbb{I}_{(0,+\infty)}(v)}{\big|\det J_g\big(g^{-1}(u,v)\big)\big|}$$

donde

$$J_g(x,y) = \begin{pmatrix} \frac{1}{\sqrt{y/k}} & \frac{u}{\sqrt{k}} \frac{1}{2\sqrt{v}} \\ 0 & 1 \end{pmatrix} \quad y \quad \big|\det J_g(x,y)\big| = \sqrt{k/y},$$

luego, sustituyendo

$$f_{U,V}(u,v) = \frac{1}{\sqrt{2\pi}} \frac{e^{-\frac{1}{2k}u^2 v} v^{\frac{k}{2}-1} e^{-\frac{v}{2}}}{\Gamma(\frac{k}{2})2^{k/2}} \mathbb{I}_{(0,+\infty)}(v) \frac{\sqrt{v}}{\sqrt{k}},$$

como $T = U$ tenemos que

$$f_U(u) = \int_{-\infty}^{+\infty} f_{U,V}(u,v)dv = \frac{1}{\sqrt{2k\pi}\Gamma(k/2)2^{\frac{k}{2}}} \int_0^{+\infty} v^{\frac{k-1}{2}} e^{-v\left(\frac{u^2}{2k}+\frac{1}{2}\right)} dv,$$

por otro lado sabemos que

$$\frac{\lambda^\alpha}{\Gamma(\alpha)}\int_0^{+\infty} x^{\alpha-1}e^{-\lambda x}dx = 1 \text{ entonces } \int_0^{+\infty} x^{\alpha-1}e^{-\lambda x}dx = \frac{\Gamma(\alpha)}{\lambda^\alpha},$$

si tomamos entonces $\alpha = \frac{k+1}{2}$ $x = v$ y $\lambda = \frac{u^2}{2k} + \frac{1}{2}$ se concluye la tesis. □

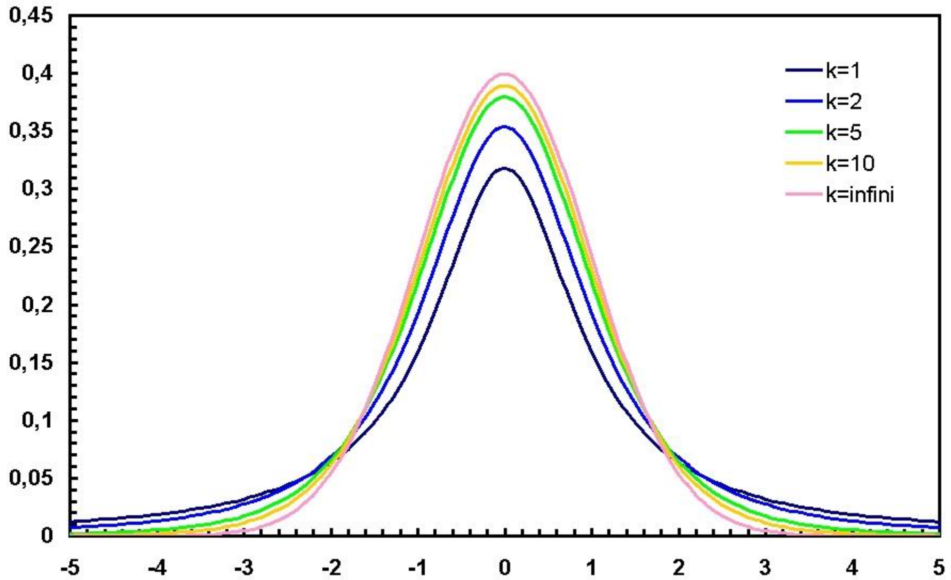

Figura 2.2: Gráfica de la densidad de una variable T_k de Student para diferentes valores de k, $k = \infty$ corresponde a la densidad de $N(0, 1)$

Teorema 2.14. *Sea $X_1, \ldots, X_n$ M.A.S. de $X \backsim N(\mu, \sigma^2)$, entonces*

1) $\overline{X_n} \backsim N\left(\mu, \dfrac{\sigma^2}{n}\right)$.

2) $\overline{X_n}$ *y* S_n^2 *son independientes.*

3) $\dfrac{n-1}{\sigma^2}S_n^2 \backsim \chi^2_{n-1}$.

4) $\sqrt{n}\dfrac{(\overline{X_n} - \mu)}{S_n} \backsim T_{n-1}$.

Demostración.

1) es inmediato

3) tomemos $\sigma = 1$, por inducción en n, para $n = 2$ tenemos que

$$\begin{aligned} S_2^2 &= \left(X_1 - \frac{X_1 + X_2}{2}\right)^2 + \left(X_2 - \frac{X_1 + X_2}{2}\right)^2 \\ &= \left(\frac{X_1 - X_2}{2}\right)^2 + \left(\frac{X_2 - X_1}{2}\right)^2 = \left(\frac{X_1 - X_2}{\sqrt{2}}\right)^2 \backsim \chi^2_1, \end{aligned}$$

Ya que $X_1 - X_2 \sim N(0,2)$. Supongamos cierto para $n-1$. Vamos a usar la igualdad

$$(n-1)S_n^2 = (n-2)S_{n-1}^2 + \frac{n-1}{n}(X_n - \overline{X_{n-1}})^2,$$

como estamos tomando $\sigma = 1$ tenemos que ver que $(n-1)S_n^2 \backsim \chi^2_{n-1}$ o lo que es lo mismo $(n-2)S_{n-1}^2 + \frac{n-1}{n}(X_n - \overline{X_{n-1}})^2 \backsim \chi^2_{n-1}$, por hipótesis de inducción $(n-2)S_{n-1}^2 \backsim \chi^2_{n-2}$, además $\frac{n-1}{n}(X_n - \overline{X_{n-1}})^2$ es independiente de $(n-2)S_{n-1}^2$ pues $\overline{X_{n-1}}$ es independiente de S_{n-1}^2 por la parte 2), y X_n es independiente de S_{n-1}^2 pues S_{n-1}^2 depende sólo de $X_1, \ldots, X_{n-1}$ y la muestra es una M.A.S.
Basta entonces ver que $\frac{n-1}{n}(X_n - \overline{X_{n-1}})^2 \backsim \chi^2_1$,

$$X_n - \overline{X_{n-1}} \backsim N\left(0, 1 + \frac{1}{n-1}\right) = N\left(0, \frac{n}{n-1}\right) \Rightarrow \frac{n-1}{n}(X_n - \overline{X_{n-1}})^2 \backsim \chi^2_1,$$

y, como la suma de χ^2 es tiene distribución χ^2 con la suma de los grados tenemos que $\chi^2_{n-2} + \chi^2_1 \backsim \chi^2_{n-1}$.

4) Es inmediato a partir de 1,2 y 3.

2)

$$\begin{aligned} S_n^2 = & \frac{1}{n-1}\Big((X_1 - \overline{X_n})^2 + \sum_{i=2}^{n}(X_i - \overline{X_n})^2\Big) \\ = & \frac{1}{n-1}\Big((\sum_{i=2}^{n}(X_i - \overline{X_n}))^2 + \sum_{i=2}^{n}\big(X_i - \overline{X_n}\big)^2\Big), \end{aligned}$$

hemos escrito entonces S_n^2 en función de $X_2 - \overline{X_n}, \ldots, X_n - \overline{X_n}$, basta demostrar entonces que $\overline{X_n}$ es independiente de $X_2 - \overline{X_n}, \ldots, X_n - \overline{X_n}$.
Consideremos $Y_1 = \overline{X_n}, Y_2 = X_2 - \overline{X_n}, \ldots, Y_n = X_n - \overline{X_n}$, $y_1 = \overline{x_n}, y_2 = x_2 - \overline{x_n}, \ldots, y_n = x_n - \overline{x_n}$ y $y = g(x_1, \ldots, x_n)$ entonces

$$J_g = \begin{pmatrix} \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ -\frac{1}{n} & 1-\frac{1}{n} & \cdots & -\frac{1}{n} \\ \vdots & \vdots & & \\ -\frac{1}{n} & -\frac{1}{n} & \cdots & 1-\frac{1}{n} \end{pmatrix}.$$

Es fácil ver que $det(J_g) = 1/n$, basta sumar la primer fila a las demas, y queda una matriz triangular superior con diagonal $1/n, 1, \ldots, 1$.
$x_2 = y_2 + y_1, \ldots, x_n = y_n + y_1$ de donde

$$g^{-1}(y) = \left(-\sum_2 y_i - y_1, y_2 + y_1, \ldots, y_n + y_1\right),$$

entonces

$$\begin{aligned} f_Y(y) = & f_X(g^{-1}(y))\frac{1}{|\det J_g(g^{-1}(y))|} \\ = & \frac{n}{(2\pi)^{n/2}}\exp\left\{-\frac{1}{2}\left(y_1^2 - 2y_1\sum_2 y_i + (\sum_2 y_i)^2 + \sum_2(y_i^2 + 2y_1y_i + y_1^2)\right)\right\} \\ = & \frac{n}{(2\pi)^{n/2}}\exp\left\{-\frac{n}{2}y_1^2\right\}\exp\left\{-\frac{1}{2}\left((\sum_2 y_i)^2 + \sum_2 y_i^2\right)\right\}. \end{aligned}$$

Por lo tanto factorizamos respecto de y_1, iterando, son independientes.

□

Observación 2.15. **Distribución F de Fisher:** Sea $X \backsim \chi^2_n$ e $Y \backsim \chi^2_m$ independientes, la distribución de

$$\frac{X/n}{Y/m}$$

se denomina distribución F de Fisher de parámetros n y m, y la anotamos $F(n,m)$.

A modo de motivación geométrica de la distribución F de Fisher, vamos a enunciar el siguiente teorema.

Teorema 2.16. *Sea A el ángulo que forma un vector $X \backsim N(0,\sigma^2)$ en $\mathbb{R}^d$ con un subespacio $\mathcal{R}$ de dimensión ρ, entonces*

$$\frac{\rho}{d-\rho}\tan^2(A) \backsim F(d-\rho,\rho)$$

Observación 2.17. *Si $Z \backsim F(n,m)$ entonces*

$$f_Z(x) = \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)}\left(\frac{n}{m}\right)^{\frac{n}{2}}\frac{x^{\frac{n}{2}-1}}{\left(1+\frac{n}{m}x\right)^{\frac{n+m}{2}}}\mathbb{I}_{(0,+\infty)}(x)$$

Teorema 2.18. *Sea $X_1,\ldots,X_n$ M.A.S. de $X \backsim N(\mu_X,\sigma^2_X)$ y $Y_1,\ldots,Y_n$ M.A.S. de $Y \backsim N(\mu_Y,\sigma^2_Y)$ X e Y independientes, entonces*

$$\frac{S^2_X/\sigma^2_X}{S^2_Y/\sigma^2_Y} \backsim F(n-1,m-1).$$

Demostración. La demostración se sigue de la parte 3) y 2) del teorema 2.14 □

2.3. Estadísticos de Orden para una M.A.S.

Definición 2.19. **Muestra Ordenada:** Sea $X_1,\ldots,X_n$ una M.A.S. de $X \backsim F_X$, definimos

$$\begin{aligned}
X_{1:n} &= \min\{X_1,\ldots,X_n\} \\
X_{2:n} &= \min\left(\{X_1,\ldots,X_n\}\setminus\{X_{1:n}\}\right) \\
&\vdots \\
X_{n:n} &= \min\left(\{X_1,\ldots,X_n\}\setminus\{X_{1:n},\ldots,X_{n-1:n}\}\right)
\end{aligned}$$

se tiene entonces que $X_{1:n} \leq \cdots \leq X_{n:n}$

Teorema 2.20. **Distribución de los percentiles:** Sea $X_1,\ldots,X_n$ una M.A.S. de X absolutamente continua, entonces

$$f_{X_{j:n}}(x) = \frac{n!}{(j-1)!(n-j)!}f_X(x)\big(F_X(x)\big)^{j-1}\big(1-F_X(x)\big)^{n-j}$$

Demostración. $F_{X_{j:n}}(x) = P(X_{j:n} \leq x)$ es decir, es la probabilidad de que al menos j variables sean menores o iguales que x. Consideremos Y la cantidad de observaciones que son menores o iguales que x, entonces $Y \backsim Bin(n,p)$ con $p = F_X(x)$.

$$P(X_{j:n} \leq x) = P(Y \geq j) = \sum_{k=j}^{n} P(Y=k) = \sum_{k=j}^{n} C^n_k\big(F_X(x)\big)^k\big(1-F_X(x)\big)^{n-k}$$

entonces, derivando y usando $q = 1 - p$

$$\begin{aligned}
f_{X_{j:n}}(x) &= \sum_{k=j}^{n} \frac{n!}{(n-k)!k!}\Big(kp^{k-1}f_X(x)q^{n-k} - f_X(x)(n-k)q^{n-k-1}p^k\Big) \\
&= f_X(x)n!\left(\sum_{k=j}^{n} \frac{1}{(n-k)!(k-1)!}p^{k-1}q^{n-k} - \sum_{k=j}^{n-1} \frac{1}{(n-k-1)!k!}p^k q^{n-k-1}\right) \\
&= f_X(x)n!\left(\sum_{k=j}^{n} \frac{1}{(n-k)!(k-1)!}p^{k-1}q^{n-k} - \sum_{k=j+1}^{n} \frac{1}{(n-k)!(k-1)!}p^{k-1}q^{n-k}\right) \\
&= f_X(x)n!\frac{1}{(n-j)!(j-1)!}p^{j-1}q^{n-j}
\end{aligned}$$

□

Observación 2.21. $f_{X_{max}}(x) = nf_X(x)\big(F_X(x)\big)^{n-1}$ *y* $f_{X_{min}}(x) = nf_X(x)\big(1 - F_X(x)\big)^{n-1}$

Definición 2.22. Distribución beta: Si X tiene densidad $f(x) = \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}\mathbb{I}_{(0,1)}(x)$ decimos que $X \backsim Beta(\alpha, \beta)$

Observación 2.23. *Si* $X \backsim Beta(\alpha,\beta)$ *entonces* $E(X) = \alpha/(\alpha+\beta)$ *y* $Var(X) = \dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

Observación 2.24. *Si* $X_1, \dots, X_n$ *es una M.A.S. de* $X \backsim U_{[0,1]}$ *entonces* $X_{j:n} \backsim Beta(j, n-j+1)$.

Capítulo 3

Métodos paramétricos de estimación

3.1. Algunas definiciones previas

Consideremos el caso en que tenemos $X_1, \dots, X_n$ M.A.S. de $X \backsim F_X(x|\theta)$ donde $\theta \in \Theta \subset \mathbb{R}^k$ es un parámetro desconocido.

Definición 3.1. Sea $\hat{\theta} : \mathbb{R}^n \longrightarrow \Theta$ medible, independiente de θ, entonces $\hat{\theta}(X_1, \dots, X_n) : \Omega \longrightarrow \Theta$ es un estimador de θ.

Ejemplo 3.2. Si $X \backsim N(\mu, \sigma^2)$ y $\theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}^+$ entonces si definimos

$$\hat{\theta}(x_1, \dots, x_n) = \left(\frac{x_1 + \cdots + x_n}{n}, \frac{1}{n-1} \sum_{i=1}^{n} x_i - \frac{x_1 + \cdots + x_n}{n} \right) : \mathbb{R}^n \longrightarrow \Theta$$

entonces

$$\hat{\theta}(X_1, \dots, X_n) = (\overline{X_n}, S_n^2)$$

es un estimador de θ.

Observemos que si bien θ es un vector, $\hat{\theta}$ es un vector aleatorio a valores en $\mathbb{R}^k$.

Definición 3.3. Si $X_1, \dots, X_n$ es una M.A.S. de $X \backsim F_X(X|\theta)$ y $\hat{\theta}$ es un estimador, decimos que $\hat{\theta}$ es debilmente consistente si $\hat{\theta} \xrightarrow{P} \theta$. Decimos que es fuertemente consistente si $\hat{\theta} \xrightarrow{c.s.} \theta$

Ejemplo 3.4. Si $X \backsim N(\mu, \sigma^2)$ y $\hat{\theta} = (\overline{X_n}, S_n^2)$ entonces $\hat{\theta}$ es fuertemente consistente.

3.2. Método de los momentos

Si $X_1, \dots, X_n$ es una M.A.S. de $X \backsim F(x|\theta)$ y $\theta = (\theta_1, \dots, \theta_k) \in \mathbb{R}^k$ y $X \in L^k$. Consideremos el sistema

$$\begin{cases} E(X) & = & \overline{X_n} \\ E(X^2) & = & \frac{1}{n} \sum_{i=1}^{n} X_i^2 \\ \vdots & & \vdots \\ E(X^k) & = & \frac{1}{n} \sum_{i=1}^{n} X_i^k \end{cases}$$

Los $E(X^k)$ se llaman momentos poblacionales y las expresiones al otro lado de la igualdad, momentos muestrales. Los θ_i aparecen en los momentos poblacionales y si despejamos las k incognitas de las k ecuaciones obtenemos los estimadores. Dicho sistema no necesariamente tiene que tener solución ni ser única. Observemos que por la ley fuerte, los estimdores que se despejan para cada θ_i son consistentes.

Ejemplo 3.5. Sea $X_1, \dots, X_n$ M.A.S. de $X \backsim \mathit{Gamma}(\alpha, 1/\beta)$, entonces $E(X_1) = \alpha\beta$ y $E(X_1^2) = \beta^\alpha \alpha(\alpha+1)$, consideremos

$$m_1 = \frac{X_1 + \cdots + X_n}{n}$$

y

$$m_2 = \frac{X_1^2 + \cdots + X_n^2}{n},$$

planteamos $m_1 = \alpha\beta$ y $m_2 = \beta^\alpha\alpha(\alpha+1)$ y obtenemos los estimadores de α y de β :

$$\alpha = \frac{m_1^2}{m_2 - m_1^2}$$

y

$$\beta = \frac{m_2 - m_1^2}{m_1}.$$

Ejemplo 3.6. Sea $X_1, \dots, X_n$ M.A.S. de $X \backsim U_{[a,b]}$ y $\theta = (a, b)$ entonces el método de los momentos es

$$\begin{cases} 1/2(b-a) & = & \overline{X_n} \\ 1/12(b-a)^2 + 1/4(a+b)^2 & = & 1/n \sum X_i^2 =: M_2 \end{cases}$$

Si despejamos b en la primer ecuación y sustituimos en la segunda obtenemos las soluciones

$$a = \overline{X_n} \pm \sqrt{3(M_2 - \overline{X_n})}, \quad b = \overline{X_n} \pm \sqrt{3(M_2 - \overline{X_n})}$$

Como $\Theta = \{(a, b) \in \mathbb{R}^2 : a < b\}$ descartamos soluciones y nos queda

$$\hat{a} = \overline{X_n} - \sqrt{3(M_2 - \overline{X_n})}, \quad \hat{b} = \overline{X_n} + \sqrt{3(M_2 - \overline{X_n})}.$$

Notación: $\frac{1}{n}\sum_{i=1}^n X_i^k =: M_k$

Teorema 3.7. Método de los momentos, existencia y unicidad de solución, consistencia: Si $F : \Theta \subset \mathbb{R}^k \longrightarrow \mathbb{R}^k$ es tal que $(E(X), E(X^2), \dots, E(X^k)) = F(\theta_1, \dots, \theta_k)$, entonces, si F es inyectiva, $F^{-1} : F(\Theta) \longrightarrow \Theta$ es continua y si $M_1, M_2, \dots, M_k \in F(\Theta)$ *c.s.* entonces los estimadores por momentos existen, son únicos y convergen *c.s.* a $\theta_1, \dots, \theta_k$.

Demostración. Los estimadores por el método de los momentos son $(\hat{\theta_1}, \dots, \hat{\theta_k}) = F^{-1}(M_1, \dots, M_k)$, como

$$\begin{aligned} M_1 = \overline{X_n} &\xrightarrow{c.s.} E(X) \\ &\vdots \\ M_k = \frac{1}{n}\sum_i X_i^k &\xrightarrow{c.s.} E(X^k) \end{aligned}$$

y F^{-1} es continua entonces

$$\begin{aligned} (\hat{\theta_1}, \dots, \hat{\theta_k}) = F^{-1}(M_1, \dots, M_k) \xrightarrow{c.s.} & \; F^{-1}\big(E(X), \dots, E(X^k)\big) \\ & = F^{-1}\big(F(\theta_1, \dots, \theta_k)\big) = (\theta_1, \dots, \theta_k), \end{aligned}$$

de donde $\hat{\theta}$ es fuertemente consistente. □

3.3. Método de Máxima Verosimilitud

Definición 3.8. Función de Verosimilitud: Dada $X_1, \ldots, X_n$ una M.A.S. de $X \backsim F(x|\theta)$ $\theta \in \Theta \subset \mathbb{R}^k$ se define $L : \Theta \times \mathbb{R}^n \longrightarrow \mathbb{R}$

$$L(\theta, \tilde{x}) = \prod_{i=1}^{n} f_X(x_i|\theta) \text{ si } X \text{ es absolutamente continua}$$
$$L(\theta, \tilde{x}) = \prod_{i=1}^{n} p_X(x_i|\theta) \text{ si es discreta}$$

donde $\tilde{x} = (x_1, \ldots, x_n)$.

El método consiste entonces en hallar $\hat{\theta} \in \Theta$ donde se realice $\text{máx}_{\theta\in\Theta} L(\theta, \tilde{x})$, dicho $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ es el estimador de máxima verosimilitud (E.M.V.) de θ. El método no asegura la existencia y/o unicidad de $\hat{\theta}$.

