Apuntes de Inferencia Estadística

Jose Antonio Lorencio Abril

1 Muestreo Aleatorio Simple

1.1 Población y muestra. Estadísticos. Distribución en el muestreo.

Dada una variable aleatoria X, cada observación de esta nos dará lugar a un dato y, en general, realizadas n observaciones tendremos n datos $(x_1, ..., x_n)$, que denominaremos **muestra de $\boldsymbol{X}$**. Fijado n (**tamaño de la muestra**) podemos obtener distintas colecciones de observaciones. A la variable X se le llama **población**.

Definition 1.1. Sea X una variable aleatoria, se dice que el vector aleatorio de dimensión n, $(X_1, ..., X_n)$, es una **muestra aleatoria simple (m.a.s.) de tamaño $\boldsymbol{n}$ de $\boldsymbol{X}$** si para cada variable X_j esta sigue la misma distribución que X y las variables X_j son independientes entre sí.

Dada una m.a.s. $(X_1, ..., X_n)$ de una variable X con función de distribución F, su **función de distribución conjunta** viene dada por

$$F(x_1, ..., x_n) = \prod_{j=1}^{n} F(x_j)$$

Si X es discreta con función puntual de probabilidad $p(x)$, la **función puntual de probabilidad conjunta** de $(X_1, ..., X_n)$ es

$$p(x_1, ..., x_n) = \prod_{j=1}^{n} p(x_i)$$

Y si X es continua con función de desnidad $f(x)$, la **función de densidad conjunta**

$$f(x_1, ..., x_n) = \prod_{j=1}^{n} f(x_j)$$

Definition 1.2. Sea $(X_1, ..., X_n)$ una m.a.s. de X, llamaremos **estadístico basado en $\boldsymbol{(X_1, ..., X_n)}$** a cualquier vector aleatorio k-dimensional $h(X_1, ..., X_n)$, donde $h : \mathbb{R}^n \to \mathbb{R}^k$ es una función medible Borel.

Como variable, una cuestión fundamental es estudiar la distribución del estadístico. No existe una metodología general para obtener la distribución del estadístico, pero en el caso en que el estadístico se puede expresear como suma de variables aleatorias destaca el método de la función característica.

1.1.1 Método de la función característica

Recordemos que en el caso de variables aleatorias independientes, se verifica

$$\phi_{\sum_{j=1}^{n} X_j}(t) = \prod_{j=1}^{n} \phi_{X_j}(t)$$

Y dado que la función característica determina unívocamente la función de distribución, si determinamos $\phi_{\sum_{j=1}^{n} X_j}(t)$ determinamos la función distribución de la suma.

1.2 Estudio de algunos estadísticos de interés

Definition 1.3. Media muestral

$$\overline{X} = \frac{\sum_{j=1}^{n} X_j}{n}$$

Proposition 1.4.

$$E\left[\overline{X}\right] = E\left[X\right]$$

$$Var\left[\overline{X}\right] = \frac{Var\left[X\right]}{n}$$

Proof.

$$E\left[\overline{X}\right] = E\left[\frac{\sum_{j=1}^{n} X_j}{n}\right] = \frac{1}{n}\sum_{j=1}^{n} E\left[X_j\right] = \frac{1}{n}\sum_{j=1}^{n} E\left[X\right] = \frac{1}{n} n \cdot E\left[X\right] = E\left[X\right]$$

$$Var\left[\overline{X}\right] = Var\left[\frac{\sum_{j=1}^{n} X_j}{n}\right] = \frac{1}{n^2}\sum_{j=1}^{n} Var\left[X_j\right] = \frac{nVar\left[X\right]}{n^2} = \frac{Var\left[X\right]}{n}$$

□

Lemma 1.5. ***Desigualdad de Tchebychev***

$$P\left(\left|Z - E\left[Z\right]\right| \geq \varepsilon\right) \leq \frac{Var\left[Z\right]}{\varepsilon^2}, \forall \varepsilon > 0$$

Lemma 1.6. ***Teorema Central del Límite***
Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias, independientes e idénticamente distribuidas con $E\left[X_i\right] = \mu$ y $Var\left[X_i\right] = \sigma^2$ finitas, entonces

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}} \to_d N\left(0, 1\right)$$

Proposition 1.7.

$$\overline{X} \to_p E\left[X\right]$$

$$\frac{\overline{X} - E\left[X\right]}{\sqrt{\frac{Var\left[X\right]}{n}}} \to_d N\left(0, 1\right)$$

Proof. Consideremos $\left\{\overline{X_n}\right\}_n$ la sucesión de medias muestrales de $X_1, ..., X_n$ y la distribución degenerada $E[X]$. Queremos ver que $\forall\varepsilon > 0, \lim_n P\left(\left|\overline{X_n} - E[X]\right| < \varepsilon\right) = 1$:

$$P\left[\left|\overline{X_n} - E[X]\right| < \varepsilon\right] = 1 - P\left(\left|\overline{X_n} - E[X]\right| \geq \varepsilon\right) \overset{E[\overline{X}]=E[X]}{=} 1 - P\left(\left|\overline{X_n} - E\left[\overline{X_n}\right]\right| \geq \varepsilon\right) \geq$$

$$\overset{Desigualdad\ Tchebychev}{\geq} 1 - \frac{Var[X]}{n\varepsilon^2} = 1 - \frac{1}{\varepsilon^2} Var\left[\overline{X_n}\right]$$

y tomando límites es

$$1 \geq \lim_n P\left(\left|\overline{X_n} - E[X]\right| < \varepsilon\right) \geq \lim_n \left(1 - \frac{Var[X]}{n\varepsilon^2}\right) = 1$$

La segunda afirmación es consecuencia del teorema central del límite, pues es

$$\frac{\overline{X} - E[X]}{\sqrt{\frac{Var[X]}{n}}} = \frac{\frac{1}{n}\sum_{i=1}^n X_i - E[X]}{\sqrt{\frac{Var[X]}{n}}} = \frac{\sum_i^n X_i - nE[X]}{n\sqrt{\frac{Var[X]}{n}}} = \frac{\sum_i^n X_i - nE[X]}{\sqrt{nVar[X]}} \overset{TCL}{\to_d} N(0,1)$$

□

Definition 1.8. Proporción muestral
Sea X una variable aleatoria con distribución Bernoulli, es decir, toma los valores 0 y 1 con probabilidades $P(X=1) = p$ y $P(X=0) = 1-p$ y sea $(X_1, ..., X_n)$ una m.a.s. de X, se llama **proporción muestral** al estadístico

$$\hat{p} = \frac{\sum_{j=1}^n X_j}{n}$$

Obsérvese que $\sum_{j=1}^n X_j$ sigue una distribución binomial $B(n,p)$.

Remark 1.9. Observar que en este caso $\overline{X} = \hat{p}$, por lo que los resultados anteriores para la media muestral son aplicables para la proporción muestral con los correspondientes parámetros.

Proposition 1.10. ***Propiedades de la proporción muestral***

1. $\sum_{j=1}^n X_j = n\hat{p} \sim B(n,p)$
2. $E[\hat{p}] = p$
3. $Var[\hat{p}] = \frac{p(1-p)}{n}$
4. $\hat{p} \to_p p$
5. $\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \to_d N(0,1)$

Definition 1.11. Función de distribución empírica
Fijado $x \in \mathbb{R}$, se llama **función de distribución empírica** en x al estadístico

$$F_n(x) = \frac{1}{n} \sum_{j=1}^{n} I_{(-\infty,x]}(X_j)$$

donde observamos que el numerador contabiliza el número de observaciones de la muestra menores o iguales que x.
Además, si pensamos en el suceso $A = (X \leq x)$, y $p = F(x) = P(X \leq x) = P(A)$, entonces $F_n(x)$ puede verse como la proporción muestral de la variable aleatoria $b(p)$.

Proposition 1.12. ***Propiedades***
Sea $x \in \mathbb{R}$ fijo, entonces

1. $nF_n(x) \sim B(n, p = F(x))$
2. $E[F_n(x)] = F(x)$
3. $Var[F_n(x)] = \frac{F(x)(1-F(x))}{n}$
4. $F_n(x) \to_p F(x)$
5. $\frac{F_n(x)-F(x)}{\sqrt{\frac{F(x)(1-F(x))}{n}}} \to_d N(0,1)$

Theorem 1.13. ***Glivenko-Cantelli***
F_n converge uniformemente a F, es decir

$$\lim_n P\left(\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| > \varepsilon\right) = 0$$

para todo $\varepsilon > 0$.

Definition 1.14. Varianza muestral

$$s^2 = \frac{\sum_{j=1}^{n} \left(X_j - \overline{X}\right)^2}{n} = \frac{\sum_{j=1}^{n} X_j^2}{n} - \left(\overline{X}\right)^2$$

Cuasi-varianza muestral

$$S^2 = \frac{\sum_{j=1}^{n} \left(X_j - \overline{X}\right)^2}{n-1} = \frac{n}{n-1} s^2$$

Proposition 1.15. ***Propiedades***

1. $E\left[s^2\right] = \frac{n-1}{n} Var[X]$
2. $E\left[S^2\right] = Var[X]$

Proof.

$$E\left[s^2\right]=E\left[\frac{\sum_{j=1}^{n}X_j^2}{n}-\left(\overline{X}\right)^2\right]=\frac{1}{n}\sum_{j=1}^{n}E\left[X_j^2\right]-E\left[\left(\overline{X}\right)^2\right]=\frac{1}{n}\sum_{j=1}^{n}E\left[X^2\right]-\left(Var\left[\overline{X}\right]+E\left[\overline{X}\right]^2\right)=$$

$$=E\left[X^2\right]-\frac{Var\left[X\right]}{n}-E\left[X\right]^2=Var\left[X\right]+\cancel{E\left[X\right]^2}-\frac{Var\left[X\right]}{n}-\cancel{E\left[X\right]^2}=\frac{nVar\left[X\right]-Var\left[X\right]}{n}=\frac{n-1}{n}Var\left[X\right]$$

$$E\left[S^2\right]=\frac{n-1}{n}E\left[s^2\right]=Var\left[X\right]$$

□

1.3 Distribución en el muestreo en poblaciones normales. Teorema de Fisher

La distribución χ^2: sean $Z_1,...,Z_d$ con $Z_i\sim N\left(0,1\right)$, independientes entre sí. Entonces

$$\sum_{j=1}^{d}Z_i^2\sim\chi_d^2$$

La distribución t de Student: sean $U\sim N\left(0,1\right)$ y $V\sim\chi_d^2$, independientes, entonces

$$\frac{U}{\sqrt{\frac{V}{d}}}\sim t_d$$

La distribución F de Snédecor: sean $U_1\sim\chi_{d_1}^2$ y $U_2\sim\chi_{d_2}^2$, independientes. Entonces

$$\frac{U_1/d_1}{U_2/d_2}\sim F_{d_1,d_2}$$

Lemma 1.16. *Sea $X=\left(X_1,...,X_n\right)$ un vector aleatorio, con $X\sim N_n\left(\mu,V\right)$, entonces $E\left[X_i\right]=\mu_i,\forall i=1,...,n$ y $Cov\left(X\right)=V$.*

Lemma 1.17. *Sea $X=\left(X_1,...,X_n\right)$un vector aleatorio, con $X\sim N_n\left(\mu,V\right)$ y sean $A\in\mathcal{M}_{m\times n},B\in\mathcal{M}_{m\times 1}$, e $Y=AX+B$. Entonces $Y\sim N_m\left(A\mu+B,AVA^T\right)$*

Theorem 1.18. *Fisher*
Sea $\left(X_1,...,X_n\right)$ una m.a.s. de $X\sim N\left(\mu,\sigma^2\right)$, entonces

1. $\overline{X}\sim N\left(\mu,\frac{\sigma^2}{n}\right)$
2. $\frac{n-1}{\sigma^2}S^2\sim\chi_{n-1}^2$
3. $\overline{X}$ *y* S^2 *son variables aleatorias independientes entre sí*

Proof. Hagamos las tres afirmaciones:

1. Sean $Y_j=\frac{X_j-\mu}{\sigma}\sim N\left(0,1\right)$ que son independientes entre sí. Sea $\mathbb{Y}=\begin{pmatrix}Y_1\\ \vdots\\ Y_n\end{pmatrix}\overset{1.16}{\sim}N_n\left(0_n,I_n\right)$

Sea

$$C=\begin{pmatrix}
\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & & & & \cdots & 0\\
\frac{1}{\sqrt{2\cdot 3}} & \frac{1}{\sqrt{2\cdot 3}} & -\frac{2}{\sqrt{2\cdot 3}} & 0 & & & \cdots & 0\\
\vdots & \vdots & \ddots & \ddots & & & & \cdots\\
\frac{1}{\sqrt{i(i+1)}} & \frac{1}{\sqrt{i(i+1)}} & \cdots & \frac{1}{\sqrt{i(i+1)}} & -\frac{i}{\sqrt{i(i+1)}} & 0 & \cdots & 0\\
\vdots & \vdots & & & \ddots & & & \vdots\\
\frac{1}{\sqrt{(n-1)n}} & \frac{1}{\sqrt{(n-1)n}} & \cdots & & \cdots & & \frac{1}{\sqrt{(n-1)n}} & -\frac{n-1}{\sqrt{(n-1)n}}\\
\frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & & & \cdots & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}}
\end{pmatrix}$$

Veamos que C es ortogonal:

- Para $i<n$, se tiene
$$C_{i,\cdot}\times C_{\cdot,i}^T:\frac{i}{i\left(i+1\right)}+\frac{i^2}{i\left(i+1\right)}=1$$
- $$C_{n,\cdot}\times C_{\cdot,n}^T:n\frac{1}{n}=1$$
- Para $i<j<n$:
$$C_{i,\cdot}\times C_{\cdot,j}:\frac{i}{\sqrt{i\left(i+1\right)}\sqrt{j\left(j+1\right)}}-\frac{i}{\sqrt{i\left(i+1\right)}\sqrt{j\left(j+1\right)}}=0$$
- $$C_{i,\cdot}\times C_{\cdot,n}:\frac{i}{\sqrt{i\left(i+1\right)}\sqrt{n}}-\frac{i}{\sqrt{n}\sqrt{i\left(i+1\right)}}=0$$

Por tanto, $CC^T=I_n$ y C es ortogonal.