Ejemplo 3.9. Sea $X_1, \ldots, X_n$ una M.A.S. de $X \backsim \exp(\lambda)$ entonces la función de verosimilitud para λ es

$$L(\lambda) = \prod_{i=1}^{n} \lambda \exp\{-\lambda x_i\} = \lambda^n \exp\Big\{ -\lambda \sum_i x_i \Big\},$$

con $x_i \geq 0\ \forall i$, derivando obtenemos

$$L'(\lambda) = \lambda^{n-1} \exp\left\{ -\lambda \sum_i x_i \left(n - \lambda \sum_i x_i \right) \right\},$$

y por lo tanto, como $\lambda \neq 0$, si hacemos $L'(\lambda) = 0$ obtenemos $\lambda = \dfrac{n}{\sum_i x_i}$, es fácil ver, mirando el signo de $L'(\lambda)$ que es un máximo. Por lo tanto $\hat{\lambda} = \frac{1}{\overline{X}_n}$ es el E.M.V. de λ.

Ejemplo 3.10. Sea $X_1, \ldots, X_n$ una M.A.S. de $X \backsim U_{[0,b]}$ $\Theta = \{b > 0\}$, la función de verosimilitud es entonces

$$L(b) = \prod_{i=1}^{n} \frac{1}{b} \mathbb{I}_{[0,b]}(x_i) = \begin{cases} \dfrac{1}{b^n} & \text{si } 0 < x_1, \ldots, x_n < b \\ 0 & \text{si no} \end{cases} = \begin{cases} \dfrac{1}{b^n} & \text{si } b > \text{máx}\{x_1, \ldots, x_n\} \\ 0 & \text{si no} \end{cases}$$

Como la función $1/b^n$ es decreciente obtenemos que $\hat{b} = x_{n:n} = \text{máx}\{x_1, \ldots, x_n\}$.

Observación 3.11. Interpretación del método: Para el caso discreto, si tenemos $X_1, \ldots, X_n$ una M.A.S. y $X \backsim p_X(X|\theta)$ entonces

$$L(\theta|\tilde{x}) = \prod_{i=1}^{n} p_X(x_i|\theta) = \prod_{i=1}^{n} P(X = x_i|\theta) = P(X_1 = x_1, \ldots, X_n = x_n|\theta),$$

esto es, la probabilidad de que salga la muestra $(x_1, \ldots, x_n)$ dado el parámetro θ. El método consiste entonces en encontrar el θ que hace que la muestra que obtuvimos sea la más probable.
Otra forma de ver esto es observar que, de la desigualdad de Jensen se deduce que

$$E_{g(x|\theta_0)} \left(\log \left(\frac{g(x|\theta_1)}{g(x|\theta_0)} \right) \right) \leq \log \left(E_{g(x|\theta_0)} \left(\frac{g(x|\theta_1)}{g(x|\theta_0)} \right) \right) = 0$$

Por lo tanto

$$E_{g(x|\theta_0)} \Big(\log(g(x|\theta_1)) \Big) \leq E_{g(x|\theta_0)} \Big(\log(g(x|\theta_0)) \Big)$$

lo cual significa que la verosimilitud bajo modelo real $g(x|\theta_0)$ es mayor o igual que bajo cualquier otro valor del parámetro.

Principio de invarianza del E.M.V.: Supongamos que tenemos un parámetro $\theta \in \Theta$ y $g : \Theta \to \mathbb{R}$, y que estamos interesados en estimar $g(\theta)$ por el método de máxima verosimilitud, es decir queremos encontrar $\hat{M} = g(\tilde{\theta})$ que haga que la muestra sea más probable. Queremos maximizar entonces

$$L^*(M|\tilde{x}) = \sup_{\{\theta : g(\theta)=M\}} L(\theta|\tilde{x}),$$

Veremos que si $\hat{M} = E.M.V.$ de $g(\theta)$, es decir donde se realiza el máximo de L^* entonces $\hat{M} = g(\hat{\theta})$ siendo $\hat{\theta} = E.M.V$ de θ. En efecto:

$$L^*(\hat{M}|\tilde{x}) = \sup_M L^*(M|\tilde{x}) = \sup_M \sup_{\{\theta:g(\theta)=M\}} L(\theta|\tilde{x}) = \sup_{\theta\in\Theta} L(\theta, \tilde{x}) = L(\hat{\theta}|\tilde{x})$$

y

$$L^*(g(\hat{\theta}), \tilde{x}) = \sup_{\{\theta:g(\theta)=g(\hat{\theta})\}} L(\theta|\tilde{x}) = L(\hat{\theta}|\tilde{x}).$$

Entonces $g(\hat{\theta})$ es E.M.V. de $g(\theta)$.

Ejemplo 3.12. Sea $X_1, \ldots, X_n \backsim Ber(p)$, el E.M.V. de p es $\hat{p} = \overline{X_n}$, como $\sigma^2 = p(1-p) = g(p)$ por el Principio de Invarianza $\hat{\sigma}^2 = g(\hat{p}) = \hat{p}(1-\hat{p})$.

Observación 3.13. *Si* $h(\theta) = \log(L(\theta)) = \sum \log(f_X(x_i|\theta))$ *podemos, dado que* $\log(x)$ *es una función creciente, tomar el* θ *que maximiza* $h(\theta)$

Teorema 3.14. Consistencia del E.M.V.: Sea $X_1, \ldots, X_n, \ldots$ $i.i.d \backsim f(x|\theta)$ y $\theta \in \Theta \subset \mathbb{R}$ donde Θ es tal que si θ_0 es el valor exacto de θ entonces $\exists \delta > 0$ tal que $(\theta_0 - \delta, \theta_0 + \delta) \subset \Theta$, si $h_n(\theta) = \log(L(\theta, \tilde{X}))$ es derivable como función de θ y además $f(x|\theta) = f(x|\theta')$ implica $\theta = \theta'$ $c.s.$ entonces

$$\text{para casi todo } \omega\ \exists \theta_{n_k} = \theta_{n_k}(\omega) \in \Theta \text{ tal que } \frac{\partial}{\partial\theta} h_{n_k}(\theta_{n_k}) = 0 \quad y \quad \theta_{n_k} \longrightarrow \theta_0$$

Demostración.

$$h_n(\theta_0 - \delta) - h_n(\theta_0) = \sum_{i=1}^{n} \log((f(X_i|\theta_0 - \delta)) - \sum_{i=1}^{n} \log((f(X_i|\theta_0))) = \sum_{i=1}^{n} \log\left(\frac{f(X_i|\theta_0 - \delta)}{f(X_i|\theta_0)}\right)$$

entonces

$$\frac{h_n(\theta_0 - \delta) - h_n(\theta_0)}{n} = \frac{1}{n} \sum \log\left(\frac{f(X_i|\theta_0 - \delta)}{f(X_i|\theta_0)}\right) \overset{L.F.G.N.}{\longrightarrow} E\left(\log\left(\frac{f(X|\theta_0 - \delta)}{f(X|\theta_0)}\right)\right) \ c.s.,$$

como $-\log$ es una función convexa estricta y f es inyectiva en θ, usando la desigualdad de Jensen

$$E\left(\log\left(\frac{f(X|\theta_0 - \delta)}{f(X|\theta_0)}\right)\right) < \log\left(E\left(\frac{f(X|\theta_0 - \delta)}{f(X|\theta_0)}\right)\right),$$

por otro lado, como $X \backsim f(x|\theta_0)$

$$E\left(\frac{f(X|\theta_0 - \delta)}{f(X|\theta_0)}\right) = \int_{-\infty}^{+\infty} \frac{f(x|\theta_0 - \delta)}{f(x|\theta_0)} f(x|\theta_0) dx = \int_{-\infty}^{+\infty} f(x|\theta_0 - \delta) dx = 1.$$

Luego el límite anterior es negativo. Lo mismo para $\theta_0 + \delta$. Definamos

$$A_\delta = \left\{\omega \in \Omega : \frac{h_n(\theta_0 \pm \delta) - h_n(\theta_0)}{n} \longrightarrow E\left(\log\left(\frac{f(X|\theta_0 \pm \delta)}{f(X|\theta_0)}\right)\right) < 0\right\}.$$

Por la L.F.G.N. $P(A_\delta) = 1$. Fijado $\omega \in A_\delta$ existe $n_0 = n_0(\omega, \delta)$ tal que $h_n(\theta_0) > h_n(\theta_0 \pm \delta)$ $\forall n \geq n_0$, y entonces existe $\theta_n \in (\theta_0 - \delta, \theta_0 + \delta)$ tal que $\dfrac{\partial}{\partial\theta} h_n(\theta_n) = 0$ $\forall n \geq n_0$ porque h_n es derivable respecto a θ. Definamos

$$B_\delta = \left\{\omega \in \Omega : \exists n_0 \ y \ \theta_n \in (\theta_0 - \delta, \theta_0 + \delta), \ \frac{\partial}{\partial\theta} h_n(\theta_n) = 0 \ \forall n \geq n_0\right\}.$$

Como $A_\delta \subset B_\delta$ tenemos que $P(B_\delta) = 1$, si tomamos $\delta = 1/k$,

$$B = \bigcap_{k=1}^{\infty} B_{1/k}, \ P(B) = 1.$$

Sea $\omega \in B$, vamos a construir θ_{n_k}:
$\omega \in B_1$, entonces $\exists \theta_{n_1} \in (\theta_0 - 1, \theta_0 + 1)$ tal que $\dfrac{\partial}{\partial\theta} h_{n_1}(\theta_{n_1}) = 0$.
Supongamos que tenemos definido $\theta_{n_{k-1}}$, $\omega \in B_{1/k}$ entonces $\exists \theta_{n_k} \in (\theta_0 - \frac{1}{k}, \theta_0 + \frac{1}{k})$ con $n_k > n_{k-1}$ tal que $\dfrac{\partial}{\partial\theta} h_{n_k}(\theta_{n_k}) = 0$.
La sucesión θ_{n_k} verifica $\theta_{n_k} \longrightarrow \theta_0$ y es cero de $\frac{\partial}{\partial\theta} h_{n_k}$.
Observe que los subíndices de la sucesión dependen de ω. □

Observación 3.15. *El teorema anterior no asegura la existencia ni la unicidad del E.M.V.*

Lema 3.16. Lema de Slutsky: Si $X_n \xrightarrow{P} c$ y $Y_n \xrightarrow{d} Y$ con c constante entonces $X_n + Y_n \xrightarrow{d} c + Y$ y $X_n Y_n \xrightarrow{d} cY$.
Recordemos que $X_n \xrightarrow{d} c \Leftrightarrow X_n \xrightarrow{P} c$.

Teorema 3.17. Normalidad asintótica del E.M.V: Sea $X_1, \ldots, X_n, \ldots$ una M.A.S. de $X \backsim f(x|\theta)$, supongamos que existe $\delta > 0$ tal que $(\theta_0 - \delta, \theta_0 + \delta) \in \Theta$ donde θ_0 es el valor exacto de θ. Si se cumplen, para todo $\theta \in (\theta_0 - \delta, \theta_0 + \delta)$

1) $\exists \{\theta_n\}$ variables aleatorias tal que $\dfrac{\partial}{\partial\theta} h(\theta_n) = 0$ $\forall n$ y $\theta_n \xrightarrow{c.s.} \theta_0$

2) $\dfrac{\partial^3}{\partial\theta^3} h(\theta|\tilde{X}) \leq Y$ con $E(Y) < \infty$.

3) $E\left(\dfrac{\frac{\partial}{\partial\theta} f(X|\theta)}{f(X|\theta)}\right) = 0.$

4) $E\left(\dfrac{\frac{\partial^2}{\partial\theta^2} f(X|\theta)}{f(X|\theta)}\right) = 0.$

5) $i(\theta) := E\left(\dfrac{\frac{\partial}{\partial\theta} f(X|\theta)}{f(X|\theta)}\right)^2 > 0$, el número i se denomina número de información de Fischer.

Entonces

$$\sqrt{n}(\theta_n - \theta_0) \xrightarrow{d} N\left(0, \frac{1}{i(\theta_0)}\right)$$

Demostración. La demostración será una consecuencia de dos afirmaciones:

Afirmación 1: $\dfrac{1}{\sqrt{n}} \dfrac{\partial}{\partial\theta} h(\theta_0) \to N(0, i(\theta_0))$

Afirmación 2: $\sqrt{n}(\theta_n - \theta_0) - \dfrac{1}{\sqrt{n} i(\theta_0)} \dfrac{\partial}{\partial\theta} h(\theta_0) \xrightarrow{P} 0.$

Veamos primero cómo, a partir de estas afirmaciones, usando el Lema de Slutsky se concluye la tesis. En efecto, podemos escribir

$$\sqrt{n}(\theta_n - \theta_0) = \left(\sqrt{n}(\theta_n - \theta_0) - \frac{1}{\sqrt{ni(\theta_0)}}\frac{\partial}{\partial\theta}h(\theta_0)\right) + \frac{1}{\sqrt{ni(\theta_0)}}\frac{\partial}{\partial\theta}h(\theta_0)$$

Veamos la demostración de la Afirmación 1:

$$\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}h(\theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\frac{\partial}{\partial\theta}f(X_i|\theta_0)}{f(X_i|\theta_0)} = \sqrt{n}\frac{1}{n}\sum_{i=1}^{n}\frac{\frac{\partial}{\partial\theta}f(X_i|\theta_0)}{f(X_i|\theta_0)} = \sqrt{n}\ \overline{Z}_n.$$

$E(Z_i) = 0$ por la hipótesis 3) y $Var(Z_i) = E(Z_i^2) - E^2(Z_i) = i(\theta_0) > 0$ por la hipótesis 5). Luego, si aplicamos el *T.C.L.* tenemos que $\sqrt{n}\ \overline{Z}_n \xrightarrow{d} N(0, i(\theta_0))$. Lo que concluye la demostración de la afirmación 1.

Veamos la demostración de la Afirmación 2: podemos escribir, usando el desarrollo de Taylor y la hipótesis 1,

$$0 = \frac{\partial}{\partial\theta}h(\theta_n) = \frac{\partial}{\partial\theta}h(\theta_0) + \frac{\partial^2}{\partial\theta^2}h(\theta_0)(\theta_n - \theta_0) + \frac{\partial^3}{\partial\theta^3}h(\hat{\theta_n})\frac{(\theta_n - \theta_0)^2}{2}$$

donde $\hat{\theta_n} \in [\theta_0, \theta_n]$, despejando obtenemos

$$\theta_n - \theta_0 = \frac{-\frac{\partial}{\partial\theta}h(\theta_0)}{\frac{\partial^2}{\partial\theta^2}h(\theta_0) + \frac{\partial^3}{\partial\theta^3}h(\hat{\theta_n})\frac{(\theta_n-\theta_0)}{2}}$$

y

$$\sqrt{n}(\theta_n - \theta_0) - \frac{1}{\sqrt{ni(\theta_0)}}\frac{\partial}{\partial\theta}h(\theta_0) = \frac{-\frac{\partial}{\partial\theta}h(\theta_n)\sqrt{n}}{\frac{\partial^2}{\partial\theta^2}h(\theta_0) + \frac{\partial^3}{\partial\theta^3}h(\hat{\theta_n})\frac{(\theta_n-\theta_0)}{2}} - \frac{1}{\sqrt{ni(\theta_0)}}\frac{\partial}{\partial\theta}h(\theta_0) =$$

$$\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}h(\theta_0)\left[\frac{-1}{\frac{1}{n}\frac{\partial^2}{\partial\theta^2}h(\theta_0) + \frac{1}{n}\frac{\partial^3}{\partial\theta^3}h(\hat{\theta_n})\frac{(\theta_n-\theta_0)}{2}} - \frac{1}{i(\theta_0)}\right] \tag{3.1}$$

Nuevamente, como $\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}h(\theta_0) \xrightarrow{d} N(0, i(\theta_0))$, por el lema de Slutsky, la afirmación 2 queda demostrada si probamos que la expresión entre [] tiende en probabilidad a 0 (o lo que es lo mismo, en distribución a 0).
Sabemos que

$$\frac{1}{n}\frac{\partial^3}{\partial\theta^3}h(\hat{\theta_n})\frac{(\theta_n - \theta_0)}{2} \xrightarrow{P} 0,$$

donde hemos usado que si $X_n \xrightarrow{P} 0$ y si $E(Y_n) \leq k\ \forall n$ entonces $X_nY_n \xrightarrow{P} 0$. (Hipótesis 2)

$$\frac{1}{n}\frac{\partial^2}{\partial\theta^2}h(\theta_0) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\left(\frac{\frac{\partial}{\partial\theta}f(X_i|\theta_0)}{f(X_i|\theta_0)}\right) = \frac{1}{n}\sum_{i=1}^{n}\frac{\left(\frac{\partial^2}{\partial\theta^2}f(X_i|\theta_0)\right)f(X_i|\theta_0) - \left(\frac{\partial}{\partial\theta}f(X_i|\theta_0)\right)^2}{\left(f(X_i|\theta_0)\right)^2}.$$

Si aplicamos ahora la *L.F.G.N* el promedio anterior tiende a su esperanza, que es, aplicando las hipótesis 4 y 5:

$$E\left(\frac{\left(\frac{\partial^2}{\partial\theta^2}f(X|\theta_0)\right)f(X|\theta_0)}{\left(f(X|\theta_0)\right)^2}\right) - E\left(\frac{\frac{\partial}{\partial\theta}f(X|\theta)}{f(X|\theta)}\right)^2 = -i(\theta_0)$$

de donde se concluye que la expresión entre [] en 3.1. converge en probabilidad a 0 como queríamos demostrar. □

Observación 3.18. Sobre las hipótesis del teorema anterior

3) $E\left(\dfrac{\frac{\partial}{\partial\theta}f(x|\theta)}{f(x|\theta)}\right) = \displaystyle\int_{-\infty}^{+\infty} \dfrac{\frac{\partial}{\partial\theta}f(x|\theta)}{f(x|\theta)} f(x|\theta)dx = \int_{-\infty}^{+\infty} \dfrac{\partial}{\partial\theta} f(x|\theta)dx$, observemos que si pudieramos aplicar convergencia dominada $\dfrac{\partial}{\partial\theta}\displaystyle\int_{-\infty}^{+\infty} f(x|\theta)dx = \dfrac{\partial}{\partial\theta}\, 1 = 0.$

4) Análogo a 3).

5) Por 3), 5) es pedir que $\dfrac{\frac{\partial}{\partial\theta}f(x|\theta)}{f(x|\theta)}$ no sea constante.

Observación 3.19. *Un estudio mas detallado del E.M.V se puede encontrar en [**?**] donde se incluye además el caso en que el parámetro θ a estimar es vectorial.*

3.4. Método de estimación por cuantiles

Definición 3.20. Cuantil o percentil p: Sea X v.a., dado $p \in (0,1)$ el cuantil p es

$$x_p = \inf\left\{x \in \mathbb{R} : F_X(x) \geq p\right\}$$

Observación 3.21. *x_p existe, y es mínimo*

Demostración. Es el infimo de un conjunto acotado inferiormente, por lo tanto existe. Si $\{x_n\}$ es tal que $F(x_n) \geq p$ y $x_n \to x_p^+$, como F es continua por derecha

$$\lim_n F(x_n) = F(\lim_n x_n) = F(x_p) \geq p.$$

□

Definición 3.22. Percentil empírico: Sea $X_1, \ldots, X_n$ M.A.S. de X, consideremos la muestra ordenada $X_1^* = X_{1:n} \leq \cdots \leq X_n^* = X_{n:n}$, entonces

$$\hat{X_p} = \begin{cases} X_{np}^* & \text{si } np \in \mathbb{N} \\ X_{[np]+1}^* & \text{si } np \notin \mathbb{N} \end{cases}$$

El método consiste en plantear la función $g(\theta) = \sum_{i=1}^k (\hat{X_{p_i}} - x_{p_i})^2$ donde los p_i y k son cualquiera. Lo que se busca es el mínimo de $g(\theta)$. El argumento que minimiza $g(\theta)$ sera $\hat{\theta}$ y dependerá de los cuantiles empíricos $\hat{X_{p_i}}$.

Ejemplo 3.23. Si $X \backsim (\mu, \sigma^2)$, entonces $f_X(x|\mu,\sigma^2) = \dfrac{1}{\pi\sigma\left(1 + \left(\frac{x-\mu}{\sigma}\right)^2\right)}$.

Es fácil ver que $E(X) = \infty$ y que su mediana es μ. Vamos a estimar $\theta = (\mu, \sigma^2)$ por el método de cuantiles. Tomamos $k = 4$, $Q_1 = \hat{X}_{0,25}$, $Q_2 = \hat{X}_{0,5}$ y $Q_3 = \hat{X}_{0,75}$, estimadores de los cuartiles. Entonces, la función a minimizar es

$$g(\mu, \sigma^2) = (Q_1 - x_{0,25})^2 + (Q_2 - x_{0,5})^2 + (Q_3 - x_{0,75})^2$$

Calculemos los cuartiles $x_{0,25}, x_{0,5}$ y $x_{0,75}$ en función de μ y σ.

$$F_X(x|\mu,\sigma^2) = \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{x-\mu}{\sigma}\right)$$

Si hacemos $F_X(x|\mu,\sigma^2) = 0{,}25$ entonces $\arctan\left(\dfrac{x-\mu}{\sigma}\right) = -\dfrac{\pi}{4}$ de donde $x = \mu - \sigma$. Análogamente $x_{0,5} = \mu$ y $x_{0,75} = \mu + \sigma$ (estos valores se calculan facilmente a partir de $x_{0,25}$ usando la paridad de f_X). Luego, la función a minimizar resulta entonces

$$g(\mu,\sigma^2) = (Q_1 - \mu + \sigma)^2 + (Q_2 - \mu)^2 + (Q_3 - \mu - \sigma)^2$$

Derivando esta función respecto de μ y de σ, el gradiente resultante se anula en

$$\hat{\mu} = \frac{Q_1 + Q_2 + Q_3}{3} \qquad \hat{\sigma} = \frac{Q_3 - Q_1}{2}$$

3.5. Estimación de la función de Distribución

Definición 3.24. Distribución Empírica: Sea $X_1,\ldots,X_n$ M.A.S. de $X \backsim F_X$ donde F_X es desconocida, la distribución empírica se define como

$$F_n^*(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{(-\infty,x]}(X_i)$$

Observemos que en cada x nos da la proporción de observaciones menores o iguales que x, y que, para x y n fijos, $F_n^*(x)$ es una v.a. Observemos ademas que si $x_i \neq x_j$ $\forall i \neq j$ los incrementos de F_n^* son n , y de tamaño $1/n$.

Proposición 3.25. $F_n^*(x) \xrightarrow{c.s.} F(x)$ $\forall x \in \mathbb{R}$.

Demostración. Es una consecuencia inmediata de la L.F.G.N a las variables $\mathbb{I}_{(-\infty,x]} \backsim Ber(p)$ con $p = F_X(x)$. □

Teorema 3.26. Teorema fundamental de la Estadística, Glivenko-Cantelli, 1937: Sea $X_1,\ldots,X_n,\ldots$ una M.A.S. de $X \backsim F_X$ entonces

$$\|F_n^* - F_X\|_\infty = \sup_{x\in\mathbb{R}} \left|F_n^*(x) - F_X(x)\right| \xrightarrow{c.s.} 0$$

.

Demostración. Para la demostración vamos a necesitar el siguiente lema:

Lema 3.27. *$Y_n = \sup_{x\in\mathbb{R}} \left|F_n^*(x) - F_X(x)\right|$ es una v.a., es decir, es medible.*

Demostración. Basta demostrar que $\{Y_n \leq \delta\} \in \mathcal{A}$ $\forall \delta > 0$.

$$\begin{aligned}\{Y_n \leq \delta\} = \left\{|F_n^*(x) - F_X(x)| \leq \delta \ \forall x \in \mathbb{R}\right\} &= \left\{F(x) - \delta \leq F_n^*(x) \leq F(x) + \delta \ \forall x \in \mathbb{R}\right\} \\ &= \bigcap_{x\in\mathbb{R}} \left\{F(x) - \delta \leq F_n^*(x) \leq F(x) + \delta\right\}\end{aligned}$$

Basta demostrar que

$$\bigcap_{x\in\mathbb{R}} \left\{F(x) - \delta \leq F_n^*(x) \leq F(x) + \delta\right\} = \bigcap_{x\in\mathbb{Q}} \left\{F(x) - \delta \leq F_n^*(x) \leq F(x) + \delta\right\}$$

Fijemos $x \in \mathbb{R}$, dado $\varepsilon > 0$ $\exists y \in \mathbb{Q}$, $y \geq x$ tal que

1) $F_n^*(y) - F_n^*(x) \leq \varepsilon$ pues F_n^* es continua por derecha.

2) $F(y) - F(x) \leq \varepsilon$ pues F es continua por derecha.

3) $F(y) - \delta \leq F_n^*(y) \leq F(y) + \delta$.

Podemos escribir entonces

$$F(x) - \delta - \varepsilon \overset{x \leq y}{\leq} F(y) - \delta - \varepsilon \overset{3}{\leq} F_n^*(y) - \varepsilon \overset{1}{\leq} F_n^*(x) \overset{x \leq y}{\leq} F_n^*(y) \overset{3}{\leq} F(y) + \delta \overset{2}{\leq} F(x) + \varepsilon + \delta$$

y por lo tanto

$$F(x) - \delta - \varepsilon \leq F_n^*(x) \leq F(x) + \varepsilon + \delta \ \forall \varepsilon > 0$$

entonces

$$F(x) - \delta \leq F_n^*(x) \leq F(x) + \delta.$$

lo cual concluye la demostración del lema. □

Veamos la demostración del teorema, para el caso continuo, dado $x \in \mathbb{R}$ sea $A_x = \{\omega \in \Omega : \lim_n F_n^*(x) = F(x)\}$. Por la proposición anterior sabemos que $P(A_x) = 1$ para todo x. Luego $P\left(\bigcap_{x \in \mathbb{Q}} A_x\right) = 1$.
Sea $A := \bigcap_{x \in \mathbb{Q}} A_x$, basta ver que $A \subset \{\omega \in \Omega : \lim_n \sup_{x \in \mathbb{R}} |F_n^*(x) - F_X(x)| = 0\}$. Sea $\varepsilon > 0$ y $\omega \in A$.
Como $\lim_{x \to +\infty} F(x) = 0$ existe $k_1 \in \mathbb{Q}$ tal que $\forall x < k_1 \ F(x) < \varepsilon$.
Como $\lim_{x \to +\infty} F(x) = 1$ existe $k_2 \in \mathbb{Q}$ tal que $\forall x > k_2 \ 1 - F(x) < \varepsilon$.
Como F es uniformemente continua en $[k_1, k_2]$ existe $k_1 = x_1 < x_2 < \ldots < x_m < x_{m+1} = k_2 \in \mathbb{Q}$ tal que $F(x_{k+1}) - F(x_k) < \varepsilon$ para todo $k = 1, \ldots, m$.
Luego si tomamos $-\infty = x_0 < x_1 < x_2 < \ldots < x_m < x_{m+1} < +\infty = x_{m+2}$ se verifica que $F(x_{k+1}) - F(x_k) < \varepsilon$ para todo $k = 0, \ldots, m+1$. Como $\omega \in A, F_n^*(x_k) \to F_n^*(x_k) \forall k = 0, \ldots, m+2$
Si $x \in \mathbb{R}$ existe $k \in \{0, \ldots, m+2\}$ tal que $x_k \leq x \leq x_{k+1}$ entonces

$$F_n^*(x) \leq F_n^*(x_{k+1}) \overset{1}{\leq} F(x_{k+1}) + \varepsilon \overset{2}{F}(x) + \varepsilon + \varepsilon = F(x) + 2\varepsilon,$$

donde 1 es porque $x_{k+1} \in \mathbb{Q}$ y hemos tomado $\omega \in A$. Esta desigualdad vale para $n > n_0$, que no depende de x. La desigualdad 2 se sigue de que $F(x_{k+1}) \leq F(x_k) + \varepsilon \leq F(x) + \varepsilon$. Razonando de forma análoga llegamos a que, para $n > n_1$, para todo x tenemos que

$$F(x) - 2\varepsilon \leq F_n^*(x) \leq F(x) + 2\varepsilon$$

de donde

$$\lim_{n \to +\infty} \sup |F_n^* - F| = 0.$$

□

3.6. Convergencia casi segura de Percentiles

Teorema 3.28. *Dado $p \in (0,1)$ tal que $\forall \varepsilon > 0 \ F(x_p + \varepsilon) > p$ entonces el percentil empírico $\hat{X_{p,n}} \xrightarrow{c.s.} x_p$.*

Demostración. Observemos que

$$F_n^*(\hat{X_{p,n}}) = \begin{cases} \frac{1}{n} np = p & \text{si } np \in \mathbb{N} \\ \frac{1}{n}([np] + 1) \to p & \text{si } np \notin \mathbb{N} \end{cases}.$$

dado $\varepsilon > 0$ sabemos que $F_n^*(x_p + \varepsilon) \xrightarrow{n} F(x_p + \varepsilon) > p \ c.s.$ y $F_n^*(\hat{X_{p.n}}) \xrightarrow{n} p$ por lo tanto $\forall n \geq n_0$ se cumple que $F_n^*(\hat{X_{p,n}}) < F_n^*(x_p + \varepsilon)$, de donde $\hat{X_{p,n}} < x_p + \varepsilon$. Además $F_n^*(x_p - \varepsilon) \xrightarrow{c.s. \ n} F(x_p - \varepsilon) < p$ y, razonando de forma análoga $x_p - \varepsilon < \hat{X_{p,n}} \ \forall n \geq n_0$. Por lo tanto $\hat{X}_{p,n} \xrightarrow{c.s} x_p$. □

Capítulo 4

Evaluación de Estimadores

Definición 4.1. Dada $X_1, \dots, X_n$ M.A.S. de $F_X(x|\theta)$ y $T = T_n(X_1, \dots, X_n)$ estimador de $g(\theta)$ con g a valores reales, conocida. Decimos que

$$T_n \text{ es insesgado si } E(T_n) = g(\theta) \quad \forall \theta \in (H)$$
$$T_n \text{ es asintóticamente insesgado si } E(T_n) \xrightarrow{n} g(\theta)$$
$$T_n \text{ es debilmente consistente si } T_n \xrightarrow{P} g(\theta)$$
$$T_n \text{ es fuertemente consistente si } T_n \xrightarrow{c.s.} g(\theta)$$

Definición 4.2. Sesgo de un estimador: Se define el sesgo de un estimador T_n como $E(T_n) - g(\theta)$

Definición 4.3. Error cuadrático medio: Se define $E.C.M(T_n) = E\big(T_n - g(\theta)\big)^2$

Es claro que si T_n es un estimador insesgado $E.C.M.(T_n) = V(T_n)$, es natural entonces, tomar estimadores con E.C.M. mínimo.

Definición 4.4. Estimador de mínima varianza: Sea T_n un estimador de $g(\theta)$ tal que $T_n \in L^2$, decimos que es insesgado en $\theta_0 \in (H)$, de varianza mínima si

i) T_n es insesgado en θ_0

ii) Si $T_n' \in L_2$ es insesgado en θ_0 $Var_{\theta_0}(T_n) \leq Var_{\theta_0}(T_n')$.

Observación 4.5. $E_\theta\big(T_n(X_1, \dots, X_n)\big) = \int_{\mathbb{R}^n} T_n(x_1, \dots, x_n) dF_X(x|\theta)$

Observación 4.6. *Si no pedimos que que T_n sea insesgado, cualquier constante es de mínima varianza.*

Teorema 4.7. *T_n es insesgado de minima varianza en θ_0 si y solo si $E_{\theta_0}(f(X_1, \dots, X_n) T_n(X_1, \dots, X_n)) = 0$ para toda $f(x_1, \dots, x_n)$ a valores reales, tal que $E_\theta(f) = 0$.*

Demostración. Para demostrar el teorema será necesario el siguiente lema

Lema 4.8. *T_n es insesgado de mínima varianza en θ_0 si y solo si T_n es insesgado y $Var_{\theta_0}(T_n) \leq Var_{\theta_0}(T_n + \lambda f)$ para todo $\lambda \in \mathbb{R}$, para todo f tal que $E_{\theta_0}(f) = 0$.*

Demostración. Veamos el directo, sea λ y f tal que $E_{\theta_0}(f) = 0$, $T_n' = T_n + \lambda f$ es insesgado pues $E(T_n + \lambda f) = E(T_n) + \lambda E(f) = E(T_n) = g(\theta)$. Como T_n es de mínima varianza $Var_{\theta_0}(T_n) \leq Var_{\theta_0}(T_n')$.
Para demostrar el recíproco consideremos T_n' insesgado, entonces $T_n' = T_n + (T_n' - T_n)$, tomemos $f = T_n' - T_n$ y $\lambda = 1$ entonces $E(f) = 0$, luego, por hipótesis

$$Var_{\theta_0}(T_n) \leq Var_{\theta_0}(T_n + \lambda f) = Var_{\theta_0}(T_n').$$

□

Veamos ahora la demostración del teorema. Por el lema basta ver que $Var_\theta(T_n) \leq Var_{\theta_0}(T_n+\lambda f)$ si y solo si $E_{\theta_0}(fT_n) = 0$.

$$\begin{aligned} Var_{\theta_0}(T_n + \lambda f) = Var(T_n) + \lambda^2 V(f) + 2\lambda cov(T_n, f) \geq Var_{\theta_0}(T_n) \ \forall \lambda \in \mathbb{R} \\ \Leftrightarrow & \lambda^2 Var_{\theta_0}(f) + 2\lambda cov(T_n, f) \geq 0 \ \forall \lambda \in \mathbb{R} \\ \Leftrightarrow & p(\lambda) = \lambda^2 Var_{\theta_0} 2\lambda cov(T_n, f) \geq 0 \Leftrightarrow cov(T_n, f) = 0, \end{aligned}$$

de lo contrario p tendrá 2 raices.

$$cov(T_n, f) = E(T_n f) - E(T_n)E(f) = 0 \Leftrightarrow E(T_n f) = 0.$$

□

Definición 4.9. Estimador insesgado de mínima varianza uniformemente: T_n es estimador I.M.V.U. si es insesgado de varianza mínima $\forall \theta \in (H)$.

Ejemplo 4.10. Sea $X_1, \ldots, X_n$ M.A.S. de $X \backsim \exp(\lambda)$, $\theta = 1/\lambda$. Un estimador de θ es $\overline{X_n}$, veamos que es de mínima varianza. Sabemos que $\overline{X_n} \xrightarrow{c.s.} E(X) = 1/\lambda = \theta$. Si f es tal que $E(f) = 0$ para todo θ.

$$E(f) = \int_{[0,+\infty)^n} f(x_1, \ldots, x_n)\lambda^n \exp\{-\lambda \sum x_i\} dx_1 \ldots dx_n = 0$$

entonces

$$\int_{[0,+\infty)^n} f(x_1, \ldots, x_n) \exp\{-\lambda \sum x_i\} dx_1 \ldots dx_n = 0 \quad \forall \lambda \in \mathbb{R}.$$

Veamos que $E(f\overline{X_n}) = 0$.

$$\begin{aligned} E(f\overline{X_n}) = & \int_{[0,+\infty)^n} f(x_1, \ldots, x_n)\frac{1}{n}(\sum_{i=1}^n x_i)\lambda^n \exp\{-\lambda \sum x_i\} dx_1 \ldots dx_n = 0 \\ \Leftrightarrow & \int_{[0,+\infty)^n} f(x_1, \ldots, x_n)\frac{1}{n}(\sum_{i=1}^n x_i) \exp\{-\lambda \sum x_i\} dx_1 \ldots dx_n = 0 \\ \Leftrightarrow & \int_{[0,+\infty)^n} \frac{\partial}{\partial\lambda}\left(f(x_1, \ldots, x_n) \exp\{-\lambda \sum x_i\}\right) dx_1 \ldots dx_n = 0 \\ \Leftrightarrow & \frac{\partial}{\partial\lambda}\int_{[0,+\infty)^n} \left(f(x_1, \ldots, x_n) \exp\{-\lambda \sum x_i\}\right) dx_1 \ldots dx_n = 0 \end{aligned}$$

Ejemplo 4.11. Sea $X_1, \ldots, X_n$ M.A.S. de $X \backsim Ber(p)$. Consideremos $\overline{X_n}$ estimador de p. Veamos que es de mínima varianza, sea f tal que $E(f) = 0$

$$\begin{aligned} E(f) = & \sum_{(x_1,\ldots,x_n)\in\{0,1\}^n} f(x_1, \ldots, x_n) \prod_{i=1}^n p(x_i|p) \\ = & \sum_{(x_1,\ldots,x_n)\in\{0,1\}^n} f(x_1, \ldots, x_n)p^{\sum x_i}(1-p)^{n-\sum x_i} \\ = & \sum_{k=0}^n \sum_{x_1+\cdots+x_n=k} f(x_1, \ldots, x_n)p^{\sum x_i}(1-p)^{n-\sum x_i} \\ = & \sum_{k=0}^n \left[\sum_{x_1+\cdots+x_n=k} f(x_1, \ldots, x_n)\right] p^k(1-p)^{n-k} = 0 \end{aligned}$$

Tenemos entonces un polinomio de grado a lo sumo n con mas de n raices, y por lo tanto todos sus coeficientes son nulos. Luego si calculamos

$$E(f\overline{X_n}) = \sum_{k=0}^{n} \left(\sum_{x_1+\cdots+x_n=k} f(x_1,\ldots,x_n) \right) \frac{k}{n} p^k (1-p)^{n-k} = 0$$

Teorema 4.12. Desigualdad de Cramer-Rao: Sea $X_1,\ldots,X_n$ M.A.S. de $X \backsim f_X(x|\theta)$. Si T_n es un estimador insesgado de $g(\theta)$. Asumiremos que estamos en las hipótesis de derivación dentro de la integral, es decir que

$$\frac{\partial}{\partial\theta} E(T_n) = \int_{\mathbb{R}^n} \frac{\partial}{\partial\theta} \left(T_n \prod_{i=1}^{n} f_X(x_i|\theta) \right) dx_1 \ldots dx_n$$

y

$$\frac{\partial}{\partial\theta} \int_{\mathbb{R}^n} \prod_{i=1}^{n} f_X(x_i|\theta) dx_i = \int_{\mathbb{R}^n} \frac{\partial}{\partial\theta} \prod_{i=1}^{n} f_X(x_i|\theta) dx_i,$$

entonces

$$Var(T_n) \geq \frac{\big(g'(\theta)\big)}{nE\left(\frac{\frac{\partial}{\partial\theta} f(x|\theta)}{f(x|\theta)}\right)^2}.$$

Además, el igual se da si y solo si existe $\lambda = \lambda(n,\theta)$ tal que

$$T_n(X_1,\ldots,X_n) - g(\theta) \stackrel{c.s.}{=} \lambda \sum_{i=1}^{n} \frac{\frac{\partial}{\partial\theta} f(x_i|\theta)}{f(x_i|\theta)}.$$

Demostración.

$$\begin{aligned}
g'(\theta) =& \frac{\partial}{\partial\theta} E(T_n) = \frac{\partial}{\partial\theta} \int_{\mathbb{R}^n} T_n(x_1,\ldots,x_n) \prod_{i=1}^{n} f_X(x|\theta) dx_1 \ldots dx_n \\
=& \int_{\mathbb{R}^n} \frac{\partial}{\partial\theta} \left[T_n(x_1,\ldots,x_n) \prod_{i=1}^{n} f_X(x_i|\theta) \right] dx_1 \ldots dx_n \\
=& \int_{\mathbb{R}^n} \big(T_n(x_1,\ldots,x_n) - g(\theta)\big) \frac{\partial}{\partial\theta} \prod_{i=1}^{n} f_X(x_i|\theta) dx_1 \ldots dx_n \\
=& \int_{\mathbb{R}^n} \big(T_n(x_1,\ldots,x_n) - g(\theta)\big) \sqrt{\prod_{i=1}^{n} f_X(x_i|\theta)} \frac{\frac{\partial}{\partial\theta} \prod_{i=1}^{n} f_X(x_i|\theta)}{\sqrt{\prod_{i=1}^{n} f_X(x_i|\theta)}} dx_1 \ldots dx_n
\end{aligned}$$

entonces, si aplicamos la desigualdad de Cauchy-Schwartz

$$\begin{aligned}
\big(g'(\theta)\big)^2 \leq & \int_{\mathbb{R}^n} \big(T_n - g(\theta)\big)^2 \prod_{i=1}^{n} f_X(x_i|\theta) \int_{\mathbb{R}^n} \frac{\big(\frac{\partial}{\partial\theta}\prod_{i=1}^{n} f_X(x_i|\theta)\big)^2}{\prod_{i=1}^{n} f_X(x_i|\theta)} \\
=& Var(T_n) \int_{\mathbb{R}^n} \frac{\big(\frac{\partial}{\partial\theta}\prod_{i=1}^{n} f_X(x_i|\theta)\big)^2}{\prod_{i=1}^{n} f_X(x_i|\theta)} \\
=& Var(T_n) E\left(\frac{\big(\frac{\partial}{\partial\theta}\prod_{i=1}^{n} f_X(x_i|\theta)\big)^2}{\prod_{i=1}^{n} f_X(x_i|\theta)}\right)^2 \\
=& Var(T_n) E\left(\frac{\partial}{\partial\theta} \log \prod_{i=1}^{n} f_X(x_i|\theta)\right)^2 \\
=& Var(T_n) E\left(\frac{\partial}{\partial\theta} \log(f_X(x_i|\theta))\right)^2 \\
=& Var(T_n) E\left(\sum_{i=1}^{n} \frac{\frac{\partial}{\partial\theta} f(x_i|\theta)}{f(x_i|\theta)}\right)^2.
\end{aligned}$$

Definamos $g(X_i) = \dfrac{\frac{\partial}{\partial\theta} f(x_i|\theta)}{f(x_i|\theta)}$.

$$E\big(\sum g(X_i)\big)^2 \;=\; E\big(\sum g^2(X_i) + 2\sum_{i\neq j} g(X_i)g(X_j)\big) \;=\; nE\big(g(X_i)^2\big) + 2\sum_{i\neq j} E\big(g(X_i)g(X_j)\big).$$

Basta ver que $E\big(g(X_i)g(X_j)\big) = 0$ para todo $i \neq j$. Como son independientes $E\big(g(X_i)g(X_j)\big) = E\big(g(X_i)\big)E\big(g(X_j)\big)$.

$$\begin{aligned}
E(g(X_i)) &= \int_{\mathbb{R}} \frac{\frac{\partial}{\partial\theta} f(x_i|\theta)}{f(x_i|\theta)} f(x_i|\theta)dx \\
&= \int_{\mathbb{R}} \frac{\partial}{\partial\theta} f(x_i|\theta)dx = 0.
\end{aligned}$$

Para ver cuando se da el igual, observemos que hemos usado la desigualdad de Cauchy-Schwartz, por lo tanto el igual se da si y solo si existe $\lambda = \lambda(n,\theta)$ independiente de $x_1, \ldots, x_n$ tal que

$$(T_n - g(\theta))\sqrt{\prod f_X(x_i|\theta)} = \lambda \frac{\frac{\partial}{\partial\theta}\prod f_X(x_i|\theta)}{\sqrt{\prod f_X(x_i|\theta)}}$$

y esto sucede si y solo si

$$\begin{aligned}
T_n - g(\theta) =& \lambda \frac{\frac{\partial}{\partial\theta}\prod f_X(x_i|\theta)}{\prod f_X(x_i|\theta)} = \lambda \frac{\partial}{\partial\theta}\Big(\log\big(\prod f_X(x_i|\theta)\big)\Big) = \lambda \sum \frac{\partial}{\partial\theta} \log\big(f_X(x_i|\theta)\big) \\
=& \lambda \sum \frac{\frac{\partial}{\partial\theta} f_X(x_i|\theta)}{f_X(x_i|\theta)}
\end{aligned}$$

□

Definición 4.13. Estimador eficiente: Si T_n es un estimador insesgado para $g(\theta)$ y cumple el igual en la desigualdad de Cramer-Rao se dice que es eficiente

Observación 4.14. *Si $\hat{\theta}$ es un estimador de θ, $\hat{\theta}$ es eficiente si y solo si*

i) $\hat{\theta}$ es insesgado

ii) $Var(\hat{\theta}) = \dfrac{1}{nE\left(\frac{\frac{\partial}{\partial\theta}f(x|\theta)}{f(x|\theta)}\right)^2}$

Observación 4.15. *Observemos que* $nE\left(\frac{\frac{\partial}{\partial\theta}f(x|\theta)}{f(x|\theta)}\right)^2$ *es el número de información de Fisher, del logaritmo de la función de verosimilitud de* $X_1,\dots,X_n$ *por lo tanto la observación anterior implica que el E.M.V es asintóticamente eficiente. Si descomponemos el error cuadratico medio* $E(\theta-\hat{\theta_n})^2$ *en sesgo y variabilidad, es decir*

$$E(\theta-\hat{\theta_n})^2 = E(\theta - E(\hat{\theta_n}))^2 + Var(\hat{\theta_n})$$

obtenemos, de la desigualdad de Cramer-Rao, que el E.M.V minimiza (entre los estimadores asintóticamente insesgados) asintóticamente el error cuadratico medio.