A partir de C, construimos el vector

$$\mathbb{Z}=\begin{pmatrix}Z_1\\ \vdots\\ Z_n\end{pmatrix}=C\mathbb{Y}\overset{1.17}{\sim}N_n\left(0,I_n\right)$$

y se tiene

$$Z_n=\sum_{j=1}^n\frac{Y_j}{\sqrt{n}}=\sum_{j=1}^n\frac{X_j-\mu}{\sigma\sqrt{n}}=\frac{\sum_{j=1}^n X_j-n\mu}{\sigma\sqrt{n}}=\frac{\sum_{j=1}^n\frac{X_j}{n}-\mu}{\sigma\frac{\sqrt{n}}{n}}=\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\overset{Z_n\sim N(0,1)}{\sim}N\left(0,1\right)$$

por tanto

$$\overline{X}\sim N\left(\mu,\frac{\sigma^2}{n}\right)$$

2.

$$\frac{n-1}{\sigma^2}S^2=\frac{\cancel{n-1}}{\sigma^2}\sum_{j=1}^n\frac{\left(X_j-\overline{X}\right)^2}{\cancel{n-1}}=\sum_{j=1}^n\left(\frac{X_j-\overline{X}}{\sigma}\right)^2=\sum_{j=1}^n\left(\frac{X_j-\mu+\mu-\overline{X}}{\sigma}\right)^2=$$

$$=\sum_{j=1}^n\left(\frac{X_j-\mu}{\sigma}-\frac{\overline{X}-\mu}{\sigma}\right)^2\overset{*}{=}\sum_{j=1}^n\left(Y_j-\overline{Y}\right)^2=n\sum_{j=1}^n\frac{\left(Y_j-\overline{Y}\right)^2}{n}=ns_Y^2=n\left(\sum_{j=1}^n\frac{Y_j^2}{n}-\overline{Y}^2\right)=$$

$$= \sum_{j=1}^{n} Y_j^2 - n\overline{Y}^2 \overset{**}{=} \mathbb{Y}^T\mathbb{Y} - Z_n^2 \overset{***}{=} \mathbb{Z}^T C \cdot C^T \mathbb{Z} - Z_n^2 = \sum_{j=1}^{n} Z_j^2 - Z_n^2 = \sum_{j=1}^{n-1} Z_j^2 \sim \chi^2_{n-1}$$

donde $*$ se debe a que $Y_j = \frac{X_j - \mu}{\sigma}$ e $\overline{Y} = \sum_{j=1}^n \frac{Y_j}{n} = \sum_{j=1}^n \frac{X_j-\mu}{n\sigma} = \frac{1}{\sigma}\left(\sum_{j=1}^n \frac{X_j}{n} - \frac{n\mu}{n}\right) = \frac{\overline{X}-\mu}{\sigma}$; $**$ se debe a que $\mathbb{Y}^T\mathbb{Y} = \sum_{j=1}^n Y_j^2$ y $n\overline{Y}^2 = n\left(\frac{\overline{X}-\mu}{\sigma}\right)^2 = Z_n^2$; y $***$ se debe a que $\mathbb{Z} = C\mathbb{Y} \implies \mathbb{Y} = C^T\mathbb{Z}$.

3. Notemos que $\overline{X} = g_1(Z_n)$ y $S^2 = g_2(Z_1, ..., Z_{n-1})$ con g_1, g_2 medibles Borel, y los Z_j son independientes entre sí. Por tanto, $g_1(Z_n)$ y $g_2(Z_1, ..., Z_{n-1})$ son independientes.

□

Corollary 1.19. ***Consecuencias del Teorema de Fisher***

1. $\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$

2. *Sea* $(X_1, ..., X_{n_1})$ *una m.a.s. de* $X \sim N(\mu_1, \sigma_1^2)$ *e* $(Y_1, ..., Y_{n_2})$ *una m.a.s. de* $Y \sim N(\mu_2, \sigma_2^2)$*, con* X *e* Y *independientes, entonces*

$$\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{n_1-1}{\sigma_1^2}S_X^2 + \frac{n_2-1}{\sigma_2^2}S_Y^2\right)\frac{\sigma_1^2/n_1 + \sigma_2^2/n_2}{n_1+n_2-2}}} \sim t_{n_1+n_2-2}$$

3. $\frac{S_X^2/\sigma_1^2}{S_Y^2/\sigma_2^2} \sim F_{n_1-1, n_2-1}$

Proof. Veamos cada consecuencia:

1. Por (1) de Fisher tenemos que $U = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ y por (2) que $V = \frac{n-1}{\sigma^2}S^2 \sim \chi^2_{n-1}$ y son independientes. Por tanto el cociente $\frac{U}{\sqrt{\frac{V}{d}}} \sim t_d$, es decir

$$\frac{\frac{\overline{X}-\mu}{\cancel{\sigma}/\sqrt{n}}}{\sqrt{\frac{\cancel{n-1}}{\cancel{\sigma^2}}\frac{S^2}{\cancel{n-1}}}} = \frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$

2. Se tiene que $\overline{X} \sim N(\mu_1, \sigma_1^2/n_1)$ e $\overline{Y} \sim N(\mu_2, \sigma_2^2/n_2)$ y son independientes, por lo que

$$\overline{X} - \overline{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

y entonces

$$U = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

Por otro lado, es $\frac{n_1-1}{\sigma_1^2}S_X^2 \sim \chi^2_{n_1-1}$ y $\frac{n_2-1}{\sigma_2^2}S_Y^2 \sim \chi^2_{n_2-1}$ (esto es, ambas son sumas cuadráticas de variables aleatorias de distribución normal e independientes, y son independientes entre sí, por lo que su suma sigue siendo una suma de este tipo), luego

$$V = \frac{n_1-1}{\sigma_1^2}S_X^2 + \frac{n_2-1}{\sigma_2^2}S_Y^2 \sim \chi_{n_1+n_2-2}$$

por tanto $\frac{U}{\sqrt{\frac{V}{n_1+n_2-2}}} \sim t_{n_1+n_2-2}$ y esto es lo que queríamos ver.

3. Como $\frac{n_1-1}{\sigma_1^2} S_X^2 \sim \chi_{n_1-1}^2$ y $\frac{n_2-1}{\sigma_2^2} S_Y^2 \sim \chi_{n_2-1}^2$, se tiene de forma inmediata.

$\square$

1.4 Distribución en el muestreo de datos ordenados

Definition 1.20. Valores máximos y mínimos de una muestra
Sea $(X_1, ..., X_n)$ una m.a.s. de una variable X, se llama **mínimo de la muestra** al estadístico

$$X_{1:n} = \min \{X_1, ..., X_n\}$$

y **máximo de la muestra** al estadístico

$$X_{n:n} = \max \{X_1, ..., X_n\}$$

Proposition 1.21. *Sea $(X_1, ..., X_n)$ una m.a.s. de una variable X con distribución F, la distribución en el muestre de $X_{1:n}$ viene dada por*

$$P(X_{1:n} \leq x) = 1 - (1 - F(x))^n, \forall x \in \mathbb{R}$$

y la de $X_{n:n}$ viene dada por

$$P(X_{n:n} \leq x) = (F(x))^n, \forall x \in \mathbb{R}$$

Proof.

$$P(X_{n:n} \leq x) = P(\max \{X_1, ..., X_n\} \leq x) = P\left(\bigcap_{j=1}^{n} \{X_j \leq x\}\right) = \prod_{j=1}^{n} P(X_j \leq x) = \prod_{j=1}^{n} F(x) = F(x)^n$$

$$P(X_{1:n} \leq x) = 1 - P(X_{1:n} > x) = 1 - P(\min \{X_1, ..., X_n\} > x) = 1 - P\left(\bigcap_{j=1}^{n} \{X_j > x\}\right) =$$

$$= 1 - \prod_{j=1}^{n} P(X_j > x) = 1 - \prod_{j=1}^{n} (1 - P(X_j \leq x)) = 1 - (1 - F(x))^n$$

$\square$

1.5 Ejercicios

1. Sea $X \sim b(p)$ y sea (X_1, X_2, X_3) una m.a.s. de X. Se pide:

1. Estudiar la distribución del vector (X_1, X_2, X_3)

   Se tiene que $ImX = \{0, 1\}$ con $P[X = 1] = p$ y $P[X = 0] = 1 - p$ con $0 < p < 1$. Buscamos su función de probabilidad conjunta, para ello podemos dar la probabilidad de cada punto:

	$(0,0,0)$	$(0,0,1)$	$(0,1,0)$	$(0,1,1)$	$(1,0,0)$	$(1,0,1)$	$(1,1,0)$	$(1,1,1)$
P	$(1-p)^3$	$(1-p)^2 p$	$(1-p)^2 p$	$(1-p) p^2$	$p(1-p)^2$	$p^2(1-p)$	$p^2(1-p)$	p^3

2. Estudiar la distribución en el muestreo del estadístico $\frac{X_1+X_2+X_3}{3}$

	0	$\frac{1}{3}$	$\frac{2}{3}$	1
P	$(1-p)^3$	$3p(1-p)^2$	$3p^2(1-p)$	p^3

2. Sea $X \sim Exp(\alpha)$. Sea $(X_1, ..., X_n)$ una m.a.s. de X, estudiar la distribución en el muestreo de $S = \sum_{j=1}^n X_j$

Vamos a usar el método de la función característica:

$$\phi_S(t) = \prod_{j=1}^n \phi_{X_j}(t) = \prod_{j=1}^n \left(1 - \frac{it}{\alpha}\right)^{-1} = \left(1 - \frac{it}{\alpha}\right)^{-n}$$

que es la función característica de una $\gamma(\alpha, n)$. Por tanto, $S \sim \gamma(\alpha, n)$ y

$$f_S(x) = \frac{\alpha^n}{\Gamma(n)} x^{n-1} e^{-\alpha x}$$

Además

$$E\left[\overline{X}\right] = \frac{1}{n} E[S] = \frac{1}{\alpha} = E[X]$$

$$Var\left[\overline{X}\right] = \frac{1}{n^2} Var(S) = \frac{1}{n\alpha^2}$$

3. Sea $X \sim N(\mu, \sigma^2)$ y $(X_1, ..., X_n)$ una m.a.s. de X. Estudiar la distribución en el muestreo de $S = \sum_{j=1}^n X_j$.

$$\phi_S(t) = \prod_{j=1}^n \phi_{X_j}(t) = \prod_{j=1}^n e^{it\mu - \frac{t^2\sigma^2}{2}} = e^{it\mu n - \frac{t^2\sigma^2}{2} n}$$

que es la característica de una $N(\mu n, \sigma^2 n)$.

$$E\left[\overline{X}\right] = \frac{1}{n} E[S] = \frac{1}{n} n\mu = \mu$$

$$Var\left(\overline{X}\right) = \frac{1}{n^2} Var(S) = \frac{1}{n^2} \sigma^2 n = \frac{\sigma^2}{n}$$

4. Sea X una variable aleatoria con función de densidad

$$f(x, \theta) = \frac{2x}{\theta} \exp\left(-\frac{x^2}{\theta}\right) \chi_{(0,\infty)}(x)$$

Obtener la distribución en el muestreo del estadístico

$$T(X_1, .., X_n) = \sum_{j=1}^n X_j^2$$

Obtener su media y su varianza.

Sea $Y = X^2$, de forma que

$$f_Y(y) = f_X(x(y)) \left|x'(y)\right| = f_X(\sqrt{y}) \frac{1}{2\sqrt{y}} = \frac{\cancel{2\sqrt{y}}}{\theta} \exp\left(-\frac{y}{\theta}\right) \chi_{(0,\infty)}(\sqrt{y}) \frac{1}{\cancel{2\sqrt{y}}} = \frac{e^{-\frac{y}{\theta}}}{\theta} \chi_{(0,\infty)}(y)$$

que es la función de densidad de una $Exp\left(\frac{1}{\theta}\right)$. Por tanto

$$T=\sum_j Y_j \overset{Ej2}{\sim} \gamma\left(\frac{1}{\theta},n\right)$$

Y entonces

$$E\left(T\right)=n\theta$$

$$Var\left(T\right)=n\theta^2$$

5. Sea $X\sim U\left(0,1\right)$. Sea $\left(X_1,X_2\right)$ una m.a.s. de X. Determinar la distribución en el muestreo de la media muestral.
Sea $S=X_1+X_2$. Para obtener la función de densidad de S, debemos hacer el producto de convolución de las funciones de X_1 y X_2:

$$f_S\left(s\right)=\int_{-\infty}^{\infty} f_{X_1}\left(y\right) f_{X_2}\left(s-y\right)dy=\int_0^1 f_{X_1}\left(y\right) f_{X_2}\left(s-y\right)dy=\int_0^1 f_{X_2}\left(s-y\right)dy$$

Y distinguimos dos casos:

- $s\in\left(0,1\right)$:

$$f_S\left(s\right)=\int_0^s f_{X_2}\left(s-y\right)dy=\int_0^s dy=s$$

- $s\in\left(1,2\right)$:

$$f_S\left(s\right)=\int_{s-1}^1 dy=1-\left(s-1\right)=2-s$$

Por tanto

$$f_S\left(s\right)=\begin{cases} s & s\in\left(0,1\right)\\ 2-s & s\in\left(1,2\right)\\ 0 & resto\end{cases}$$