Observación 4.16. *Observemos que si* $\hat{\theta}$ *es eficiente, es de mínima varianza (entre el conjunto de estimadores que están el las hipótesis del Teorema de Cramer-Rao). Podría no existir un estimador eficiente, además, existen estimadores de mínima varianza que no cumplen la igualdad.*

Ejemplo 4.17. Sea $X_1,\dots,X_n$ M.A.S. de $X \backsim Ber(p)$, $\overline{X_n}$ es insesgado y además

$$nE\left(\frac{\frac{\partial}{\partial p}p(x|p)}{p(x|p)}\right)^2 = n\left(\frac{1}{p^2}p + \left(\frac{-1}{1-p}\right)^2(1-p)\right) = n\frac{1}{p(1-p)} = \frac{1}{Var(X)},$$

por lo tanto $\hat{p} = \overline{X_n}$ es eficiente. Como $\hat{X_n}$ es eficiente es de mínima varianza varianza ya que X es de recorrido finito.

Definición 4.18. Estimador Suficiente: Dada $X_1,\dots,X_n$ M.A.S. de $X \backsim F(x|\theta)$ y $T_n = T(X_1,\dots,X_n)$ estimador, decimos que T_n es suficiente para θ si y solo si $F_{X_1,\dots,X_n|T_n}$ no depende de θ.

Ejemplo 4.19. Sea $X_1,\dots,X_n$ M.A.S. tal que $X \backsim Ber(p)$ entonces $T = \sum_{i=1}^n X_i$ es un estimador suficiente para estimar p.

Demostración.

$$\begin{aligned}
p_{X_1,\dots,X_n|T=t}(x_1,\dots,x_n) &= P(X_1=x_1,\dots,X_n=x_n|T=t)\\
&= \frac{P(X_1=x_1,\dots,X_n=x_n,T=t)}{P(T=t)} =\\
&\begin{cases} 0 & si \quad t\neq \sum x_i \\ \frac{P(X_1=x_1)\dots P(X_n=x_n)}{P(T=t)} & si \quad t=\sum x_i \end{cases}\\
&= \frac{p^{\sum x_i}(1-p)^{n-\sum x_i}}{C_t^n p^t(1-p)^{n-t}}\\
&= \frac{1}{C_t^n}.
\end{aligned}$$

Que no depende de p, hemos usado que $T \backsim Bin(n,p)$. □

Teorema 4.20. *T es suficiente para* θ *si y solo si* $L(\tilde{x}|\theta) = \prod_{i=1}^n f(x_i|\theta) = g\big(T(\tilde{x}),\theta\big)h(\tilde{x})$

Demostración. (Caso discreto:)

$$L(\tilde{x}|\theta) = \prod_{i=1}^n p_X(x_i|\theta) = \prod_{i=1}^n P(X=x_i|\theta) = P(X_1=x_1,\dots,X_n=x_n|\theta) =$$
$$P_\theta(X_1=x_1,\dots,X_n=x_n|T=t)P(T=t) = h(\tilde{x})g(T(\tilde{x},\theta))$$

Veamos el recíproco, supongamos que $P(T=t)>0$.

$$\begin{aligned} P(X_1=x_1,\dots,X_n=x_n|T=t) &= \frac{P(X_1=x_1,\dots,X_n=x_n,T=t)}{P(T=t)} \\ &= \begin{cases} 0 & si \quad t\neq T(\tilde{x}) \\ \frac{P(X_1=x_1,\dots,X_n=x_n)}{P(T=t)} & si \quad t=T(\tilde{x}) \end{cases} \end{aligned}$$

Para el caso en que $t=T(\tilde{x})$

$$P(X_1=x_1,\dots,X_n=x_n|T=t)=\frac{P(X_1=x_1,\dots,X_n=x_n)}{\sum_{\tilde{y}:T(\tilde{y})=t}P(X_1=y_1,\dots,X_n=y_n)}=\frac{g(T(\tilde{x}))h(\tilde{x})}{\sum_{\tilde{y}:T(\tilde{y})=t}g(T(\tilde{y}),\theta)h(\tilde{y})}.$$

□

Observemos que, dado que estamos en el caso $T(\tilde{x})=t$ y $g(T(\tilde{x}),\theta)=g(t,\theta)=g(T(\tilde{y}),\theta)$. Por lo tanto

$$P(X_1=x_1,\dots,X_n=x_n|T=t)=\frac{h(\tilde{x})}{\sum_{\tilde{y}:T(\tilde{y})=t}h(\tilde{y})}.$$

Que no depende de θ.

Ejemplo 4.21. Sea $X_1,\dots,X_n$ M.A.S. de $X\backsim N(\mu,\sigma^2)$. Estimamos μ y σ^2, consideremos $T(\tilde{x})=(\sum x_i,\sum x_i^2)=(T_1,T_2)$.

$$\begin{aligned} L(\tilde{x}|(\mu,\sigma)) &= \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{\sigma^2}(x_i-\mu)^2\right\} \\ &= \frac{1}{(2\pi)^{n/2}\sigma^n}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{\sum_{i=1}^{n}x_i^2-2\mu\sum_{i=1}^{n}x_i+n\mu^2\right\} \\ &= (2\pi)^{-\frac{n}{2}}\sigma^{-n}\exp\left\{-\frac{1}{2\sigma^2}\big(T_2-2\mu T_1+n\mu^2\big)\right\}. \end{aligned}$$

Por lo tanto si definimos

$$h(\tilde{x})=(2\pi)^{-\frac{n}{2}} \ y \ g\big(T(\tilde{x}),(\mu,\sigma^2)\big)=\sigma^{-n}\exp\left\{-\frac{1}{2\sigma^2}\big(T_2-2\mu T_1+n\mu^2\big)\right\}.$$

De donde T es suficiente.

Observación 4.22. *Siempre existe un estimador suficiente, basta tomar $T(\tilde{x})=\tilde{x}$ y h constante. Esto significa que tener toda la muestra es suficiente.*

Ejemplo 4.23. Si $X_1,\dots,X_n$ es una M.A.S. de $X\backsim U[a,b]$, estimamos (a,b).

$$\begin{aligned} L\big(\tilde{x}|(a,b)\big) &= \begin{cases} \prod\frac{1}{b-a} & si \quad a<x_i<b \\ 0 & si\ no \end{cases} \\ &= \begin{cases} (b-a)^n & si \quad a<x_i<b \\ 0 & si\ no \end{cases} \\ &= \begin{cases} (b-a)^n & si \quad a<x_{1:i};\ x_{n:n}<b \\ 0 & si\ no \end{cases} \\ &= \begin{cases} (b-a)^n & si \quad a<T_1;\ T_2<b \\ 0 & si\ no \end{cases} \end{aligned}$$

Luego $T(\tilde{x})=(T_1,T_2)$ es suficiente.

Observación 4.24. *Si T es fuciente, el E.M.V. es función de un estimador suficiente, ya que en este caso $L(\tilde{x}|\theta) = g(T(\tilde{x}), \theta)h(\tilde{x})$, y, al maximizar en θ como h no varía, podemos maximizar solamente en $g(T(\tilde{x}), \theta)$*

Definición 4.25. **Estimador suficiente minimal:** T estimador suficiente, es minimal si para todo T' estimador suficiente, T es función de T'.

Teorema 4.26. *Sea $X_1, \ldots, X_n$ M.A.S. de $X \backsim F_X(x|\theta)$, si T es un estimador que cumple:*

$$\frac{L(\tilde{x}|\theta)}{L(\tilde{y}|\theta)} \text{ no depende de } \theta \Leftrightarrow T(\tilde{x}) = T(\tilde{y})$$

entonces T es suficiente minimal.

Demostración. Veamos primero que T es suficiente, podemos escribir, tomando $\tilde{y}$ tal que $T(\tilde{y}) = T(\tilde{x})$

$$L(\tilde{x}|\theta) = \frac{L(\tilde{x}|\theta)}{L(\tilde{y}|\theta)} L(\tilde{y}|\theta) = h(\tilde{x})L(\tilde{y}|\theta) = h(\tilde{x})g\big(T(\tilde{x}), \theta)\big).$$

Por lo tanto, por el teorema anterior, T es suficiente ya que hemos podido factorizar la función de verosimilitud.
Veamos que T es minimal, sea T' suficiente, podemos escribir entonces $L(\tilde{x}|\theta) = g'\big(T'(\tilde{x}), \theta\big)h'(\tilde{x})$. Sea $\tilde{x}$ y $\tilde{y}$, $T'(\tilde{x}) = T'(\tilde{y})$ entonces

$$\frac{L(\tilde{x}|\theta)}{L(\tilde{y}|\theta)} = \frac{g'(T'(\tilde{x}), \theta)h'(\tilde{x})}{g'(T'(\tilde{y}), \theta)h'(\tilde{y})} = \frac{h'(\tilde{x})}{h'(\tilde{y})},$$

que no depende de θ, entonces, usando el directo de nuestra hipótesis tenemos que $T(\tilde{x}) = T(\tilde{y})$. Hemos demostrado que cada ves que $T'(\tilde{x}) = T'(\tilde{y})$ entonces $T(\tilde{x}) = T(\tilde{y})$. Veamos que esto implica que $T = f(T')$. Definimos para $z \notin Im(T')$ $f(z)$ cualquier cosa, y para $z \in Im(T')$ entonces $z = T'(x)$ y $f(z) := T(x)$. □

Definición 4.27. **Estadístico Completo:** T se dice completo si toda vez que tenga una función g tal que $E_\theta\big(g(T)\big) = 0$ para todo $\theta \in (H)$ implica que $g(T) = 0$ *c.s.*

Ejemplo 4.28. Sea $X_1, \ldots, X_n$ M.A.S. de $X \backsim U(0, \theta)$ veamos que $\hat{\theta} = x_{n:n}$ es completo.

$$E(g(T)) = \int_{-\infty}^{+\infty} g(t)f_T(t)dt = \int_0^\theta g(t)n\frac{t^{n-1}}{\theta^{n-1}}\frac{1}{\theta}dt = \frac{n}{\theta^n}\int_0^\theta t^{n-1}g(t)dt,$$

luego $E(g(T)) = 0$ si y solo si $\int_0^\theta t^{n-1}g(t)dt = 0$ lo cual implica que $g(t) = 0$, ya que esta integral es derivable c.s., $\theta^{n-1}g(\theta) = 0$ entonces $g(\theta) = 0$ para todo θ.

Definición 4.29. **Función de pérdida:** Sea (H) y $L : (H) \times (H) \longrightarrow \mathbb{R}$ que verifica

i) $L(u, v) = L(v, u)$ para todo $u, v \in (H)$.

ii) $L(u, v) = 0$ si y solo si $u = v$.

iii) L es convexa, es decir,

$$\text{para todo } p, q \in (H) \times (H) \ L(\lambda p + (1-\lambda)q) \leq \lambda L(p) + (1-\lambda)L(q).$$

se denomina función de pérdida.

Observación 4.30. *Si L es C^2 es convexa si y solo si $H_{(x,y)}L$ es semidefinido positivo*

Definición 4.31. **función de riesgo:** Sea $X_1, \ldots, X_n$ M.A.S. de $X \backsim F_X(x|\theta)$ y $\theta \in (H)$ desconocida, dado $T(X_1, \ldots, X_n)$ estimador de θ y L una función de perdida, definimos la función de riesgo

$$R(\theta, T) = E\big(L(\theta, T)\big).$$

Definición 4.32. Estimador de riesgo mínimo, uniformemente entre los insesgados: T es E.R.M.U entre los insesgados si dado T' estimador insesgado se cumple que

$$R(\theta, T) \leq R(\theta, T') \quad \forall\, \theta \in (H).$$

Teorema 4.33. Rao-Blackwell: Si $\sigma(X_1, \dots, X_n)$ es insesgado y $T(X_1, \dots, X_n)$ es suficiente, entonces

$$\eta(X_1, \dots, X_n) = E\big(\sigma(X_1, \dots, X_n)|T(X_1, \dots, X_n)\big),$$

entonces

$$R(\theta, \eta) \leq R(\theta, \sigma).$$

Demostración.

$$\begin{aligned} R(\theta, \eta) = E\big(L(\theta, \eta)\big) = E\big(L(\theta, E(\sigma|T))\big) = E\big(L(E(\theta, \sigma|T))\big) \\ \leq E\big(E(L(\theta, \sigma)|T)\big) = E\big(L(\theta, \sigma)\big). \end{aligned}$$

Donde hemos usado la desigualdad de Jensen. □

Observación 4.34. *En la demostración anterior, la hipótesis de que T es suficiente es necesaria para que η sea un estimador de θ.*

Observación 4.35. *η es insesgado $E(\eta) = E(E(\sigma|T)) = E(\sigma) = \theta$.*

Lema 4.36. *Sea T suficiente, y $\psi\big(T(X_1, \dots, X_n)\big)$ tal que si $f\big(T(X_1, \dots, X_n)\big)$ es una función de T insesgada entonces $\psi\big(T(X_1, \dots, X_n)\big) = f\big(T(X_1, \dots, X_n)\big)$ c.s. entonces $\psi(T)$ es uniformemente de mínimo riesgo entre los insesgados.*

Demostración. Sea σ insesgado, por Rao-Blackwell, como T es suficiente $R(\theta, \eta) \leq R(\theta, \sigma)$, sea $\eta = E(\sigma|T)$ es una función de T y es insesgado entonces por hipótesis $f(T) = \psi(T)$ *c.s.*. Entonces $\eta = \psi(T)$, y $R(\theta, \psi(T)) \leq R(\theta, \sigma)$, donde σ es arbitrario dentro de los insesgados, por lo tanto ψ es uniformemente de mínimo riesgo. □

Lema 4.37. *Si T es completo y $f(T(X_1, \dots, X_n))$ $\psi(T(X_1, \dots, X_n))$ son insesgados entonces entonces*

$$f(T(X_1, \dots, X_n)) = \psi(T(X_1, \dots, X_n)) \text{ c.s.}$$

Demostración. $E(f(T) - \psi(T)) = 0$ para todo $\theta \in (H)$, como T es completo, tomamos $g(T) = f(T) - \psi(T)$ entonces $E(g(T)) = 0$ para todo $\theta \in (H)$, entonces $g = 0$ *c.s.*. □

Teorema 4.38.

1) *Si T es suficiente y completo y σ es insesgado entonces $E(\sigma|T)$ minimiza el riesgo uniformemente entre los insesgados.*

2) *Si T es suficiente, completo e insesgado entonces T minimiza el riesgo uniformemente entre los insesgados.*

Demostración.

1) Sea $\psi(T) = E(\sigma|T)$, entonces ψ es insesgado ya que σ lo es. Si $f(T)$ es insesgado, por el Lema 4.37 $f(T) = \psi(T)$ *c.s.*, entonces, por el Lema 4.36 $\psi(T)$ minimiza el riesgo uniformemente entre los insesgados.

2) Tomamos $\sigma = E(T|T) = T$ y se concluye usando la parte anterior.

□

Ejemplo 4.39. Sea $X_1, \ldots, X_n$ M.A.S. de $X \backsim Ber(p)$. Entonces $\hat{p} = \overline{X_n}$ es uniformemente de mínimo riesgo entre los insesgados. Como ya vimos $\hat{p}$ es insesgado y suficiente como ya vimos, veamos que es completo.

$$\begin{aligned}
0 = E(g(\hat{p})) &= \sum_{x_1,\ldots,x_n \in \{0,1\}} g(\overline{x}) P(X_1 = x_1, \ldots, X_n = x_n) \\
&= \sum_{k=0}^{n} \sum_{x_1+\cdots+x_n=k} g\left(\frac{k}{n}\right) p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n} g\left(\frac{k}{n}\right) p^k (1-p)^{n-k} A_k^n = 0 \ \forall p \\
&= (1-p)^n \sum g\left(\frac{k}{n}\right) \left(\frac{p}{1-p}\right)^k \frac{k!}{(n-k)!}.
\end{aligned}$$

Como $p \in (0,1)$ y tomamos $t = p/(1-p)$. Luego, tenemos un polinomio de grado n, en t con infinitas raices, entonces $g(k/n) = 0$, para todo k, y para todo n, entonces $g(T) = 0$ es 0 $c.s.$

Capítulo 5

Estimación por intervalos de confianza

Definición 5.1. Intervalo de confianza: Dada $X_1, \dots, X_n$ M.A.S. de $X \backsim F_X(x|\theta)$ con θ desconocido, $\theta \in \mathbb{R}$. Un intervalo de confianza al nivel $1-\alpha$ con $\alpha \in (0,1)$ es

$$I = \left[L(X_1, \dots, X_n), U(X_1, \dots, X_n)\right],$$

donde L y U son estimadores y $P(\theta \in I) = 1 - \alpha$.

Ejemplo 5.2. Construcción de intervalos de confianza: Sea $X \backsim N(\mu, \sigma^2)$ con σ^2 conocido, tomamos $\theta = \mu$. Buscamos un intervalo de la forma

$$\left[\overline{X}_n - k, \overline{X}_n + k\right].$$

Debemos hallar k tal que $P(\mu \in I) = 1 - \alpha$, entonces

$$\begin{aligned} 1-\alpha &= P(\overline{X}_n - k \leq \mu \leq \overline{X}_n + k) \\ &= P(\mu - k \leq \overline{X}_n \leq \mu + k) \\ &= \Phi\left(\frac{\mu + k - \mu}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{\mu - k - \mu}{\sigma/\sqrt{n}}\right) \\ &= \Phi\left(\frac{\sqrt{n}k}{\sigma}\right) - \Phi\left(\frac{-\sqrt{n}k}{\sigma}\right) \\ &= 2\Phi\left(\frac{\sqrt{n}k}{\sigma}\right) - 1, \end{aligned}$$

donde en la tercer igualdad hemos usado que $X \backsim N(\mu, \sigma^2/n)$ y en la última la paridad de Φ. Por lo tanto obtuvimos que

$$1 - \alpha/2 = \Phi\left(\frac{\sqrt{n}k}{\sigma}\right) \quad \text{entonces} \frac{\sqrt{n}k}{\sigma} = \phi^{-1}(1-\alpha/2),$$

y por lo tanto tomamos

$$k = \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\alpha/2).$$

Notación: Anotaremos $Z_p = \Phi^{-1}(p)$, con esta notación el intervalo de confianza del ejemplo anterior es

$$\left[\overline{X}_n - \frac{\sigma}{\sqrt{n}}Z_{1-\alpha/2}, \overline{X}_n + \frac{\sigma}{\sqrt{n}}Z_{1-\alpha/2}\right].$$

Ejemplo 5.3. Se $X \backsim N(\mu, \sigma^2)$ con σ^2 desconocido, y $\theta = \mu$, buscamos un intervalo de la forma

$$\left[\overline{X}_n - kS_n, \overline{X}_n + kS_n\right].$$

$$P(\mu \in I) = P\big(|\overline{X}_n - \mu| \leq kS_n\big) = P\left(\frac{\sqrt{n}|\overline{X}_n - \mu|}{S_n} \leq \sqrt{n}k\right).$$

Recordemos que

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{S_n} \backsim T_{n-1},$$

entonces

$$\begin{aligned} P(\mu \in I) =& P\left(-\sqrt{n}k \leq T \leq \sqrt{n}k\right) \\ =& F_T(\sqrt{n}k) - F_T(-\sqrt{n}k) \\ =& 2F_T(\sqrt{n}k) - 1 = 1 - \alpha, \end{aligned}$$

donde hemos usado la simetría de F. Despejando obtenemos

$$k = \frac{F_T^{-1}(1-\alpha/2)}{\sqrt{n}} = \frac{t_{1-\alpha/2}(n-1)}{\sqrt{n}},$$

donde usamos la notación $F_T^{-1}(p) = t_p(n-1)$ siendo $n-1$ son los grados de libertad. Por lo tanto el intervalo de confianza para μ al nivel $1-\alpha$ es

$$I = \left[\overline{X}_n - \frac{S_n}{\sqrt{n}} t_{1-\alpha/2}(n-1), \overline{X}_n + \frac{S_n}{\sqrt{n}} t_{1-\alpha/2}(n-1)\right].$$

Obervemos que como $S_n \xrightarrow{c.s.} \sigma$ entonces

$$T_n = \frac{\sqrt{n}(\overline{X}_n - \mu)}{S_n} \xrightarrow{d} N(0,1), \quad t_p(n-1) \to Z_p.$$