Y buscamos la distribución de $T=\frac{S}{2}$ (se tiene $S=2T, S'=2$), que será

$$f_T\left(t\right)=\begin{cases} 4t & t\in\left(0,\frac{1}{2}\right)\\ \left(2-2t\right)2=4-4t & t\in\left(\frac{1}{2}.1\right)\\ 0 & resto\end{cases}$$

6. Sea X una variable aleatoria con función de densidad

$$f\left(x,\theta\right)=\frac{\theta}{\left(1+x\right)^{1+\theta}}\chi_{\left(0,\infty\right)}\left(x\right)$$

Obtener la distribución en el muestreo del estadístico

$$T\left(X_1,...,X_n\right)=\frac{\sum_{j=1}^n \log\left(1+X_j\right)}{n}$$

Obtener su media y su varianza.
Primero observamos que $\log\left(1+x\right):\left(0,\infty\right)\overset{biy}{\to}\left(0,\infty\right)$, y definimos $Y=\log\left(1+X\right)$, de forma que $X=e^Y-1$ y $X'=e^Y$. Entonces

$$f_Y\left(y\right)=f_X\left(x\left(y\right)\right)\left|x'\left(y\right)\right|=\frac{\theta}{e^{y\left(\cancel{1}+\theta\right)}}\chi_{\left(0,\infty\right)}\left(e^y\right)\cancel{e^y}=\frac{\theta}{e^{y\theta}}$$

que es la función de densidad de una $Exp(\theta)$.
Ahora usamos el método de la función característica a $S = \sum_{j=1}^{n} Y_j$:

$$\phi_S(t) = \prod_{j=1}^{n} \phi_{X_j}(t) = \left(1 - \frac{it}{\theta}\right)^{-n}$$

que es la característica de una $\gamma(\theta, n)$.
Por último, tenemos $T = \frac{S}{n}$, y es $S = nt$ y $S' = n$. Luego

$$f_T(t) = f_S(nt)\, n = \frac{\theta^n}{\Gamma(n)} (nt)^{n-1} e^{-\theta n t} n = \frac{(\theta n)^n}{\Gamma(n)} t^{n-1} e^{-(\theta n)t}$$

que es la densidad de una $\gamma(\theta n.n)$.
Es decir, $T \sim \gamma(\theta n, n)$ y

$$E(T) = \frac{n}{n\theta} = \frac{1}{\theta}$$

$$Var(T) = \frac{n}{(n\theta)^2} = \frac{1}{n\theta^2}$$

7. Sea X una variable aleatoria con función de densidad

$$f(x, \theta) = \exp(-(x-\theta)) \cdot \exp(-\exp(-(x-\theta)))$$

Obtener la distribución en el muestreo del estadístico

$$T(X_1, ..., X_n) = \frac{\sum_{j=1}^{n} \exp(-X_j)}{n}$$

Obtener su media y su varianza.
Sea $Y = \exp(-X)$, entonces $X = -\log(Y)$ y $X' = -\frac{1}{Y}$. Además $\mathbb{R} \overset{\exp}{\to} (0, \infty)$. Entonces es

$$f_Y(y) = f_X(-\log(y)) \frac{1}{y} = e^{\log y + \theta} e^{-e^{\log y + \theta}} \frac{1}{y} = \cancel{y} e^{\theta} e^{-ye^{\theta}} \frac{1}{\cancel{y}} = e^{\theta} e^{-ye^{\theta}}$$

que es la densidad de una $Exp\left(e^{\theta}\right)$.
Ahora hacemos $S = \sum_j Y_j$ y aplicamos el método de la función característica, que por el ejercicio 2 nos da que es $S \sim \gamma\left(e^{\theta}, n\right)$.
Por último, tenemos $T = \frac{S}{n}$, luego, como en el ejercicio 6, sale $T \sim \gamma\left(e^{\theta} n, n\right)$.

$$E(T) = \frac{n}{ne^{\theta}} = \frac{1}{e^{\theta}}$$

$$Var(T) = \frac{n}{\left(ne^{\theta}\right)^2} = \frac{1}{ne^{2\theta}}$$

8. Sea $(X_1, ..., X_n)$ una m.a.s. de $X \sim N\left(\mu, \sigma^2\right)$. Sean $\overline{X}$ y S^2 su media y cuasivarianza muestrales, respectivamente. Sea X_{n+1} una nueva observación de X independiente de $(X_1, ..., X_n)$. Obtener la distribución en el muestreo del estadístico

$$\frac{X_{n+1} - \overline{X}}{S} \sqrt{\frac{n}{n+1}}$$

Tenemos que $X_{n+1} \sim N\left(\mu, \sigma^2\right), \overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ y son independientes entre sí.

Por tanto

$$X_{n+1} - \overline{X} \sim N\left(0, \sigma^2 \frac{n+1}{n}\right)$$

por lo que

$$\frac{X_{n+1} - \overline{X}}{\sigma\sqrt{\frac{n+1}{n}}} \sim N(0,1)$$

Por otro lado, $\frac{n-1}{\sigma^2}S^2 \sim \chi^2_{n-1}$. Y entonces

$$\frac{\frac{X_{n+1}-\overline{X}}{\cancel{\sigma}\sqrt{\frac{n+1}{n}}}}{\frac{\cancel{\sqrt{n-1}}}{\cancel{\sigma}}S\frac{1}{\cancel{\sqrt{n-1}}}} = \frac{X_{n+1} - \overline{X}}{S}\sqrt{\frac{n}{n+1}} \sim t_{n-1}$$

9. Sean $(X_1, ..., X_{n_1})$ e $(Y_1, ..., Y_{n_2})$ m.a.s.'s independientes de dos poblaciones $X \sim N\left(\mu_1, \sigma^2\right)$ e $Y \sim N\left(\mu, \sigma^2\right)$, respectivamente. Obtener la distribución en el muestreo del estadístico

$$\frac{\alpha\left(\overline{X} - \mu_1\right) + \beta\left(\overline{Y} - \mu_2\right)}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\frac{\alpha^2}{n_1} + \frac{\beta^2}{n_2}}}$$

con $\alpha, \beta \in \mathbb{R}$.
Se tiene

$$\overline{X} \sim N\left(\mu_1, \frac{\sigma^2}{n_1}\right) \implies \overline{X} - \mu_1 \sim N\left(0, \frac{\sigma^2}{n_1}\right) \implies \alpha\left(\overline{X} - \mu\right) \sim N\left(0, \frac{\alpha^2\sigma^2}{n_1}\right)$$

De igual forma

$$\beta\left(\overline{Y} - \mu_2\right) \sim N\left(0, \frac{\beta^2\sigma^2}{n_2}\right)$$

Entonces, como X e Y son independientes, es

$$\alpha\left(\overline{X} - \mu_1\right) + \beta\left(\overline{Y} - \mu_2\right) \sim N\left(0, \sigma^2\left(\frac{\alpha^2}{n_1} + \frac{\beta^2}{n_2}\right)\right)$$

Por otro lado

$$\frac{n_1 - 1}{\sigma^2}S_1^2 \sim \chi^2_{n_1-1}$$
$$\frac{n_2 - 1}{\sigma^2}S_2^2 \sim \chi^2_{n_2-1}$$

y como son independientes, es

$$\frac{n_1 - 1}{\sigma^2}S_1^2 + \frac{n_2 - 1}{\sigma^2}S_2^2 = \frac{1}{\sigma^2}\left((n_1 - 1)S_1^2 + (n_2 - 1)S_2^2\right) \sim \chi^2_{n_1+n_2-2}$$

Y entonces

$$\frac{\frac{\alpha\left(\overline{X}-\mu_1\right)+\beta\left(\overline{Y}-\mu_2\right)}{\sigma\sqrt{\frac{\alpha^2}{n_1}+\frac{\beta^2}{n_2}}}}{\frac{1}{\sigma}\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}} \sim t_{n_1+n_2-2}$$

2 Estimación Paramétrica Puntual. Cotas para la Varianza de un Estimador

2.1 Introducción

Un fenómeno aleatorio, como variable aleatoria, tendrá una función de distribución F, que dependerá de unos ciertos parámetros, que asumiremos desconocidos y que toman valores dentro de un determinado espacio. Esto se denota $\boldsymbol{X \sim F(x, \theta)}$, donde la forma funcional de esta distribución será conocida, pero sus parámetros no.
Lo que buscamos es, a través de una m.a.s. de esta variable aleatoria, obtener información acerca del parámetro. La **estimación puntual** lo que hace, en general, es obtener un estadístico que, al tomar un valor concreto para una muestra, nos dé un valor lo más aproximado posible a θ. Llamaremos **estimador** a un estadístico utilizado para estimar un parámetro.

2.2 Estimador Insesgado de Mínima Varianza (EIMV). Cota de Cramer-Rao

Dado un estadístico $T = T(X_1, ..., X_n)$, podríamos considerar la diferencia entre este estadístico y el parámetro θ, $T - \theta$. Esta diferencia será aleatoria y una forma de medir la variabilidad de esa aleatoriedad es a través de la expresión

$$E\left[(T-\theta)^2\right]$$

Lo interesante sería encontrar un estimador que hiciese esta expresión lo más pequeña posible, puesto que eso nos indicaría que el promedio de las diferencias entre T y θ es muy pequeño y por lo tanto los valores que toma el estadístico T para las distintas muestras se encuentran muy próximos al parámetro θ.

$$E\left[(T-\theta)^2\right] = Var[T] + (E[T]-\theta)^2$$

pues

$$E\left[(T-\theta)^2\right] = E\left[(T-E[T]+E[T]-\theta)^2\right] =$$

$$= E\left[(T-E[T])^2\right] + 2E[(T-E[T])(E[T]-\theta)] + E\left[(E[T]-\theta)^2\right] =$$

$$\overset{*}{=} Var[T] + (E[T]-\theta)^2$$

Donde $*$ se debe a que $Var[T] = E\left[(T-E[T])^2\right]$, $(E[T]-\theta)$ es una constante, y $E[(T-E[T])(E[T]-\theta)] = (E[T]-\theta)E[T-E[T]] = (E[T]-\theta)(E[T]-E[T]) = 0$.

Definition 2.1. Un estadístico T es un **estimador insesgado (EI) del parámetro $\boldsymbol{\theta}$** si $E[T] = \theta$.
Si $E[T] = \theta + b(\theta)$, $b(\theta)$ se denomina **sesgo del estimador**.
T es un **estimador insesgado de mínima varianza (EIMV) del parámetro $\boldsymbol{\theta}$** si T es EI de θ y $Var[T] \leq Var[T']$ para todo T' EI de θ.

Definition 2.2. Sea $(X_1, ..., X_n)$ una m.a.s. de la variable aleatoria X. Llamaremos **función de verosimilitud** a la función de densidad (o puntual de probabilidad) conjunta de $(X_1, ..., X_n)$, es decir, a

$$L(x_1, ..., x_n, \theta) = \prod_{j=1}^{n} f(x_j, \theta)$$

donde $f(x, \theta)$ es la función de densidad (o puntual de probabilidad) de X.

Theorem 2.3. ***Cota de Cramer-Rao***
Sea $X \sim F(x,\theta), \theta \in \Theta \subset \mathbb{R}^k, (X_1,...,X_n)$ *una m.a.s. de* X *y* $\hat{\theta}$ *un estadístico con* $E\left[\hat{\theta}\right] = \theta + b(\theta)$.
Supongamos que se verifica:

i) *El conjunto* $\Psi = Sop(X_1,...,X_n) = \{(x_1,...,x_n) \in \mathbb{R}^n | L(x_1,...,x_n,\theta) > 0\}$ *no depende de* θ.

ii) *Existe* $\frac{\partial L}{\partial \theta}$ *para todo* θ *y para todo* $(x_1,...,x_n) \in \Psi$

iii) *Tanto*
$$E\left[\hat{\theta}\right] = \theta + b(\theta) = \int ... \int_{\Psi} \hat{\theta}(x_1,...,x_n) \cdot L(x_1,...,x_n,\theta)\, dx_1...dx_n$$
como
$$\int ... \int_{\Psi} L(x_1,...,x_n,\theta)\, dx_1...dx_n$$
son derivables bajo el signo integral respecto de θ.

iv)
$$E\left[\left(\frac{\partial \ln L}{\partial \theta}\right)^2\right]$$
es finita y estrictamente positiva, para todo θ.

Entonces se verifica
$$Var\left(\hat{\theta}\right) \geq \frac{\left(\frac{\partial E(\hat{\theta})}{\partial \theta}\right)^2}{E\left[\left(\frac{\partial \ln L}{\partial \theta}\right)^2\right]} \qquad \textbf{\textit{Cota de Cramer} - \textit{Rao}}$$

La idea de la prueba es calcular el coeficiente de corelación entre $\hat{\theta}$ y $\frac{\partial \ln L}{\partial \theta}$ y si tenemos
$$0 \leq \left(\frac{cov\left[\hat{\theta}, \frac{\partial \ln L}{\partial \theta}\right]}{\sqrt{Var\left[\hat{\theta}\right] Var\left[\frac{\partial \ln L}{\partial \theta}\right]}}\right)^2 \leq 1$$
entonces se dará
$$Var\left[\hat{\theta}\right] \geq \frac{cov\left[\hat{\theta}, \frac{\partial \ln L}{\partial \theta}\right]}{Var\left[\frac{\partial \ln L}{\partial \theta}\right]}$$

Proof. Vamos a calcular todas las cosas mencionadas en la idea:

a) Como $\int ... \int_{\Psi} L(\boldsymbol{x},\theta)\, d\boldsymbol{x} = 1$ donde $\boldsymbol{x} = (x_1,...,x_n)$, entonces
$$0 \overset{i),iii).}{=} \int ... \int_{\Psi} \frac{\partial}{\partial \theta} L(\boldsymbol{x},\theta)\, d\boldsymbol{x} \overset{*}{=} \int ... \int_{\Psi} \left[\frac{\partial}{\partial \theta} \ln L(\boldsymbol{x},\theta)\right] L(\boldsymbol{x},\theta)\, d\boldsymbol{x} = E\left[\frac{\partial}{\partial \theta} \ln L(\boldsymbol{x},\theta)\right]$$
donde $*$ se debe a que $\frac{\partial}{\partial \theta} \ln L(\boldsymbol{x},\theta) \overset{ii)}{=} \frac{\frac{\partial}{\partial \theta} L(\boldsymbol{x},\theta)}{L(\boldsymbol{x},\theta)}$ (1).