Ejemplo 5.4. Si $X \in L^2$ cualquiera con $E(X) = \mu$ y $Var(X) = \sigma^2$, si n es *grande*, en vista de las observaciones anteriores, un intervalo de confianza aproximado, para μ al nivel $1-\alpha$ es

$$\left[\overline{X}_n - \frac{S_n}{\sqrt{n}} Z_{1-\alpha/2}, \overline{X}_n + \frac{S_n}{\sqrt{n}} Z_{1-\alpha/2}\right].$$

Ejemplo 5.5. Si $X \backsim N(\mu, \sigma^2)$ con μ desconocido, tomamos $\theta = \sigma^2$, busquemos a y b tal que

$$P\big(aS_n^2 \leq \sigma^2 \leq bS_n^2\big) = 1 - \alpha,$$

Recordemos que

$$(n-1)\frac{S_n^2}{\sigma^2} \backsim \chi^2_{n-1},$$

entonces

$$P(\sigma^2/b \leq S_n^2 \leq \sigma^2/a) = P\left(\frac{(n-1)}{b} \leq \frac{(n-1)S_n^2}{\sigma^2} \leq \frac{n-1}{a}\right) = F\left(\frac{n-1}{a}\right) - F\left(\frac{n-1}{b}\right),$$

Basta elegir a tal que $F((n-1)/a) = 1\alpha/2$ y b tal que $F((n-1)/b) = \alpha/2$, de donde

$$a = \frac{n-1}{\chi^2_{1-\alpha/2}(n-1)} \qquad b = \frac{n-1}{\chi^2_{\alpha/2}(n-1)},$$

donde hemos usado la notación $F_{\chi^2}^{-1}(p) = \chi^2_p(n-1)$, para la distribución χ^2 con $(n-1)$ grados de libertad. Luego el intervalo es

$$I = \left[\frac{n-1}{\chi^2_{1-\alpha/2}(n-1)} S_n^2, \frac{n-1}{\chi^2_{\alpha/2}(n-1)} S_n^2\right].$$

Ejemplo 5.6. Sea $X \backsim Ber(p)$ con n *grande* tomemos $\theta = p$, si aproximamos usando el T.C.L. es fácil ver, como $\sigma^2 = p(1-p)$ y $S_n = \sqrt{\overline{X}_n(1-\overline{X_n})}$, nos queda el intervalo

$$I = \left[\overline{X_n} - \frac{\sqrt{\overline{X_n}(1-\overline{X_n})}}{\sqrt{n}} Z_{1-\alpha/2}, \overline{X_n} + \frac{\sqrt{\overline{X_n}(1-\overline{X_n})}}{\sqrt{n}} Z_{1-\alpha/2} \right]$$

Ejemplo 5.7. Aplicacin del T.C.L.: Intervalos de confianza aproximados para $\mu = E(X)$ cuando $\sigma^2 = f(\mu)$. Consideremos $X_1, \ldots, X_n$ M.A.S. de $X \in L^2$ y $g : \mathbb{R} \longrightarrow \mathbb{R}$ clase C^1. Si $g'(\mu) \neq 0$, veamos que $\sqrt{n}(g(\overline{X_n} - g(\mu)) \xrightarrow{d} N(0, (\sigma g'(\mu))^2)$:

$$\sqrt{n}(g(\overline{X_n} - g(\mu)) = \sqrt{n} g'(C_n)(\overline{X_n} - \mu) = g'(C_n)\sqrt{n}(\overline{X_n} - \mu),$$

con $C_n \in [\overline{X_n}, \mu]$ o $C_n \in [\mu, \overline{X_n}]$, sabemos que $g'(C_n) \xrightarrow{c.s.} g'(\mu)$ y $\sqrt{n}(\overline{X_n} - \mu) \xrightarrow{d} N(0, \sigma^2)$, por lo tanto usando el lema de Slutsky se concluye.

Capítulo 6

Pruebas de hipótesis

Supongamos que queremos saber si una moneda está balanceada o no. Se tira 100 veces y obtenemos 54 caras, debemos tomar una decisión entre

$$\begin{aligned} &H_0 : p = 1/2 \quad \text{donde } p = P(\text{cara}) \\ &H_1 : p \neq 1/2. \end{aligned}$$

Definición 6.1. Test de hipótesis: Dada $X_1, \ldots, X_n$ M.A.S. de $F_X(x|\theta)$ con θ desconocido, un test de hipótesis es decidir entre 2 hipótesis;

$$\begin{aligned} &H_0 : \theta \in A \quad \text{hipótesis nula} \\ &H_1 : \theta \in B \quad \text{hipótesis alternativa} \end{aligned}$$

donde suponemos que $A, B \subset (H)$ y $A \cap B = \emptyset$.

Definición 6.2. Región Crítica: La región crítica, que anotaremos como $RC \subset \mathbb{R}^n$ con n el tamaño de la muestra, es la zona de rechazo de H_0.

Definición 6.3. Regla de decisión: Si $(x_1, \ldots, x_n) \in RC$ entonces rechazo H_0, en caso contrario si $(x_1, \ldots, x_n) \notin RC$ no rechazo H_0 (acepto H_0).

Ejemplo 6.4. En nuestro ejemplo de la moneda es natural tomar

$$RC = \big\{(x_1, \ldots, x_n) \in \{0,1\}^n \subset \mathbb{R}^n : |\overline{X_n} - 1/2| \geq k\big\}$$

Definición 6.5. Errores de tipo 1 y 2:

* error tipo 1: rechazar H_0 siendo cierta.

* error tipo 2: aceptar H_0 siendo falsa, H_1 es cierta.

Definición 6.6. Significación de una prueba:

$$\alpha = \sup_{\theta \in A} P_\theta\big((X_1, \ldots, X_n) \in RC\big) = P(\text{ error tipo 1 }).$$

Definición 6.7. Probabilidad del error tipo 2: definimos, para $\theta \in B$

$$\beta(\theta) = P_{\theta \in B}\big((X_1, \ldots, X_n) \notin RC\big) = P(\text{ error tipo 2 }).$$

Definición 6.8. Potencia de la prueba: se define como

$$\pi(\theta) = P\big((X_1, \ldots, X_n) \in RC\big) \quad \forall \theta$$

Observación 6.9. *$\pi(\theta) = 1-\beta(\theta)$ si $\theta \in B$ y $\pi(\theta) \leq \alpha$ si $\theta \in A$. En particular si A es θ_0 $\pi(\theta_0) = \alpha$.*

Ejemplo 6.10. Para el caso de la moneda, si tomamos $\alpha = 0{,}05$ es decir el 5 %, entonces $\sigma^2 = 1/4$.

$$\begin{aligned} \alpha &= P_{1/2}\big((X_1,\dots,X_n) \in [1/2-k, 1/2+k]^c\big) \\ &= P_{1/2}\big(\overline{X_n} \in [1/2-k, 1/2+k]^c\big) \\ &= P\big(20(\overline{X_n} - 1/2) \in [-20k, 20k]^c\big) \\ &= 1 - \Phi(20k) + \Phi(-20k) = 2 - 2\Phi(20k) \end{aligned}$$

donde hemos usado la aproximación de $\frac{\sqrt{100}}{\sigma}(\overline{X_n} - 1/2)$ por una $N(0,1)$. Obtenemos entonces $20k = Z_{0{,}975}$ de donde $k = 0{,}098$. Tenemos entonces la región crítica

$$RC = \big\{(x_1,\dots,x_n) \in \mathbb{R}^n : |\overline{x_n} - 1/2| \geq 0{,}098\big\}.$$

Como $|0,54 - 1/2|$ no es mayor o igual que 0,98 no rechazo H_0 al nivel 5 %.

Observación 6.11. *La decisión depende fuertemente del nivel al que trabajo. Concretamente si elegimos $\alpha = 0$, es decir, la probabilidad de rechazar H_0 siendo cierto es 0, siempre acepeto H_0.*

Calculemos $\beta(p)$ con $p \in H_1 = \{1/2\}^c$ con

$$\begin{aligned} \beta(p) &= P_p(RC^c) \\ &= P_p(|\overline{X_n} - 1/2| < 0{,}098) \\ &= P(0{,}402 < \overline{X_n} < 0{,}598) \\ &\cong \Phi\left(\frac{0{,}598 - p}{\sqrt{\frac{p(1-p)}{100}}}\right) - \Phi\left(\frac{0{,}402 - p}{\sqrt{\frac{p(1-p)}{100}}}\right) \end{aligned}$$

Donde hemos usado que $\overline{X_n} \backsim N\left(p, \frac{p(1-p)}{100}\right)$.

Observación 6.12. *Si construimos una RC con un nivel dado α entonces puedo controlar el error de tipo 1, y no el error de tipo 2, podria decirse entonces que el error de tipo 1 es mas grave.*

Observación 6.13. *En general, uno define la región crítica a partir de un estimador insesgado $RC = \{|\hat{\theta} - \theta_0| \geq k\}$.*

Observación 6.14. *Al permitir variar el tamaño de la muestra uno puede fijar los errores α y β y hallar un n que verifique las igualdades.*

Observación 6.15. *Como el error de tipo 1 es más grave, al rechazar H_0 uno debe estar seguro (tener evidencia) de que H_0 es falso. No rechazar H_0 implica que no hay suficiente evidencia emprica para decir que H_0 es falso. No es que se acepte H_1.*

6.1. Región crítica óptima, Teorema de Neyman-Pearson.

Teorema 6.16. Neyman-Pearson: Sea $X_1,\dots,X_n$ M.A.S. de $X \backsim F_X(x|\theta)$ absolutamente continua, y el test

$$\begin{aligned} H_0: &\quad \theta = \theta_0 \\ H_1: &\quad \theta = \theta_1 \end{aligned}$$

Sea $S_k = \left\{\prod_{i=1}^n \frac{f(x_i,\theta_1)}{f(x_i|\theta_0)} \geq k\right\}$, si k es tal que

$$P_{H_0}(S_k) = P_{H_0}\big((X_1,\dots,X_n) \in S_k\big) = \alpha,$$

entonces S_k es entre todas las RC de nivel α la que tiene menor β (máxima potencia).

Demostración. Sea $\beta = P_{\theta_1}(S_k^c)$ y $\beta_0 = P_{\theta_1}(S_0^c)$ donde S_0 es otra RC de nivel α, entonces

$$\begin{aligned}
\beta - \beta_0 &= P_{\theta_1}(S_k^c) - P_{\theta_1}(S_0^c) \\
&= \int_{S_k^c} \prod_{i=1}^n f(x_i|\theta_1)dx_1 \dots dx_n - \int_{S_0^c} \prod_{i=1}^n f(x_i|\theta_1)dx_1 \dots dx_n \\
&= \int_{S_k^c \setminus S_0^c} \prod_{i=1}^n f(x_i|\theta_1)dx_1 \dots dx_n - \int_{S_0^c \cap S_k} \prod_{i=1}^n f(x_i|\theta_1)dx_1 \dots dx_n \\
&\leq k \left[\int_{S_k^c \setminus S_0^c} \prod_{i=1}^n f(x_i|\theta_0)dx_1 \dots dx_n - \int_{S_0^c \cap S_k} \prod_{i=1}^n f(x_i|\theta_0)dx_1 \dots dx_n \right] \\
&= k \left[\int_{S_k^c} \prod_{i=1}^n f(x_i|\theta_0)dx_1 \dots dx_n - \int_{S_0^c} \prod_{i=1}^n f(x_i|\theta_0)dx_1 \dots dx_n \right] \\
&= k\big[P_{\theta_0}(S_k^c) - P_{\theta_0}(S_0^c)\big] = k[1 - \alpha - (1 - \alpha)] = 0.
\end{aligned}$$

Luego $\beta \leq \beta_0$, como β_0 es arbitrario β es mínimo. □

Ejemplo 6.17. Hallar la forma de la RC óptima para el caso $X \backsim N(\mu, \sigma^2)$ y el problema

$$\begin{aligned}
H_0: &\quad \mu = \mu_0 \\
H_1: &\quad \mu = \mu_1
\end{aligned}$$

con $\mu_1 > \mu_0$

Por el teorema de Neyman-Pearson planteamos

$$R_{NP} = \left\{ \prod_{i=1}^n \frac{e^{-\frac{(X_i-\mu_1)^2}{2\sigma^2}}}{e^{-\frac{(X_i-\mu_0)^2}{2\sigma^2}}} \geq k \right\}$$

donde k es tal que $P_{H_0}(R_{NP}) = \alpha$,

$$\begin{aligned}
R_{NP} &= \left\{ e^{-\sum_{i=1}^n \frac{(X_i-\mu_1)^2-(X_i-\mu_0)^2}{2\sigma^2}} \geq k \right\} \\
&= \left\{ -\sum_{i=1}^n \frac{(X_i - \mu_1)^2 - (X_i - \mu_0)^2}{2\sigma^2} \geq \log(k) \right\} \\
&= \left\{ -\frac{\sum_{i=1}^n X_i^2 - 2\mu_1 \sum_{i=1}^n X_i + n\mu_1^2 - \sum_{i=1}^n X_i^2 + 2\mu_0 \sum_{i=1}^n X_i - n\mu_0^2}{2\sigma^2} \geq \log(k) \right\} \\
&= \left\{ \frac{\mu_1 - \mu_0}{\sigma^2} \sum_{i=1} X_i + \frac{n(\mu_0^2 - \mu_1^2)}{2\sigma^2} \geq \log(k) \right\} \\
&= \left\{ \frac{2n(\mu_1 - \mu_0)\overline{X_n} - n(\mu_1 - \mu_0)(\mu_1 + \mu_0)}{2\sigma^2} \geq \log(k) \right\} \\
&= \left\{ \overline{X_n} \geq \frac{2\sigma^2 \log(k) + n(\mu_1 - \mu_0)(\mu_1 + \mu_0)}{2n(\mu_1 - \mu_0)} \right\} = \left\{ \overline{X_n} \geq k' \right\}
\end{aligned}$$

donde

$$k' = \frac{\sigma^2 \log(k)}{n(\mu_1 - \mu_0)} + \frac{(\mu_1 + \mu_0)}{2},$$

Observemos que bajo H_0 $\overline{X_n} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right)$ entonces

$$\alpha = P_{H_0}\left(\frac{\sqrt{n}(\overline{X_n} - \mu_0)}{\sigma} \geq \frac{\sqrt{n}(k' - \mu_0)}{\sigma} \right) = P\left(N(0,1) \geq \frac{\sqrt{n}(k' - \mu_0)}{\sigma} \right)$$

por lo tanto

$$z_\alpha = \frac{\sqrt{n}(k' - \mu_0)}{\sigma}$$

y

$$k' = \mu_0 + \frac{\sigma z_\alpha}{\sqrt{n}}$$

finalmente la región crítica es

$$\left\{ \overline{X_n} \geq \mu_0 + \frac{\sigma z_\alpha}{\sqrt{n}} \right\}.$$

Vamos a calcular para esta prueba la probabilidad del error de tipo II, esto es:

$$\beta = P_{H_1}\left(\overline{X_n} \leq \mu_0 + \frac{\sigma z_\alpha}{\sqrt{n}} \right) = P_{H_1}\left(\frac{\sqrt{n}(\overline{X_n} - \mu_1)}{\sigma} \leq \frac{\sqrt{n}(\mu_0 - \mu_1)}{\sigma} + z_\alpha \right) = \Phi\left\{ z_\alpha - \frac{\sqrt{n}(\mu_1 - \mu_0)}{\sigma} \right\}$$

Por ejemplo si $\sigma = 1$, $\alpha = 5\,\%$, $z_\alpha = 1,645$, $\mu_0 = 0$, $\mu_1 = 0,5$, tenemos la siguiente variación de β según n

n	β
4	0,740
9	0,558
16	0,361
25	0,196
36	0,088

Es decir que por ejemplo para $n = 9$ ningún test de nivel $5\,\%$ para éste test tiene potencia mayor que $44,2\,\%$, esto quiere decir que es muy probable que aceptemos H_0 de forma erronea con estas muestras pequeas.

Observación 6.18. *Análogamente se demuestra que si $\mu_1 < \mu_0$ la prueba*

$$\begin{aligned} H_0 : &\quad \mu = \mu_0 \\ H_1 : &\quad \mu = \mu_1 \end{aligned}$$

tiene como region crítica

$$R_\alpha = \left\{ \overline{X_n} < \mu_0 - \frac{\sigma z_\alpha}{\sqrt{n}} \right\}$$

Ejemplo 6.19. Consideremos $X_1, \ldots, X_n$ i.i.d. con distribución de Poisson de parámetro λ, que anotaremos $\mathcal{P}(\lambda)$ y

$$\begin{aligned} H_0 : &\quad \lambda = 100 \\ H_1 : &\quad \lambda = 120 \end{aligned}$$

entonces $R_{NP} = \{\overline{X_n} > c_{\alpha,n}\}$ donde $c_{\alpha,n}$ es tal que $P\big(\mathcal{P}(n100) > nc_{\alpha,n}\big) = \alpha$. Observemos que usando el T.C.L, sabemos que si $Z \sim \mathcal{P}(n100)$ entonces

$$\frac{Z - n100}{\sqrt{n100}} \approx N(0,1),$$

de donde

$$\alpha = P\left(\frac{Z - n100}{\sqrt{n100}} > \frac{n(c_{\alpha,n} - 100)}{\sqrt{n100}} \right) \approx P\left(N(0,1) > \frac{\sqrt{n}}{10}(c_{\alpha,n} - 100) \right)$$

despejando

$$c_{\alpha,n} \approx \frac{10 z_\alpha}{\sqrt{n}} + 100.$$

Ejemplo 6.20. Consideremos $X_1, \ldots, X_n$ i.i.d. $\sim Ber(p)$ y deseamos testear

$$\begin{aligned} H_0 : &\quad p = p_0 \\ H_1 : &\quad p = p_1 \end{aligned}$$

con $p_1 > p_0$ dados. Tenemos $R_{NP} = \{\overline{X_n} > c_{\alpha,n}\}$ donde

- $c_{\alpha,n}$ lo deducimos de la tabla de la $Bin(n, p_0)$ si n es moderado y $P(Bin(n, p_0) > nc_{\alpha,n}) = \alpha$.
- $c_{\alpha,n}$ lo deducimos de la tabla de $\mathcal{P}(np_0)$ si n es grande y p_0 muy pequeo, donde $P(\mathcal{P}(np_0) > nc_{\alpha,n}) = \alpha$.
- $c_{\alpha,n} = p_0 + \frac{\sqrt{p_0(1-p_0)}}{\sqrt{n}} z_\alpha$ para el caso en que n es grande, y p_0 no muy pequeo, aqui usamos el T.C.L.

Las regiones críticas para $p_1 < p_0$ son análogas.

Corolario 6.21. Corolario de Neyman-Pearson, en las hipótesis del teorema, $\alpha + \beta \leq 1$

Demostración.

$$\beta = P_{H_1}(S^c) = \int_{S^c} \prod_{i=1}^{n} f(x_i|\theta_1) dx_1 \ldots dx_n \leq k \int_{S^c} \prod_{i=1}^{n} f(x_i|\theta_0) dx_1 \ldots dx_n =$$
$$k\big(1 - \int_S \prod_{i=1}^{n} f(x_i|\theta_1) dx_1 \ldots dx_n\big) = k(1-\alpha),$$

si $k \leq 1$ entonces $\beta \leq 1 - \alpha$ de donde $\alpha + \beta \leq 1$,
si $k \geq 1$

$$1 - \beta = P_{H_1}(S) = \int_S \prod_{i=1}^{n} f(x_i|\theta_1) dx_1 \ldots dx_n \geq k \int_S \prod_{i=1}^{n} f(x_i|\theta_0) dx_1 \ldots dx_n = k\alpha,$$

como $k \geq 1$ entonces $1 - \beta \geq k\alpha \geq \alpha$ de donde $\alpha + \beta \leq 1$. ☐

Teorema 6.22. *Consideremos $X_1, \ldots, X_n$ una M.A.S. de $X \backsim F_X(x|\theta)$ absolutamente continua, y la prueba*

$$\begin{aligned} H_0 : &\quad \mu = \mu_0 \\ H_1 : &\quad \mu = \mu_1 \end{aligned}$$

y $k = k_n$ es tal que $P_{H_0}\left(\prod_{i=1}^{n} \frac{f(X_i|\theta_1)}{f(X_i|\theta_0)} \geq k_n \right) = \alpha$ entonces $\beta_n \to 0$.

Demostración. $\log\left(\prod_{i=1}^{n} \frac{f(X_i|\theta_1)}{f(X_i|\theta_0)} \right) = \sum_{i=1}^{n} \log\left(\frac{f(X_i|\theta_1)}{f(Xx_i|\theta_0)} \right)$ y por la L.F.G.N.

$$\begin{aligned} \frac{1}{n} \log\left(\prod_{i=1}^{n} \frac{f(X_i|\theta_1)}{f(X_i|\theta_0)} \right) &\xrightarrow{c.s.} E\left(\log\left(\frac{f(X|\theta_1)}{f(X|\theta_0)} \right) \right) \\ &< \log\left(E\left(\frac{f(X|\theta_1)}{f(X|\theta_0)} \right) \right) \\ &= \log\left(\int \frac{f(x|\theta_1)}{f(x|\theta_0)} f(x|\theta_0) dx \right) = \log(1) = 0. \end{aligned}$$

Donde en la primera desigualdad usamos Jensen (estricta porque log es estrictamente cóncava), y en la siguiente igualdad hicimos el supuesto de H_0 cierto, es decir $\theta = \theta_0$. Tenemos entonces que

$$\sum_{i=1}^{n} \log\left(\frac{f(X_i|\theta_1)}{f(X_i|\theta_0)}\right) \xrightarrow{c.s.} -\infty.$$

Luego, para todo $\varepsilon > 0$ tomando $\alpha = \varepsilon$, y para todo $m \in \mathbb{N}$ existe n_0 tal que $\forall n \geq n_0$

$$P\left(\sum_{i=1}^{n} \log\left(\frac{f(X_i|\theta_1)}{f(X_i|\theta_0)}\right) < -m\right) \geq 1 - \varepsilon = 1 - \alpha.$$

Llamemos

$$S_n = \left\{\prod_{i=1}^{n} \frac{f(X_i|\theta_1)}{f(X_i|\theta)} \geq k_n\right\}$$

y

$$A_{n,m} = \left\{\sum_{i=1}^{n} \log\left(\frac{f(X_i|\theta)}{f(X_i|\theta)}\right) < -m\right\}.$$

Si tomamos $\omega \in A_{n,m} \cap S_n$ entonces

$$\log(K_n) \leq \sum_{i=1}^{n} \log\left(\frac{f(X_i(\omega)|\theta_1)}{f(X_i(\omega)|\theta_0)}\right) < -m,$$

luego, tenemos que $\forall\ m \in \mathbb{N},\ \exists n_0\quad$ tal que $\forall n \geq n_0\ \log(k_n) < -m$ de donde $k_n \to 0$.
Observemos que $S_n \cap A_{n,m} \neq \emptyset$ ya que $P(S_n) + P(A_{n,m}) > \alpha + 1 - \alpha > 1$, luego, se intersectan.
Como $\beta_n = P_{H_1}(S_n^c) = \int_{S_n^c} \prod f(x_i|\theta)dx \leq k_n \int_n^c \prod f(x_i|\theta)dx \to 0$. □

Corolario 6.23. *Consideremos el caso particular*

$$\begin{aligned} H_0 : &\quad \theta = \theta_0 \\ H_1 : &\quad \theta = \theta_1 \end{aligned}$$

Sea $S_n = \left\{\prod_{i=1}^{n} \frac{f(x_i|\theta_1)}{f(x_i|\theta_0)} \geq 1\right\}$ *entonces* $\alpha_n + \beta_n \to 0$

Demostración. Si H_0 es cierto entonces $\sum_{i=1}^{n} \log\left(\frac{f(X_i|\theta_1)}{f(X_i|\theta_0)}\right) \xrightarrow{P} -\infty$, de donde $\alpha_n = P_{H_0}(S_n) = P_{H_0}\left(\sum_{i=1}^{n} \log\left(\frac{f(X_i|\theta_1)}{f(X_i|\theta_0)}\right) \geq 0\right) \xrightarrow{n} 0$.
Si H_1 es cierto entonces $\sum_{i=1}^{n} \log\left(\frac{f(X_i|\theta_0)}{f(X_i|\theta_1)}\right) \xrightarrow{P} -\infty$.

$$\beta_n = P_{H_1}\left(\sum_{i=1}^{n} \log\left(\frac{f(X_i|\theta_1)}{f(X_i|\theta_0)}\right) \leq 0\right) \longrightarrow 0.$$

□

Observemos que éste resultado nos dice que si pudiésemos disponer de muestras arbitrariamente grandes, tanto la probabilidad de error de tipo I, como la de tipo II, podrían hacerse arbitrariamente pequeas. No obstante se cumple el siguiente teorema:

Teorema 6.24. *Si $X_1, \ldots, X_n$ son i.i.d. con densidad f_θ , consideramos la prueba*

$$\begin{aligned} H_0 : &\quad \theta = \theta_0 \\ H_1 : &\quad \theta = \theta_1 \end{aligned}$$

y suponemos además que $\forall\ A_k$ sucesión de sucesos tal que $P_{H_0}(A_k) \to 1$ el límite inferior en k de $P_{H_1}(A_k)$ es positivo, entonces existe $\delta_n > 0$ tal que para cualquier región crítica RC se tiene que si $P(\text{error tipo I}) = P_{H_0}(RC) \quad P(\text{error tipo II}) = P_(H_1)(RC^c)$ entonces

$$P(\text{error tipo I}) + P(\text{error tipo II}) \geq \delta_n$$

La hipotesis sobre los A_k es técnica y se cumple en la mayoría de los casos, además puede verse que que en los casos en que las densidades correspondientes a ambas hipótesis tienen el mismo soporte, esa hipótesis es válida. Observemos que lo que nos esta dando este teorema es una cota para la velocidad de convergencia de la suma de las probabilidades.

Ejemplo 6.25. Supongamos que $X_1, \ldots, X_n$ son i.i.d. $\sim N(\mu, \sigma^2)$, con σ conocida y queremos testear:

$$\begin{aligned} H_0: &\quad \mu \leq \mu_0 \\ H_1: &\quad \mu > \mu_0 \end{aligned}$$

dada, en tal caso tenemos que si RC es la región crítica,

$$\alpha = \sup_{\mu \leq \mu_0} P\big((X_1, \ldots, X_n) \in RC\big)$$

Dado α propondremos la región crítica (para cada alternativa $\mu > \mu_0$ fija, es la mejor segun Neyman Pearson)

$$RC = \left\{ \overline{X_n} > \mu_0 + \frac{z_\alpha \sigma}{\sqrt{n}} \right\},$$

y verifiquemos que su nivel es α, en efecto

$$\begin{aligned} \sup_{\mu \leq \mu_0} P\left(\overline{X_n} > \mu_0 + \frac{z_\alpha \sigma}{\sqrt{n}} \right) = \sup_{\mu \leq \mu_0} P\left(\sqrt{n}\frac{(\overline{X_n} - \mu)}{\sigma} > \frac{\sqrt{n}(\mu_0 - \mu)}{\sigma} + z_\alpha \right) &= \sup_{\mu \leq \mu_0} 1 - \Phi\left(\frac{\sqrt{n}(\mu_0 - \mu)}{\sigma} + z_\alpha \right) \\ &= 1 - \Phi(z_\alpha) = \alpha \end{aligned}$$

Observemos que en este caso el error de tipo II puede ser muy apreciable. Finalmente, puede demostrarse que, en este tipo de ejemplos (test sobre la media de poblaciones guassianas) si σ es desconocida, todos los test antes vistos funcionan de igual modo si se reemplaza σ por S_n y z_α por $t_\alpha(n-1)$. Nótese que si $X_1, \ldots, X_n$ son i.i.d. $\sim N(\mu, \sigma^2)$ sea σ conocida o no, los test

$$\begin{array}{lll} H_0: \ \mu \leq \mu_0 & \qquad H_0: \ \mu = \mu_0 & \qquad H_0: \ \mu = \mu_0 \\ H_1: \ \mu > \mu_0 & \qquad H_1: \ \mu > \mu_0 & \qquad H_1: \ \mu = \mu_1 \end{array}$$

Tienen la misma región crítica ya que el segundo caso contiene la peor comparación del primero , y la región crítica de Neyman Pearson del tercero, no depende del valor de μ_1 como se observó.

La prueba

$$\begin{aligned} H_0: &\quad \mu = \mu_0 \\ H_1: &\quad \mu \neq \mu_0 \end{aligned}$$

tiene región crítica

$$RC = \left\{ \left| \frac{\sqrt{n}(\overline{X_n} - \mu_0)}{\sigma} \right| \geq z_{\alpha/2} \right\}$$

Vamos a presentar, sin demostración las regiones críticas para el caso en que tanto σ como μ son desconocidos, para un tratamiento mas detallado de estos temas pueden verse, [3] o [2].

Para

$$\begin{array}{lll} H_0: \ \sigma = \sigma_0 & \qquad H_0: \ \sigma = \sigma_0 & \qquad H_0: \ \sigma \leq \sigma_0 \\ H_1: \ \sigma = \sigma_1 & \qquad H_1: \ \sigma > \sigma_0 & \qquad H_1: \ \sigma > \sigma_0 \end{array}$$

con $\sigma_1 > \sigma_0$, tenemos

$$RC_\alpha = \left\{ (n-1)\frac{S_n^2}{\sigma_0^2} \geq \chi^2_\alpha(n-1) \right\},$$

y para

$$\begin{array}{llll} H_0: & \sigma & = \sigma_0 \\ H_1: & \sigma & = \sigma_1 \end{array} \qquad \begin{array}{llll} H_0: & \sigma & = \sigma_0 \\ H_1: & \sigma & < \sigma_0 \end{array} \qquad \begin{array}{llll} H_0: & \sigma & \leq \sigma_0 \\ H_1: & \sigma & > \sigma_0 \end{array}$$

con $\sigma_1 < \sigma_0$ es

$$RC_\alpha = \left\{ (n-1)\frac{S_n^2}{\sigma_0^2} \geq \chi^2_{1-\alpha}(n-1) \right\},$$

finalmente para

$$\begin{array}{llll} H_0: & \sigma & = \sigma_0 \\ H_1: & \sigma & \neq \sigma_1 \end{array}$$

tenemos

$$RC_\alpha = \left\{ (n-1)\frac{S_n^2}{\sigma_0^2} \notin \left(\chi^2_{1-\alpha/2}(n-1), \chi^2_{\alpha/2}(n-1) \right) \right\}.$$

6.2. Familias con cociente de verosimilitud monótono

Definición 6.26. Familia con C.V.M.: Una familia de densidades $f(\cdot|\theta)$ con $\theta \in (H) \subset \mathbb{R}$ tiene C.V.M. si

$$\frac{L(\tilde{x}|\theta)}{L(\tilde{x}|\theta')} = \frac{\prod_{i=1}^n f(x_i|\theta)}{\prod_{i=1}^n f(x_i|\theta')} = g\big(T(\tilde{x})\big),$$

donde $g : \mathbb{R} \longrightarrow \mathbb{R}$ es estrictamente creciente, $\theta > \theta'$, y $T = T_n$ es un estimador. Observemos que g depende de n de θ y de θ'

Ejemplo 6.27. $f(\cdot|\theta)$ es una familia exponencial (para $\theta \in (H) \subset \mathbb{R}$) si

$$\prod_{i=1}^n f(x_i|\theta) \;=\; C_n e^{Q(\theta)t(\tilde{x})} h(\tilde{x}) \quad \text{con } C_n(\theta) > 0,$$

si Q es estrictamente creciente la familia tiene C.V.M.:

$$\frac{L(\tilde{x}|\theta)}{L(\tilde{x}|\theta')} = \frac{C_n(\theta)}{C_n(\theta')}\frac{e^{Q(\theta)t(\tilde{x})}}{e^{Q(\theta')t(\tilde{x})}}\frac{h(\tilde{x})}{h(\tilde{x})} = \frac{C_n(\theta)}{C_n(\theta')}e^{t(\tilde{x})\left(Q(\theta)-Q(\theta')\right)} = g(t(\tilde{x}))$$

con $g(s) = \frac{C_n(\theta)}{C_n(\theta')}e^{s\left(Q(\theta)-Q(\theta')\right)}$, luego, g es una función creciente de s.

Teorema 6.28. *Sea $X_1, \ldots, X_n$ M.A.S. de X con densidad $f(\cdot|\theta)$ perteneciente a una familia con C.V.M, sea $T(\tilde{x})$ absolutamente continua y $\theta \in (H) \subset \mathbb{R}$, consideremos*

$$\begin{aligned} H_0 &: \theta \leq \theta_0 \\ H_1 &: \theta > \theta_0 \end{aligned}$$

Si $R = \{\tilde{x} \in \mathbb{R}^n : T(\tilde{x}) \geq k\}$ donde k es tal que R sea R.C. de nivel α, entonces R es R.C. uniformemente de máxima potencia.

Demostración. En el conjutno $\{\theta : \theta \leq \theta_0\} \subset (H)$ defino $\alpha_k(\theta) = \alpha(\theta) = P_\theta(R)$. Probaremos que α es creciente y por lo tanto $\sup_{\theta \in H_0} \alpha(\theta) = \alpha(\theta_0)$, de donde el k de la hipótesis es tal que $P_{\theta_0}(T(\tilde{x}) \geq k) = \alpha$. Consideremos la prueba

$$\begin{aligned} H_0 &: \theta = \theta' \\ H_1 &: \theta = \theta'' \end{aligned}$$

Con $\theta'' > \theta'$. Por lo tanto aplicando el teorema de Neyman Pearson a esta prueba obtenemos la región crítica óptima
$$\left\{\prod_{i=1}^{n}\frac{f(x_i|\theta'')}{f(x_i|\theta')} \geq k'\right\} = \{T(\tilde{x}) \geq g^{-1}(k')\},$$
en esta igualdad hemos usado que g es creciente, llamemos $k'' = g^{-1}(k')$. Para esta prueba $\alpha+\beta \leq 1$, $\alpha = P_{\theta'}(\{T(\tilde{x}) \geq)\}) = \alpha(\theta')$ y $\beta = P_{\theta''}(\{T(\tilde{x}) \geq k''\}^c) = 1 - P_{\theta''}(\{T(\tilde{x}) \geq k''\}) = 1 - \alpha(\theta'')$. Entonces $\alpha(\theta')+1-\alpha(\theta'') \leq 1$ y por lo tanto $\alpha(\theta') \leq \alpha(\theta'')$. Como θ' y θ'' son arbitrarios se deduce que α creciente.
Veamos ahora que R es óptima, es decir, uniformemente de máxima potencia. Supongamos por absurdo, que existe otra S RC de nivel α tal que existe $\hat{\theta} > \theta_0$ y $\beta_S(\hat{\theta}) < \beta_R(\hat{\theta})$, sabemos que $\sup_{\theta\leq\theta_0} \alpha_S(\theta) = \alpha$ ya que hemos supuesto que S es RC de nivel α, por lo tanto $\alpha_S(\theta_0) \leq \alpha$. Consideremos la prueba
$$\begin{aligned} H_0 &: \ \theta = \theta_0 \\ H_1 &: \ \theta > \theta_0 \end{aligned} \tag{6.1}$$
Sea $S' = \{T(\tilde{x}) \geq k\}$ con k' tal que $\alpha_{S'}(\theta_0) = \alpha_S(\theta_0)$, (tal k' existe porque hemos supuesto que T es absolutamente continua). Como hemos supuesto que la familia tiene C.V.M. sabemos por el teorema de Neyman Person que S' es uniformemente de máxima potencia para la prueba 6.1. Entonces $\beta_{S'}(\theta) \leq \beta_S(\theta)\ \forall\theta \geq \theta_0$. En particular $\beta_{S'}(\hat{\theta}) \leq \beta_S(\hat{\theta})$. Como $\alpha_S(\theta_0) \leq \alpha = \alpha_R(\hat{\theta_0})$ o lo que es lo mismo $P_{\theta_0}(T(\tilde{x}) \geq k') \leq P_{\theta_0}(T(\tilde{x} \geq k)$ obtenemos que $k \leq k'$, pero esto contradice $\beta_{S'}(\hat{\theta}) \leq \beta_S(\hat{\theta})$ ya que esto es equivalente a que $P_{\hat{\theta}}(\{T(\tilde{x}) \geq k'\}^c) < P_{\hat{\theta}}(\{T(\tilde{x}) \geq k\}^c)$ ya que esto implica $k \geq k'$. □

6.3. Método de la razón de verosimilitud para RC:

Consideremos $X_1,\ldots,X_n$ M.A.S. de $X \backsim F_X(x|\theta)$ con $\theta \in (H) \subset \mathbb{R}^k$ y la prueba
$$\begin{aligned} H_0 &: \theta \in A \subset (H) \\ H_1 &: \theta \notin A \end{aligned}$$

Planteamos una RC de la forma
$$R = \left\{\tilde{x} \in \mathbb{R}^n : \ \frac{\sup_{\theta\in A} L(\tilde{x}|\theta)}{\sup_{\theta\in H} L(\tilde{x}|\theta)} \leq k\right\}.$$

Observemos que para hipótesis simples $H_0 : \theta = \theta_0$ y $H_1 : \theta = \theta_1$ se obtiene
$$\sup_{\theta\in A} L(\tilde{x}|\theta) = L(\tilde{x}|\theta_0) = \prod_{i=1}^{n} f(x_i|\theta_0)$$
y
$$\sup_{\theta\in(H)} L(\tilde{x}|\theta) = \begin{cases} L(\tilde{x}|\theta_0) = \displaystyle\prod_{i=1}^{n} f(x_i|\theta_0) & \text{de donde } R = \emptyset \\ L(\tilde{x}|\theta_1) = \displaystyle\prod_{i=1}^{n} f(x_i|\theta_1) & \end{cases}$$

Entonces, la RC de de la razón de verosimilitud queda
$$\left\{\tilde{x} \in \mathbb{R}^n : \ \prod_{i=1}^{n}\frac{f(x_i|\theta_0)}{f(x_i|\theta_1)} \leq k\right\}$$
que es la R.C .O. del teorema de Neyman Pearson.

Ejemplo 6.29. Sea $X_1, \ldots, X_n$ M.A.S. de $X \backsim N(\mu, 1)$ y la prueba

$$\begin{aligned} H_0 &: \ \mu = \mu_0 \\ H_1 &: \ \mu \neq \mu_0 \end{aligned}$$

Hallaremos la RC de la razón de verosimilitud. Tenemos que

$$\sup_{\mu \in \mathbb{R}} L(\tilde{x}|\mu) = L(\tilde{x}|\overline{x})$$

y

$$L(\tilde{x}|\mu) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} x_i^2 - \frac{n}{2}\mu^2\right\} \exp\{n\mu\overline{x}\}$$

entonces

$$\frac{L(\tilde{x}|\theta_0)}{L(\tilde{x}|\overline{x})} = \frac{e^{-\frac{n\mu_0^2}{2} + n\mu_0\overline{x}}}{e^{-\frac{n\overline{x}^2}{2} + n\overline{x}^2}} = e^{-\frac{n\mu_0^2}{2} + n\mu_0\overline{x} - \frac{n\overline{x}^2}{2}} = e^{-\frac{n}{2}(\mu_0 - \overline{x})^2}$$

si planteamos la región crítica

$$\frac{L(\tilde{x}|\theta_0)}{L(\tilde{x}|\overline{x})} \leq k \Leftrightarrow -\frac{n}{2}(\overline{x} - \mu_0)^2 \leq L(k) = k' \Leftrightarrow |\overline{x} - \mu_0| \geq k''$$

por lo tanto la región crítica es de la forma

$$RC = \{\tilde{x} \in \mathbb{R}^n : |\overline{x} - \mu_0| \geq k\}$$

Proposición 6.30. *Consideremos la prueba*

$$\begin{aligned} H_0 &: \ \theta \in A \subset (H) \\ H_1 &: \ \theta \notin A \end{aligned}$$

$$\alpha(\tilde{x}) = \frac{\sup_{\theta \in A} L(\tilde{x}|\theta)}{\sup_{\theta \in (H)} L(\tilde{x}|\theta)} = \frac{\sup_{\theta \in A} g(T(\tilde{x}), \theta)h(\tilde{x})}{\sup_{\theta \in (H)} g(T(\tilde{x}), \theta)h(\tilde{x})} = \frac{\sup_{\theta \in A} g(T(\tilde{x}), \theta)}{\sup_{\theta \in (H)} g(T(\tilde{x}, \theta))} = \beta(T(\tilde{x}))$$

6.4. Pruebas de Bondad de ajuste

Se tiene una M.A.S. $X_1, \ldots, X_n$ de $X \backsim F_X$ desconocida. Dada F_0 una distribución, (conocida o no) se quiere tomar una decisión acerca de si X distribuye como F_0 o no, es decir,

$$\begin{aligned} H_0 &: \ F_X = F_0 \\ H_1 &: \ F_X \neq F_0 \end{aligned}$$

6.4.1. Test de χ^2:

Consideremos la prueba

$$\begin{aligned} H_0 &: F_X = F_0 \\ H_1 &: F_X \neq F_0 \end{aligned}$$

Dado $k \in \mathbb{N}$ elijo $I_1, \ldots, I_k$ intervalos en $\mathbb{R}$ tal que $I_i = (a_{i-1}, a_i]$, $I_1 = (-\infty, a_1]$ y $I_k = (a_k, +\infty]$ tal que $I_i \cap I_j = \emptyset$ si $i \neq j$, y $\cup_{i=1}^{k} I_i = \mathbb{R}$. Si H_0 es cierto $P(X \in I_j) = F_0(a_j) - F_0(a_{j-1}) = F_0(I_j)$, dada $X_1, \ldots, X_n$ M.A.S. de X definimos F_n^* la distribución emprica, sabemos que $F_n^*(I_j) \xrightarrow{c.s.} F_X(I_j)$. Sea b_j la cantidad de observaciones en I_j. Si tomo los valores esperados (bajo H_0 cierto) en el intervalo $I_j := E_j = nF_0(I_j)$, consideremos $T = \sum_{i=1}^{k}(b_j - E_j)^2$. Es razonable entonces construir la $RC = \{T \geq k\}$.
Si definimos $T_n = \frac{\sum(b_j - E_j)^2}{E_j}$, siendo b_j la cantidad de observaciones X_i que cayeron en el intervalo I_j, bajo la hipótesis H_0, se prueba que $T_n \xrightarrow{d} \chi^2_{k-1}$. Luego si $\alpha = P_{H_0}(T_n \geq k)$, se aproxima con la distribución de una χ^2_{k-1} y se halla un k aproximado.

6.4.2. Test de Kolmogorov-Smirnov

Consideremos

$$
\begin{aligned}
&H_0 : F_X = F_0 \;\; \text{completamente conocida}\\
&H_1 : F_X \neq F_0
\end{aligned}
$$

tomemos $RC = \{\sup_{x\in\mathbb{R}} |F_n^*(x) - F_0(x)| \geq k\}$, por $Gilvenco-Cantelli$ F_n^* converge uniformemente a $F_0(x)$. Para conocer la distribución de $\sup_{x\in\mathbb{R}} |F_n^*(x) - F_0(x)|$ tenemos el siguiente teorema.