b) Por otro lado

$$\frac{\partial}{\partial\theta}E\left[\hat{\theta}\right] \overset{iii)}{=} 1+b'(\theta) \overset{i),iii)}{=} \int \dots \int_{\Psi} \hat{\theta}(\boldsymbol{x}) \cdot \frac{\partial}{\partial\theta}L(\boldsymbol{x},\theta)\, d\boldsymbol{x} \overset{(1)}{=}$$

$$= \int \dots \int_{\Psi} \hat{\theta}(\boldsymbol{x}) \cdot \left[\frac{\partial}{\partial\theta}\ln L(\boldsymbol{x},\theta)\right] \cdot L(\boldsymbol{x},\theta)\, d\boldsymbol{x} = E\left[\hat{\theta} \cdot \frac{\partial}{\partial\theta}\ln L(\boldsymbol{x},\theta)\right]$$

c) Y ahora calculamos el coeficiente de correlación:

$$\rho_{\hat{\theta}, \frac{\partial}{\partial\theta}\ln L(\boldsymbol{x},\theta)} = \frac{cov\left[\hat{\theta}, \frac{\partial \ln L}{\partial\theta}\right]}{\sqrt{Var\left[\hat{\theta}\right] Var\left[\frac{\partial \ln L}{\partial\theta}\right]}} = \frac{E\left[\hat{\theta}\cdot\frac{\partial}{\partial\theta}\ln L(\boldsymbol{x},\theta)\right] - E\left[\hat{\theta}\right] E\left[\frac{\partial}{\partial\theta}\ln L(\boldsymbol{x},\theta)\right]}{\sqrt{Var\left[\hat{\theta}\right]\left(E\left[\left(\frac{\partial}{\partial\theta}\ln L(\boldsymbol{x},\theta)\right)^2\right] - \left(E\left[\frac{\partial}{\partial\theta}\ln L(\boldsymbol{x},\theta)\right]\right)^2\right)}} \overset{iv),a)}{=}$$

$$= \frac{E\left[\hat{\theta}\cdot\frac{\partial}{\partial\theta}\ln L(\boldsymbol{x},\theta)\right]}{\sqrt{Var\left[\hat{\theta}\right]\cdot E\left[\left(\frac{\partial}{\partial\theta}\ln L(\boldsymbol{x},\theta)\right)^2\right]}}$$

$$\overset{\rho^2\leq 1}{\Longleftrightarrow} Var\left[\hat{\theta}\right] \overset{b)}{\geq} \frac{\left(\frac{\partial}{\partial\theta}E\left[\hat{\theta}\right]\right)^2}{E\left[\left(\frac{\partial}{\partial\theta}\ln L(\boldsymbol{x},\theta)\right)^2\right]}$$

□

Remark 2.4. Para el desarrollo anterior, no es necesaria la independencia de las variables X_i, aunque trabajaremos con m.a.s. y por tanto serán independientes.

Remark 2.5. Según la notación del resultado, la variable manejada es continua, pero el resultado sirve también cambiando la integración por sumatorios en el caso discreto.

Corollary 2.6. *En las condiciones del teorema de Cramer-Rao, si $\hat{\theta}$ es un estimador insesgado de θ, entonces se tiene que*

$$Var\left[\hat{\theta}\right] \geq \frac{1}{E\left[\left(\frac{\partial}{\partial\theta}\ln L(\boldsymbol{x},\theta)\right)^2\right]}$$

Remark 2.7. El resultado anterior es cierto si tenemos un estimador insesgado, con varianza finita, que verifica la propiedad de que su esperanza es derivable bajo el signo de la integral.

Remark 2.8. **Criterio para estudiar si un estimador es E.I.M.V. de un parámetro**

1. Vemos que sea insesgado
2. Comprobamos que estamos en las condiciones del teorema de la cota de Cramer-Rao, las condiciones $i), ii), iii), iv)$ se denominan **condiciones de regularidad de Cramer-Rao**
3. Calculamos la cota de Cramer-Rao y la varianza del estimador
4. Si esos dos valores coinciden, entonces el estimador será E.I.M.V

Definition 2.9. Se denomina **cantidad de información de Fisher** a

$$I_n(\theta) = E\left[\left(\frac{\partial}{\partial\theta}\ln L(\boldsymbol{x},\theta)\right)^2\right]$$

Remark 2.10. Entonces la cota de Cramer-Rao se puede escribir como

$$Var\left[\hat{\theta}\right] \geq \frac{\left(\frac{\partial}{\partial\theta}E\left[\hat{\theta}\right]\right)^2}{I_n\left(\theta\right)}$$

Proposition 2.11. *Bajo las condiciones de regularidad de Cramer-Rao, dada $(X_1, ..., X_n)$ una m.a.s. de X se tiene que la cantidad de información de Fisher se puede escribir como*

$$I_n\left(\theta\right) = nE\left[\left(\frac{\partial}{\partial\theta}\ln f\left(x,\theta\right)\right)^2\right]$$

siendo $f\left(x,\theta\right)$ la función de densidad o función puntual de probabilidad de X.

Proof. Primero, notamos que

$$L\left(\boldsymbol{x},\theta\right) = \prod_{j=1}^{n} f\left(x_j,\theta\right) \implies \ln L\left(\boldsymbol{x},\theta\right) = \sum_{j=1}^{n}\ln f\left(x_j,\theta\right)$$

y por tanto

$$\frac{\partial}{\partial\theta}\ln L\left(\boldsymbol{x},\theta\right) = \sum_{j=1}^{n}\frac{\partial}{\partial\theta}\ln f\left(x_j,\theta\right)$$

Y entonces

$$I_n\left(\theta\right) = E\left[\left(\sum_{j=1}^{n}\frac{\partial}{\partial\theta}\ln f\left(x_j,\theta\right)\right)^2\right] = E\left[\sum_{j=1}^{n}\left(\frac{\partial}{\partial\theta}\ln f\left(x_j,\theta\right)\right)^2 + \sum_{i\neq j}\frac{\partial}{\partial\theta}\ln f\left(x_i,\theta\right)\cdot\frac{\partial}{\partial\theta}\ln f\left(x_j,\theta\right)\right] =$$

$$\overset{X_i,X_j\ indep\ \forall i\neq j}{=} \sum_{j=1}^{n}E\left[\left(\frac{\partial}{\partial\theta}\ln f\left(x_j,\theta\right)\right)^2\right] + \sum_{i\neq j}E\left[\frac{\partial}{\partial\theta}\ln f\left(x_i,\theta\right)\right]\cdot E\left[\frac{\partial}{\partial\theta}\ln f\left(x_j,\theta\right)\right]$$