Teorema 6.31. Kolmogorov: Si $D_n = \sup_{x\in\mathbb{R}} |F_n^*(x) - F_0(x)|$ entonces, si F_0 es continua

$$
\lim_{n\to+\infty} P\big(\sqrt{n}D_n \leq z\big) = 1 - 2\sum_{n=1}^{\infty}(-1)^{n-1}e^{-2n^2z^2}
$$

Definición 6.32. Dada una prueba de hipótesis

$$
\begin{aligned}
&H_0 :\; \theta \in A\\
&H_1 :\; \theta \notin A
\end{aligned}
$$

cuya región crítica sea $RC = \{T \geq k\}$ con $T = T(X_1, \dots, X_n)$ estimador de θ, el $p - valor$ es

$$
\sup_{\theta\in A} P\big(T(\tilde{X}) \geq T(\tilde{x})\big)
$$

Ejemplo 6.33. Sea $X_1, \dots, X_n$ M.A.S. de $X \backsim N(\mu, 1)$, consideremos la prueba

$$
\begin{aligned}
&H_0 :\; \mu = \mu_0 = 0\\
&H_1 :\; \mu \neq \mu_0 = 0
\end{aligned}
$$

Sabemos que $RC = \{|\overline{x_n}| \geq k\}$ entonces $T(X_1, \dots, X_n) = |\overline{X_n}|$ el $p - valor$ es

$$
P_{H_0}(|\overline{X_n}| \;\geq\; |\overline{x}|) \;=\; 1 - P_{H_0}(|\overline{X_n}| \;\leq\; |\overline{x}|) \;=\; 1 - \Phi(\sqrt{n}|\overline{x}|) + \Phi(-\sqrt{n}|\overline{x}|) \;=\; 2\big(1 - \Phi(\sqrt{n}|\overline{x}|)\big)
$$

Proposición 6.34. *Si los supremos se realizan en un mismo $\theta_0 \in A$, $\alpha < p - valor \Leftrightarrow$ no rechazo H_0 al nivel α.*

Demostración. Si $\alpha < p - valor$ entonces hallamos k tal que $\alpha = \sup_{\theta\in A} P(\big(T(\tilde{X} \geq k)\big)$,

$$
\alpha = \sup_{\theta\in A} P\big(T(\tilde{X}) \geq k\big) < \sup_{\theta\in A} P\big(T(\tilde{X}) \geq T(\tilde{x})\big)
$$

$$
\alpha = P_{\theta_0}\big(T(\tilde{X}) \geq k\big) < P_{\theta_0}\big(T(\tilde{X}) \geq T(\tilde{x})\big)
$$

de donde $T(\tilde{x}) < k$ por lo tanto $\tilde{x} \notin RC$ y no rechado H_0. El razonamiento es análogo si $\alpha > p - valor$. □

Observación 6.35. *La propiedad se cumple si H_0 es simple ($\theta = \theta_0$), o en el caso de concientes de verosimilitud monótonos.*

6.5. Análisis de Varianza, (ANOVA)

Supongamos que tenemos $\{Y_{ij}\}$ observaciones, con $i = \{1, \dots, k\}$ y $j = \{1, \dots, n_j\}$ y que $Y_{ij} \backsim N(\theta_i, \sigma^2)$ para todo i, j. Queremos testear si los θ_i son todos iguales o no. El supuesto de que σ^2 es la misma se llama homocedasticidad. Supongamos que las variables Y_{ij} son independientes. Para cada $i \in \{1, \dots, k\}$ definimos

$$
\overline{Y_i} = \frac{1}{n_i}\sum_{j=1}^{n} Y_{ij},
$$

y

$$S_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y_{ij}})^2.$$

Sabemos que

$$\overline{Y_i} \backsim N(\theta_i, \sigma^2/n_i) \qquad \frac{(n_i-1)}{\sigma^2}S_i^2 \backsim \chi^2_{n_i-1}$$

Observación 6.36. *Si $\mathcal{A} = \{a = (a_1, \ldots, a_k) \in \mathbb{R}^k : \sum a_i = 0\}$ entonces*

$$\theta_1 = \cdots = \theta_k \Leftrightarrow \ \forall a \in \mathcal{A}, \ \sum a_i\theta_i = 0$$

Demostración. El directo es inmediato, veamos el recíproco, tomemos $a_1 = 1, a_2 = -1, a_3 = \cdots = a_k = 0$ entonces $\theta_1 - \theta_2 = 0$ y as sucesivamente $\theta_1 = \cdots = \theta_k$. □

Observación 6.37. *Si defino $S_p^2 = \frac{1}{N-k}\sum_{i=1}^k (n_i-1)S_i^k$ con $N = \sum n_i$. entonces*

$$\frac{N-k}{\sigma^2}S_p^2 = \sum_{i=1}^k \frac{(n_i-1)}{\sigma^2}S_i^2 \backsim \chi^2_{N-k}.$$

Adems

$$\sum_{i=1}^k a_i\overline{Y_i} \backsim N\left(\sum_{i=1}^k a_i\theta_i, \frac{\sum_{i=1}^k a_i^2\sigma^2}{n_i}\right).$$

Se puede demostrar que S_p^2 y $\sum a_i\overline{Y_i}$ son independientes, luego, si recordamos que si $X \backsim N(0,1)$ es independiente de χ_n^2 entonces

$$\frac{X}{\sqrt{\chi_n^2/n}} \backsim t_n$$

obtenemos que

$$\frac{\dfrac{\sum_{i=1}^k a_i\overline{Y_i} \ - \ \sum_{i=1}^k a_i\theta_i}{\sigma\sqrt{\sum_{i=1}^k a_i^2/n_i}}}{\sqrt{\frac{(N-K)}{\sigma^2}S_p^2/(N-k)}} = \frac{\sum_{i=1}^k a_i(\overline{Y_i} - \theta_i)}{S_p\sqrt{\sum_{i=1}^k a_i^2/n_i}} \backsim t_{N-k}$$

Supongamos que $a \in \mathcal{A}$ fijo, y $\alpha \in (0,1)$ tenemos

$$H_0 : \sum a_i\theta_i = 0$$
$$H_1 : no\ H_0$$

Consideremos la región crítica,

$$RC = \left\{\frac{|\sum a_i\overline{Y_i}|}{S_p\sqrt{\sum a_i/n_i}} > m\right\},$$

$$\alpha = P_{H_0}(RC) = P_{H_0}\left(\frac{|\sum a_i\overline{Y_i}|}{S_p\sqrt{\sum a_i/n_i}} > k\right),$$

como estamos bajo H_0 si utilizamos la observación anterior

$$\alpha = 1 - P(-m < T < m), \text{ con } T \backsim t_{N-k},$$

y por lo tanto $1 - \alpha/2 = F(m)$, $k = t_{1-\alpha/2}(N-k)$.

Nos planteamos ahora la siguiente prueba

$$H_0 : \theta_1 = \cdots = \theta_k$$
$$H_1 : noH_0$$

y esto es si y solo si

$$H_0 : \sum a_i\theta_i = 0 \quad \forall a \in \mathcal{A}$$
$$H_1 : noH_0$$

Tomo el estadístico $T_a = \dfrac{\sum a_i\overline{Y_i}}{S_p\sqrt{\sum a_i^2/n_i}}$, resulta natural plantear la región crítica $RC = \left\{\sup_{a\in\mathcal{A}} T_a^2 > k\right\}$. Debemos entonces hallar la distribución de $\sup_{a\in\mathcal{A}} T_a$ bajo la hipótesis H_0 cierto. Llamemos $C_i = \overline{Y_i}$ y $\overline{C_i} = \frac{\sum n_iC_i}{N}$.

$$\sup_{a\in\mathcal{A}} T_a^2 = \frac{1}{S_p}\sup_{a\in\mathcal{A}}\frac{\big(\sum a_iC_i\big)^2}{\sum a_i/n_i} = \sup_{a\in\mathcal{A}}\frac{\big(\sum \frac{a_i}{\sqrt{n_i}}(C_i-\overline{C})\sqrt{n_i}\big)^2}{\sum a_i/n_i},$$

donde hemos usado que $\sum a_i\overline{C} = 0$, si aplicamos la desigualdad de Cauchy-Schwartz

$$\sup_{a\in\mathcal{A}}\frac{\big(\sum \frac{a_i}{\sqrt{n_i}}(C_i-\overline{C})\sqrt{n_i}\big)^2}{\sum a_i/n_i} \leq \sup_{a\in\mathcal{A}}\frac{\sum a_i^2/n_i\sum n_i(C_i-\overline{C})^2}{\sum a_i/n_i} = \sum n_i(C_i-\overline{C})^2.$$

Obtuvimos una cota para el supremo, veamos que se alcanza, si tomamos $a_i = cten_i(C_i-\overline{C})$ es claro que $\sum a_i = 0$, entonces el supremo se alcanza. (Basta observar que la igualdad en Cauchy-Schwartz se da en ese caso).

$$\sup_{a\in\mathcal{A}} T_a^2 = \frac{\sum_{i=1}^k n_i(\overline{Y_i}-\overline{\overline{Y}})^2}{S_p^2} \qquad \text{donde} \qquad \overline{\overline{Y}} = \frac{\sum_{i=1}^k n_i\overline{Y_i}}{N},$$

recordemos que

$$\frac{\chi_n^2/n}{\chi_m^2/m} \backsim F(n,m),$$

se puede demostrar que $\displaystyle\sum_{i=1}^k n_i(\overline{Y_i}-\overline{\overline{Y}})^2 \backsim \chi_{k-1}^2$ y por lo tanto

$$\sup_{a\in\mathcal{A}}\frac{\sigma^2\chi_{k-1}^2}{\sigma^2\chi_{N-k}^2/(N-k)} > cte \quad \Leftrightarrow \frac{\chi_{k-1}^2/(k-1)}{\chi_{N-k}^2/(N-k)} \backsim F(k-1,N-k) \geq cte/(k-1).$$

Planteamos

$$\alpha = P_{H_0}(RC) = 1 - P_{H_0}\big(F(k-1,N-k) \leq \frac{cte}{k-1}\big) \quad entonces \quad cte = F_{1-\alpha}(k-1,N-k)(k-1).$$

Finalmente, obtuvimos la región crítica

$$RC = \left\{\frac{1}{S_p}\sum_{i=1}^k n_i(\overline{Y_i}-\overline{\overline{Y_i}})^2 \geq F_{1-\alpha}(k-1,N-k)(k-1)\right\}.$$

Capítulo 7

Modelos Lineales

7.1. Variable Normal Multivariada

Definición 7.1. Dado un vector aleatorio $(X_1, \ldots, X_n)$ recordemos que el vector de medias $\mu = (\mu_1, \ldots, \mu_n) := (E(X_1), \ldots, E(X_n))$, y la matriz de covarianzas es

$$\Sigma_{n\times n} = \begin{pmatrix} Var(X_1) & cov(X_1X_2) & \ldots & cov(X_1X_n) \\ cov(X_2X_1) & Var(X_2) & & \\ \vdots & \vdots & \ddots & \\ cov(X_nX_1) & \ldots & & Var(X_n) \end{pmatrix}$$

Anotamos $\mu = E(X)$ y $Var(X) = \sigma_{n\times n}$.

Observación 7.2. *Veamos algunas propiedades*

1) Si $A \in M_{k\times n}$ es constante entonces $E(AX) = AE(X)$.

2) $Var(AX) = A\Sigma_{n\times n}A^t$.

3) Si $X \in \mathbb{R}^n$ es un vector aleatorio A es una matriz $k \times n$ y b un vector $k \times 1$ constante entonces

$$E(AX + b) = AE(X) + b \quad y \quad Var(AX + b) = A\Sigma_X A^t.$$

4) Si X es un vector aleatorio en $\mathbb{R}^n$, σ_X es semidefinida positiva.

Demostración.

2) Es inmediato a partir de observar que $Var(X) = E\big((X - E(X))(X - E(X))^t\big)$.

4) Tenemos que ver que para todo $\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n$ entonces $\lambda\Sigma\lambda^t \geq 0$, y esto se sigue de que $\lambda\Sigma\lambda^t = Var(\sum \lambda_i X_i)$.

□

Definición 7.3. Normal típica en $\mathbb{R}^n$: Decimos que el vector $U = (U_1, \ldots, U_n)$ tiene distribución normal típica en $\mathbb{R}^n$ si las $U_i \backsim N(0, 1)$ y son independientes.

Observación 7.4. *La densidad conjunta de U es*

$$f_U(x) = \frac{e^{-\frac{1}{2}\|x\|^2}}{(2\pi)^{\frac{n}{2}}}.$$

Definición 7.5. Normal multivariada Decimos que X tiene distribución normal multivariada si existe una matriz $n \times k$ C y un vector μ $n \times 1$ tal que $X = CU + \mu$.

Observación 7.6. *Observemos que si X tiene distribución normal multivariada entonces $E(X) = \mu$ y $\Sigma_X = CC^t$*

Proposición 7.7. *Veamos algunas propiedades de la normal multivariada*

1) Si $C_{n\times n}$ es invertible, X es absolutamente coninua y

$$f_X(x) = \frac{e^{-\frac{1}{2}(x-\mu)^t\Sigma^{-1}(x-\mu)}}{(2\pi)^{n/2}|\det \Sigma|^{1/2}} \quad \Sigma = CC^t.$$

Demostración. $X = CU + \mu = g(U)$, $g : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ es invertible ya que C lo es.

$$\begin{aligned} f_X(x) = f_{g(U)}(x) = \; & f_U((g^{-1}(x))\frac{1}{|\det J_g(g^{-1}(x))|} = f_U(C^{-1}(x-\mu))\frac{1}{|\det C|} \\ = \; & \frac{e^{-\frac{1}{2}(x-\mu)(C^{-1})^tC^{-1}(x-\mu)}}{(2\pi)^{n/2}|\det \Sigma|^{1/2}} \end{aligned}$$

□

2) La distribución normal tpica es invariante bajo tranformaciones ortogonales. De hecho es la nica distribución que depende solamente de la norma, y que es invariante bajo transformaciones ortogonales (a menos de multiplicarla por constantes). Que es invariante bajo transformaciones ortogonales se sigue de la definición y de la propiedad anterior.

3) Si X es normal multivariada, entonces $AX + b$ tambien lo es, con $A_{m\times n}$ y $b_{m\times 1}$ constantes.

4) Si $X = CU + \mu$ y C es sobreyectiva entonces X es absolutamente continua.

Definición 7.8. Normal multivariada degenerada: Si $X = CU + \mu$ con U normal tpica, decimos que es degenerada si C no es sobreyectiva

Observación 7.9. *Si X es degenerada entonces no es absolutamente conitnua.*

Demostración. Supongamos por absurdo que existe una densidad f_X. Recordemos que C no es sobre si y solo si $\det(CC^t) = \det(\Sigma) = 0$, si $\det(\Sigma) = 0$ entonces $t\Sigma t^t = Var(tX) = 0$ entonces tX es c.s. constante, de donde se sigue que esta contenida en un hiperplano S, si existiese $f_X(x_1, \dots, x_n)$, al integrarla en S obtendramos que debera dar 1 porque X esta contenida ahi, pero 0 porque S tiene medida nula, absurdo. □

Observación 7.10. *Si $X \backsim N(\mu, \Sigma)$ cualquier subvector de X tambien es normal multivariado. Esto es obvio de hecho de que si X es normal multivariado, AX tambien lo es, basta tomar A adecuadamente.*

Observación 7.11. *Si $(X_1, \dots, X_k, Y_1, \dots, Y_k) \backsim N(\mu, \Sigma)$ entonces si $cov(X_i, Y_j) = 0 \quad \forall i, j$ enontces $(X_1, \dots, X_k)$ y $(Y_1, \dots, Y_k)$ son independientes.*

Demostración. Si Σ es invertible, entonces

$$\Sigma^{-1} = \begin{pmatrix} \Sigma_X^{-1} & 0 \\ 0 & \Sigma_Y^{-1} \end{pmatrix}$$

□

y es fácil ver que $f_{X,Y}(x) = g_X(x)g_Y(y)$.
Si Σ no es invertible, y Σ_Y si, entonces $(X_1, \dots, X_k) \in S$, variedad lineal de dimensión $\alpha - k$, supongamos que S esta generado por $X_1, \dots, X_\alpha$ entonces
$(X_1, \dots, X_\alpha, Y_1, \dots, Y_m)$ est en las hipótesis anteriores y por lo tanto son independientes, de donde $(X_1, \dots, X_k, Y_1, \dots, Y_m)$ lo son. El caso en que Σ_Y tampoco es invertible es anlogo.

7.2. Modelos Lineales

Se desea estimar $Y = g(X_1, \dots, X_k)$, se observan medidas de las variables $X_1, \dots, X_k$ y se desea estimar g. A las variables X_i se las denomina explicativas y a la Y explicada. Se plantea entonces $g(\tilde{x}) = g(\tilde{x}, \theta) = \theta_1 X_1 + \dots + \theta_k X_k$, $\theta = (\theta_1, \dots, \theta_k)$. Para estimar g estimamos θ. Se plantea entonces

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} X_{11} & X_{12} & \dots & X_{1k} \\ \vdots & \vdots & \ddots & \\ X_{n1} & X_{n2} & \dots & X_{nk} \end{pmatrix} \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_k \end{pmatrix}$$

Se observan n muestras de Y,

$$X = \begin{pmatrix} X_{11} & X_{12} & \dots & X_{1k} \\ \vdots & \vdots & \ddots & \\ X_{n1} & X_{n2} & \dots & X_{nk} \end{pmatrix},$$

es la matriz de diseño (constante y conocida).

En el modelo lineal planteamos $Y = X\theta + e$ donde, X es una matriz de diseño, y

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \quad \theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_k \end{pmatrix} \quad e = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix},$$

e aleatorio (vector de errores).

Ejemplo 7.12. Análisis de varianza: $Y_{ij} = \theta_i + e_{ij}$, en este caso la matriz X es un vector $n \times 1$ con entradas todas iguales a 1.

Ejemplo 7.13. Modelo lineal simple: $Y = \alpha + \beta X + e$, tomamos $(Y_1, X_1), \dots, (Y_n, X_n)$ y $\theta = (\alpha, \beta)$, y como matriz de diseño la matriz

$$X = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \\ 1 & X_n \end{pmatrix},$$

lo que se busca es entonces ajustar una recta a los datos.

Ejemplo 7.14. Ajuste de un polinomio de grado k: De forma análoga al ejemplo anterior, si $Y = \alpha + \beta_1 x + \beta_2 x^2 + \dots + \beta_k x^k + e$, planteamos la matriz de diseño

$$X = \begin{pmatrix} 1 & X_1 & X_1^2 & \dots & X_1^k \\ \vdots & \vdots & & & \\ 1 & X_n & X_n^2 & \dots & X_n^k \end{pmatrix}.$$

Observación 7.15. *Observemos que, en vistas del ejemplo anterior, la función $y = g(x_1, \dots, x_n, \theta)$ es lineal en θ pero no en $\tilde{x} = (x_1, \dots, x_n)$, podrá ser $x_3 = cos(x_1)$ etc.*

7.3. Hipótesis del modelo

1) *$Rango(g(X)) = k$.*

2) Los errores tienen media 0, $E(e_i) = 0$ para todo i.

3) Homocedasticidad: $Var(e_i) = \sigma^2$ para todo i.

3') $cov(e_i, e_j) = 0$ para todo $i \neq j$.

4) el vector e de errores tiene distribución $N(0, \sigma^2 I)$ en este caso se cumplen $2), 3)$ y $3')$

Para estimar $\theta \in \mathbb{R}^k$ se utiliza el método de los mínimos cuadrados, consiste en hallar $\theta \in \mathbb{R}^k$ donde se realize

$$\min_{\theta \in \mathbb{R}^k} \|Y - X\theta\|.$$

Teorema 7.16. *Bajo la hipótesis 1 se cumple que $(X^tX)^{-1}X^tY$ es el estimador por mínimos cuadrados de θ.*

Demostración. Sea $\hat{\theta}$ el valor donde se obtiene el mínimo, es decir

$$\|Y - X\hat{\theta}\|^2 \leq \|Y - X\theta\|^2 \quad \forall \theta \in \mathbb{R}^k,$$

si consideramos la multiplicación por X como una transformación lineal de $\mathbb{R}^k$ en $\mathbb{R}^n$ entonces $X\hat{\theta}$ es la proyección de Y sobre la imagen de X, entonces $Y - X\hat{\theta} \perp X\theta$ para todo $\theta \in \mathbb{R}^k$, esto es $0 = (X\theta)^t(Y - X\hat{\theta})$, o lo que es lo mismo

$$\theta^t X^t Y = \theta^t X^t X \hat{\theta} \quad \forall \theta \in \mathbb{R}^k,$$

luego las transformaciones lineales X^tY y $X^tX\theta$ son iguales, de donde $\hat{\theta} = (X^tX)^{-1}X^tY$. $\square$

Teorema 7.17.

a) Bajo las hipótesis $1)$ *y* $2)$*, $\hat{\theta}$ es insesgado.*

b) Bajo las hipótesis $1), 2)$ *y* $3)$*, $\Sigma_{\hat{\theta}} = \sigma^2(X^tX)^{-1}$.*

Demostración.

a) $E(\hat{\theta}) = (X^tX)^{-1}X^tE(Y) = (X^tX)^{-1}X^tX\theta = \theta$.

b) $\Sigma_{\hat{\theta}} = \Sigma_{(X^tX)^{-1}X^te+\theta} = (X^tX)^{-1}X^t(\sigma^2 Id)X(X^tX)^{-1}$, donde usamos que $\Sigma_{AX+C} = A\sigma_X A^t$, finalmente se obtiene, $\Sigma_{\hat{\theta}} = \sigma^2(X^tX)^{-1}$, ya que transponer e invertir conmutan.

$\square$

Teorema 7.18. *Bajo las hipótesis* $1)$ *a* $4)$ *el E.M.V. de θ coincide con el de mínimos cuadrados y adems el E.M.V. de σ es $\frac{1}{n}\|Y - X\hat{\theta}\|$.*

.

Teorema 7.19. *Bajo las hiptesis* $1)$ *a* $4)$ *$\hat{\theta}$ es insesgado de mnima varianza, uniformemente.*

Demostración. Veamos que es suficiente:

$$L(y_1, \ldots, y_n|\theta, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\|Y - X\hat{\theta}\|^2\right\} \exp\left\{-\frac{1}{2\sigma^2}\|X\hat{\theta} - X\theta\|^2\right\} = h(\tilde{y})g(\hat{\theta}, \theta)$$

donde hemos usado que $Y - X\hat{\theta}$ es perpendicular a $X\hat{\theta} - X\theta$. Es fácil ver que es completo y por lo tanto minimiza el riesgo uniformemente entre los insesgados, considerando como función de riesgo $\|\|^2$. $\square$

Teorema 7.20. *Bajo* $1)$ *,* $2)$ *y* $3)$*, si los e_i son independientes (no necesariamente con distribución Noramal), entonces $\hat{\theta}$ es uniformemente de mínima varianza entre los estimadores lineales e insesgados, (es decir los $\tilde{\hat{}} = CY$).*

Teorema 7.21. *Bajo los supuestos 1) a 4):*

a) $\dfrac{n\hat{\sigma}^2}{\sigma^2} = \dfrac{\|Y - X\hat{\theta}\|^2}{\sigma^2} \backsim \chi^2_{(n-k)}$

b) $s^2 = \dfrac{n\hat{\sigma}^2}{n-k} = \dfrac{\|Y - X\hat{\theta}\|}{n-k}$ *es insesgado (de donde $\hat{\sigma}^2$ es asintóticamente insesgado).*

c) $\dfrac{\|X(\hat{\theta} - \theta)\|^2}{ks^2} \backsim F(k, n-k)$

d) $\dfrac{\lambda_1(\hat{\theta}_1 - \theta_1) + \lambda(\hat{\theta}_2 - \theta_2) + \cdots + \lambda_n(\hat{\theta}_n - \theta_n)}{s\sqrt{\lambda^t(X^tX)^{-1}\lambda}} \backsim t_{n-k} \quad \forall \lambda \in \mathbb{R}^n$

Demostración. a) Sea $H = \{v_1, \ldots, v_n\}$ base ortonormal de $\mathbb{R}^n$ tal que $\{v_1, \ldots, v_k\}$ es base ortonormal de $S = Im(X)$, tenemos entonces que existen $Z_1, \ldots, Z_n$ variables aleatorias tal que $Y = \sum_{i=1}^n Z_i v_i$. Si B es la matriz de cambio de base de la base H a la base cannica, B es ortogonal y

$Y = BZ$ de donde $Z = B^{-1}Y = B^tY \backsim N(B^tX\theta, B^t\sigma^2 IdB)$ y por lo tanto Z es normal multivariado y $\Sigma_Z = \sigma^2 Id$, adems Z_i son variables aleatorias independientes con distribución $N(\gamma_1, \sigma^2)$.

$$\|Y - X\hat{\theta}\|^2 = \left\|\sum_{i=1}^n Z_i v_i - \sum_{j=1}^k Z_j v_j\right\|^2 = \left\|\sum_{k=1}^n Z_i v_i\right\|^2 = \sum_{i=k+1}^n Z_i^2,$$

$$\frac{\|Y - X\hat{\theta}\|^2}{\sigma^2} = \sum_{i=k+1}^n \left(\frac{Z_i}{\sigma}\right)^2$$

$$\frac{Z_i}{\sigma^2} \backsim N(\gamma_i, 1),$$

bastaría entones demostrar que todos los γ_i para $i = k+1$ son 0. Observemos que $E(Y) = X\theta \in S$ y $E(Y) = \sum_{i=1}^n \gamma_i v_i$.

b) $E(s^2) = \dfrac{1}{n-k} E\big(\|Y - X\hat{\theta}\|\big) = \dfrac{\sigma^2}{n-k} E\left(\dfrac{\|Y - X\hat{\theta}\|}{\sigma^2}\right) = \sigma^2$

c) $\|X\hat{\theta} - X\theta\|^2 = \left\|\sum_{i=1}^k Z_i v_i - \sum_{i=1}^k \gamma_i v_i\right\|^2 = \sum_{i=1}^k (Z_i - \gamma_i)^2$. entonces

$$\frac{\|X\hat{\theta}X - \theta\|}{ks^2} = \frac{\sum \left(\frac{Z_i - \gamma_i}{\sigma}\right)^2}{\frac{k}{\sigma^2}\frac{1}{n-k}\|Y - X\hat{\theta}\|} \backsim \frac{\chi^2_k / k}{\chi^2_{n-k}/(n-k)} \backsim F(k, n-k).$$

d) $\lambda_1(\hat{\theta}_1 - \theta_1) + \cdots + \lambda_n(\hat{\theta}_n - \theta_n) = \lambda^t(\hat{\theta} - \theta)$, como $\hat{\theta} \backsim N(\theta, \sigma^2(X^tX)^{-1})$, entonces $\lambda^t(\hat{\theta} - \theta) \backsim N(0, \lambda^t\sigma^2(X^tX)^{-1}\lambda)$,

$$\frac{\lambda^t(\hat{\theta} - \theta)}{s\sqrt{\lambda^t(X^tX)^{-1}\lambda}} = \frac{\frac{\lambda^t(\hat{\theta} - \theta)}{\sigma\sqrt{\lambda(X^tX)^{-1}\lambda}}}{\frac{s}{\sigma}},$$

por lo tanto si usamos la parte *b)* solo basta ver que son independientes, esto se sigue de que $\|Y - X\theta\|^2$ depende de $Z_{k+1}, \ldots, Z_n$ y $X\hat{\theta}$ de $Z_{k+1}, \ldots, Z_k$. □

7.4. Aplicación:

Construccón de intervalos de confianza para $\lambda^t\theta$. Consideremos

$$I = \left[\lambda^t\hat{\theta} - ks, \lambda^t\hat{\theta} + ks\right],$$

$$1-\alpha = P(\lambda^t\theta \in I) = P\left(\left|\frac{\lambda^t(\hat{\theta}-\theta)}{s}\right| \leq k\right) = P\left(\left|\frac{\lambda^t(\hat{\theta}-\theta)}{s\sqrt{\lambda^t(X^tX)_{-1}\lambda}}\right| \leq \frac{k}{\sqrt{\lambda^t(X^tX)^{-1}\lambda}}\right),$$

de donde, por la parte *d)* $k = t_{1-\alpha/2}(n-k)\sqrt{\lambda^t(X^tX)^{-1}\lambda}$.
Observemos que en particular tomando $\lambda = (1, \ldots, 0)$ obtenemos un intervalo de confianza para θ_1.

Capítulo 8

Test de Aleatoriedad

8.1. Introducción

En éste capítulo veremos algunos test que permiten chequear cuándo una muestra $X_1, \dots, X_n$ a valores en $\mathbb{R}$ cumple las hipótesis de ser independiente e idénticamente distribuida, como no haremos suposiciones respecto de la distribución de las X_i veremos métodos muy básicos que lo que *miden* es la forma en que las variables están ordenadas.

8.2. Test de Rachas para muestras de 2 tipos

En esta sección vamos a suponer que tenemos n_1 objetos de un cierto tipo, y n_2 de otro, y estamos interesados en la forma en que estos objetos se distribuyen. Podemos pensar que estamos observando el género de las personas que forman una cola de espera. Una posible observación sería $M, F, M, F, M, F, M, F, M, F$, en este caso es evidente que el orden de llegada de los géneros no es aleatorio, lo mismo pasa si observamos $M, M, M, M, M, M, F, F, F, F, F, F$. Dada una secuencia de objetos de dos tipos, una racha es una sucesión de objetos de un tipo, seguida y precedida por objetos del otro tipo, o por ningún objeto, por ejemplo en F, M, F tenemos 2 rachas de objetos de tipo F y 1 de objetos de tipo M, en F, F, F tenemos una sola racha. Intuitivamente, si observamos muchas rachas o muy pocas estamos ante un caso en el que los objetos no se distribuyen de forma independiente. Para ser más rigurosos podemos pensar que tenemos $X_1, \dots, X_n$ variables que toman valores 0 o 1, y estamos testeando si son i.i.d., condicionado a que $\sum X_i = n_1$.

8.2.1. Test basados en el número total de rachas

Supongamos que tenemos n objetos, n_1 de un tipo y n_2 de otro, $(n_1 + n_2 = n)$, anotaremos r_1 al número de rachas de tipo 1, y r_2 al número de rachas de tipo 2. El número total de rachas es $r = r_1 + r_2$, veremos un test basado en la variable aleatoria R que nos da el numero total de rachas, vamos a calcular la distribución de R, para el caso en que se cumple la hipótesis de ser i.i.d.. Bajo dicha hipótesis, cualquier ordenación de los n objetos tiene la misma probabilidad, veamos con un ejemplo sencillo como calcular la distribución de R, supongamos que tenemos los objetos $*, *$ y $+, +$, las posibles reordenaciones son $++**$, $*+*+$, $**++$, $+**+$, $+*+*$ y $*++*$, y el numero total de rachas es 2,4,2,4,3,3 respectivamente, por lo tanto la probabilidad de obtener por ejemplo 2 rachas es $2/6 = 1/3$. Para el caso general, el número total de posibles reordenaciones es $n!/(n_1!n_2!)$, por lo tanto para calcular por ejemplo la probabilidad de tener r_1 rachas de tipo 1, y r_2 de tipo 2, tenemos que contar la cantidad de formas de intercalar los n objetos, que dan como resultado r_1 rachas de tipo 1, y r_2 de tipo 2. Observemos que si tenemos n_1 objetos de tipo 1 y queremos contar la cantidad de ordenaciones que se pueden formar con dichos objetos de modo de generar r_1 rachas, y con la hipótesis adicional de que siempre se comience con un objeto de tipo 1, es

$$\binom{n_1-1}{r_1-1}$$

para visualizar esto basta observar que tenemos n_1-1 lugares donde poner divisiones de los n_1 objetos y que tomando r_1-1 de dichas divisiones genero r_1 rachas, por ejemplo si tenemos

$$*\,*\,*\,*\,*\,*\,*\,*\,*\,**$$

y queremos tener 5 rachas, y que se comience con un $*$ algunas posibilidades de separación son

$$**\,|\,**|\,*\,|\,****|\,* \quad *|\,*****\,|\,*\,|\,*\,|\,** \quad *\,|\,*\,|\,*\,|\,*\,|\,******,$$

si además queremos tener r_2 rachas de tipo 2, tenemos que colocar en los $|$ los objetos de tipo 2, esto nos da otras $\binom{n_2-1}{r_2-1}$, de forma análoga si comenzamos con objetos de tipo 2. Observemos que como los bloques de objetos de tipo 1 y 2 se alternan solo tenemos 3 casos, o bien $r_1=r_2+1$ como en el ejemplo anterior, o $r_1=r_2-1$, o $r_1=r_2$. Observemos que si $r_1=r_2+1$ la secuencia tiene que empezar necesariamente con un objeto de tipo 1, si $r_2=r_1+1$ tiene que empezar necesariamente con objetos de tipo 2, mientras que si $r_1=r_2$ puede empezar o bien con objetos de tipo 1, o 2, (observar en el ejemplo de $**$ y $++$). Finalmente hemos demostrado el siguiente teorema.

Teorema 8.1. *Sean R_1 el número de rachas de objetos de tipo 1, y R_2 el número de rachas de tipo 2, en una muestra aleatoria de $n=n_1+n_2$ objetos, la distribución conjunta de R_1 y R_2 es*

$$f_{R_1,R_2}(r_1,r_2)=\frac{c\binom{n_1-1}{r_1-1}\binom{n_2-1}{r_2-1}}{\binom{n_1+n_2}{n_1}}$$

para $r_1=1,\dots,n_1 \quad r_2=1,\dots,n_2$ donde $c=2$ si $r_1=r_2$ y $c=1$ si $r_1=r_2\pm 1$.

Corolario 8.2. *La distribución marginal de probabilidades de R_1 es*

$$f_{R_1}(r_1)=\frac{\binom{n_1-1}{r_1-1}\binom{n_2-1}{r_1}}{\binom{n_1+n_2}{n_1}}$$

donde $r_1=1,\dots,n_1$.

Demostración. Del teorema anterior, sólo tenemos que calcular los tres sumandos:

$$f_{R_1}(r_1)=\sum_{r_1} f_{R_1,R_2}(r_1,r_2),$$

$$\begin{aligned}
\binom{n_1+n_2}{n_1}f_{R_1}(r_1) &= 2\binom{n_1-1}{r_1-1}\binom{n_2-1}{r_1-1}+\binom{n_1-1}{r_1-1}\binom{n_2-1}{r_1-2}+\binom{n_1-1}{r_1-1}\binom{n_2-1}{r_1}\\
&=\binom{n_1-1}{r_1-1}\left[\binom{n_2-1}{r_1-1}+\binom{n_2-1}{r_1-2}+\binom{n_2-1}{r_1-1}+\binom{n_2-1}{r_1}\right]\\
&=\binom{n_1-1}{r_1-1}\left[\binom{n_2}{r_1-1}+\binom{n_2}{r_1}\right]\\
&=\binom{n_1-1}{r_1-1}\binom{n_2+1}{r_1}
\end{aligned}$$

$\square$

Teorema 8.3. *La distribución de R, el número total de rachas de $n=n_1+n_2$ objetos, n_1 de tipo 1 y n_2 de tipo 2, en una muestra aleatoria es*

$$f_R(r)=\begin{cases}\dfrac{2\binom{n_1-1}{r/2-1}\binom{n_2-1}{r/2-1}}{\binom{n_1+n_2}{n_1}} & \text{si } r \text{ es par}\\[2ex] \dfrac{\binom{n_1-1}{(r-1)/2}\binom{n_2-1}{(r-3)/2}+\binom{n_1-1}{(r-3)/2}\binom{n_2-1}{(r-1)/2}}{\binom{n_1+n_2}{n_1}} & \text{si } r \text{ es impar}\end{cases}$$

Momentos de R

Veamos ahora como aplicar los resultados anteriores para calcular los momentos de R bajo la hipótesis de que las observaciones son i.i.d.

$$\begin{aligned}
E(R^k) &= \sum_r r^k f_R(r) \\
&= \frac{\sum_{\{r \text{ par}\}} 2r^k \binom{n_1-1}{r/2-1}\binom{n_2-1}{r/2-1}}{\binom{n_1+n_2}{n_1}} + \\
&\quad \frac{\sum_{\{r \text{ impar}\}} r^k \left[\binom{n_1-1}{(r-1)/2}\binom{n_2-1}{(r-3)/2} + \binom{n_1-1}{(r-3)/2}\binom{n_2-1}{(r-1)/2}\right]}{\binom{n_1+n_2}{n_1}}
\end{aligned}$$

El valor más pequeño para r es siempre 2. Si $n_1 = n_2$ el valor más grande para r es $2n_1$, si $n_1 < n_2$ el máximo valor para r es $2n_1 + 1$. Si asumimos sin pérdida de generalidad que $n_1 \leq n_2$, el rango de r es $2 \leq r \leq 2n_1 + 1$. Si tomamos $r = 2i$ para r par (y $r = 2i + 1$ para r impar) i varía entre $1 \leq i \leq n_1$, por ejemplo para la media de R podemos escribir las sumatorias anteriores como

$$\begin{aligned}
\binom{n_1+n_2}{n_1} E(R) = \sum_{i=1}^{n_1} 4i \binom{n_1-1}{i-1}\binom{n_2-1}{i-1} + \sum_{i=1}^{n_1} (2i+1)\binom{n_1-1}{i}\binom{n_2-1}{i-1} \\
+ \sum_{i=1}^{n_1} (2i+1)\binom{n_1-1}{i-1}\binom{n_2-1}{i}
\end{aligned}$$

Para calcular estas sumatorias son útiles los siguientes lemas:

Lema 8.4.

$$\sum_{r=0}^{c} \binom{m}{r}\binom{n}{r} = \binom{m+n}{m} \qquad \textit{con } c = \text{mín}\{m, n\}$$

Lema 8.5.

$$\sum_{r=0}^{c} \binom{m}{r}\binom{n}{r+1} = \binom{m+n}{m+1} \qquad \textit{con } c = \text{mín}\{m, n-1\}$$

Una forma más simple de calcular el valor esperado y la varianza de R es observar que si $n = n_1 + n_2$, entonces

$$R = 1 + I_2 + I_3 + \cdots + I_n$$

donde,

$$I_k = \begin{cases} 1 & \text{si el } k - \text{ésimo elemento es distnto del } k-1 \\ 0 & \text{en caso contrario} \end{cases}$$

Observemos que I_k es una variable aleatoria con distribución Bernoulli de parámetro $p = 2n_1n_2(n-2)!/n! = n_1n_2/\binom{n}{2}$, por lo tanto

$$E(I_k) = E(I_k^2) = \frac{2n_1n_2}{n(n-1)},$$

de donde

$$E(R) = 1 + \frac{2n_1n_2}{n_1+n_2}.$$

De forma análoga se puede demostrar que

$$Var(R) = \frac{2n_1n_2(2n_1n_2 - n)}{n^2(n-1)}.$$

Distribución asintótica

Para el caso en que n es muy grande, las cuentas para la distribución de R se vuelven muy engorrosas, para el caso en que la hipótesis nula es cierta, se puede usar la siguiente aproximacion: Si suponemos que $\lambda = n_1/n$ y $1 - \lambda = n_2/n$ se mantienen constantes, se puede demostrar que

$$Z_n = \frac{R - 2n\lambda(1-\lambda)}{2\sqrt{n}\lambda(1-\lambda)},$$

tiende en distribución a una $N(0,1)$, en virtud de eso, rechazamos la hipótesis nula si

$$\left| \frac{R - 2n\lambda(1-\lambda)}{2\sqrt{n}\lambda(1-\lambda)} \right| \geq z_{\alpha/2}.$$

8.3. Test de Rachas de subidas y bajadas

Supongamos que tenemos variables aleatorias $X_1, X_2, \ldots, X_n$ y queremos chequear si son i.i.d, para eso definimos las variables $Y_1, \ldots, Y_{n-1}$ de la siguiente forma

$$Y_i = \mathbb{I}_{\{X_i < X_{i+1}\}}.$$

Lo que haremos es estudiar el número de rachas totales de $Y_1, \ldots, Y_{n-1}$ o lo que es lo mismo

$$R = 1 + \sum_{i=1}^{n-2} \mathbb{I}_{\{Y_i \neq Y_{i+1}\}}.$$

Si n es chico, al igual que antes, la distribución de R esta tabulada, y rechazamos la hipótesis de ser i.i.d., a nivel α, si el valor observado $\hat{R}$ cumple que $|\hat{R}| > R_{\alpha/2}$, para valores grandes de n se cumple que

$$Z_n = \frac{R - (2n-1)/3}{\sqrt{\frac{16n-29}{90}}}$$

converge en distribución a una variable con distribución normal con media 0 y varianza 1.

8.4. Test de Spearman

Consideremos una variable aleatoria X y una muestra $X_1, \ldots, X_n$ de ella, a partir de la cual se puede construir el estadístico ordenado $X_{(1)}, \ldots, X_{(n)}$, y el estadístico de rangos $R_1, \ldots, R_n$ donde

$$R_i = \sum_{j=1}^{n} \mathbb{I}_{\{X_j \leq X_i\}}.$$

Para visualizar lo que estamos haciendo, consideremos $X_1; X_2; X_3; X_4 = 1{,}3; 7{,}4; 6{,}2; 2{,}3$, en este caso la muestra ordenada es $1{,}3; 2{,}3; 6{,}2; 7{,}4$, y el el estadistico de rancos es $1; 4; 3; 2$. Llamemos ρ_s al coeficiente de correlacion entre el vector $P = (1, 2, \ldots, n)$ y el vector formado por los rangos, $R = (R_1, \ldots, R_n)$. ρ_s se denomina coeficiente de correlación de rangos de Spearman. Teniendo en cuenta que la media y varianza de P estan fijas y valen $(n+1)/2$ y $(n^2-1)/12$ y que además coinciden con la media y varianza de los rangos $(n+1)/2$ y $(n^2-1)/12$ y que además coinciden con la media y varianza de los rangos, ya que el vector R es una permutación de P. Por lo tanto

$$\rho_s = \frac{\frac{\sum_{i=1}^{n} iR_i}{n} - \left(\frac{n+1}{2}\right)^2}{\frac{n^2-1}{12}} = 1 - \frac{6D}{n(n^2-1)},$$

con $D = \sum_{i=1}^{n}(R_i - i)^2$. Bajo la hipótesis de que la muestra es aleatoria simple, lo valores de X podrían estar ordenados de cualquier forma posible, con la misma probabilidad, es decir, todas las permutacioens de los valores de X serían igualmenten probables, en consecuencia la variable R toma cualquier valor entre 1 y n con la misma probabildad, con éste dato, se puede calcular la distribución de ρ_s bajo la hipótesis nula. De esta forma se obtiene que:

- ρ_s es una variable discreta que tiene una distribución simétrica entre -1 y 1.
- $E(\rho_s)$=0.
- $Var(\rho_s) = 1/(n-1)$.

Definimos la región crítica $RC = \{(X_1, \ldots, X_n)/|\rho_s| > c\}$. Para valores de n menores que 20 existen tablas con la distribución de ρ_s, para muestras de tamaño grande se cumple que $\sqrt{n-1}\rho_s$ converge en distribución a una variable $N(0, 1)$.

Bibliografía

[1] Lehmann, E.L, Casella, G. Theory of Point Estimation. Springer.

[2] Borokov, A.A. (1988). Estadística Matemática, Editorial Mir, Moscuú.

[3] Peña, Daniel. (2001). Fundamentos de estadística. Alianza.

Índice alfabético