Y aquí tenemos que $\frac{\partial}{\partial\theta}\ln f\left(x_j,\theta\right)$ presenta la misma distribución que $\frac{\partial}{\partial\theta}\ln f\left(x,\theta\right)$, luego $E\left[\left(\frac{\partial}{\partial\theta}\ln f\left(x_j,\theta\right)\right)^2\right] = E\left[\left(\frac{\partial}{\partial\theta}\ln f\left(x,\theta\right)\right)^2\right], \forall j = 1, ..., n$.
Por otro lado, por a) de la demostración de la cota de Cramer-Rao se tiene que $E\left[\frac{\partial}{\partial\theta}\ln f\left(x_j,\theta\right)\right] = 0, \forall j = 1, ..., n$, por lo que los sumandos de la derecha son todos nulos.
Entonces

$$I_n\left(\theta\right) = \sum_{j=1}^{n}E\left[\left(\frac{\partial}{\partial\theta}\ln f\left(x,\theta\right)\right)^2\right] = nE\left[\left(\frac{\partial}{\partial\theta}\ln f\left(x,\theta\right)\right)^2\right]$$

□

Proposition 2.12. *Bajo las condiciones de regularidad de Cramer-Rao, si se verifica que*

$$E\left[\frac{\partial}{\partial\theta}\ln L\left(\boldsymbol{x},\theta\right)\right]$$

es derivable bajo el signo integral respecto de θ, entonces

$$I_n\left(\theta\right)=-E\left[\frac{\partial^2}{\partial\theta^2}\ln L\left(\boldsymbol{x},\theta\right)\right]=-nE\left[\frac{\partial^2}{\partial\theta^2}\ln f\left(x,\theta\right)\right]$$

siendo $f\left(x,\theta\right)$ la función de densidad o función puntual de probabilidad de X.

Proof. Viendo la demostración de la cota de Cramer-Rao, por a) tenemos que $E\left[\frac{\partial}{\partial\theta}\ln L\left(\boldsymbol{x},\theta\right)\right]=0$. Por tanto:

$$0=\frac{\partial}{\partial\theta}E\left[\frac{\partial}{\partial\theta}\ln L\left(\boldsymbol{x},\theta\right)\right]=\int\ldots\int_{\Psi}\frac{\partial}{\partial\theta}\left(\frac{\partial}{\partial\theta}\ln L\left(\boldsymbol{x},\theta\right)\cdot L\left(\boldsymbol{x},\theta\right)\right)d\boldsymbol{x}=$$

$$=\int\ldots\int_{\Psi}\left[\frac{\partial^2}{\partial\theta^2}\ln L\left(\boldsymbol{x},\theta\right)\cdot L\left(\boldsymbol{x},\theta\right)+\frac{\partial}{\partial\theta}\ln L\left(\boldsymbol{x},\theta\right)\cdot\frac{\partial}{\partial\theta}L\left(\boldsymbol{x},\theta\right)\right]d\boldsymbol{x}=$$

$$=E\left[\frac{\partial^2}{\partial\theta^2}\ln L\left(\boldsymbol{x},\theta\right)\right]+\int\ldots\int_{\Psi}\left(\frac{\partial}{\partial\theta}\ln L\left(\boldsymbol{x},\theta\right)\right)^2\cdot L\left(\boldsymbol{x},\theta\right)d\boldsymbol{x}=E\left[\frac{\partial^2}{\partial\theta^2}\ln L\left(\boldsymbol{x},\theta\right)\right]+I_n\left(\theta\right)$$

donde el segundo sumando sale de que $\frac{\partial}{\partial\theta}\ln L\left(\boldsymbol{x},\theta\right)=\frac{\frac{\partial}{\partial\theta}L(\boldsymbol{x},\theta)}{L(\boldsymbol{x},\theta)}$.
Por tanto

$$I_n\left(\theta\right)=-E\left[\frac{\partial^2}{\partial\theta^2}\ln L\left(\boldsymbol{x},\theta\right)\right]$$

Para ver la otra forma:

$$-E\left[\frac{\partial^2}{\partial\theta^2}\ln L\left(\boldsymbol{x},\theta\right)\right]=-E\left[\frac{\partial^2}{\partial\theta^2}\ln\left(\prod_{j=1}^{n}f\left(x_j,\theta\right)\right)\right]=-E\left[\frac{\partial^2}{\partial\theta^2}\sum_{j=1}^{n}\ln f\left(x_j,\theta\right)\right]=$$

$$=-E\left[\sum_{j=1}^{n}\frac{\partial^2}{\partial\theta^2}\ln f\left(x_j,\theta\right)\right]=-\sum_{j=1}^{n}E\left[\frac{\partial^2}{\partial\theta^2}\ln f\left(x_j,\theta\right)\right]==-\sum_{j=1}^{n}E\left[\frac{\partial^2}{\partial\theta^2}\ln f\left(x,\theta\right)\right]=$$

$$=-nE\left[\frac{\partial^2}{\partial\theta^2}\ln f\left(x,\theta\right)\right]$$

□

Definition 2.13. Dado un estadístico $\hat{\theta}$, se dice que es **eficiente-c (ef-c) para el parámetro $\boldsymbol{\theta}$** si su varianza alcanza la cota de Cramer-Rao.

Remark 2.14. Una consecuencia inmediata es que todo estimador insesgado eficiente-c es un EIMV.

Proposition 2.15. *Bajo las condiciones de regularidad de Cramer-Rao y dado un estadístico $\hat{\theta}$, se tiene que $\hat{\theta}$ es ef-c para θ si, y solo si,*

$$L\left(\boldsymbol{x},\theta\right)=\exp\left\{A\left(\theta\right)\hat{\theta}+B\left(\theta\right)+h\left(\boldsymbol{x}\right)\right\}$$

para todo $\boldsymbol{x}\in\Psi$. Donde A,B son funciones solo de θ, y h es función solo de la muestra.

Proof. Suponamos que $X \sim F(x,\theta)$ y que $(X_1, ..., X_n)$ es una m.a.s. de X, y sea $\hat{\theta} = \hat{\theta}(\boldsymbol{X})$. Entonces

$$Var\left[\hat{\theta}\right] \text{ alcanza la cota de Cramer-Rao} \overset{dem\ Cramer-Rao}{\iff} \rho^2_{\hat{\theta}, \frac{\partial}{\partial\theta}\ln L(\boldsymbol{x},\theta)} = 1$$

Esto sucede, por la definición de $\rho^2_{X,Y}$ (si es 1, la relación entre X e Y es lineal: $Y = aX + b$), si, y solo si, fijado $\theta \in \Theta \subset \mathbb{R}$ es

$$\frac{\partial}{\partial\theta}\ln L(\boldsymbol{x},\theta) = a(\theta)\hat{\theta} + b(\theta) \ \ c.s.$$

Y esto se da si, y solo si,

$$\ln L(\boldsymbol{x},\theta) = \int \left[a(\theta)\hat{\theta} + b(\theta)\right] d\theta = \hat{\theta}\int a(\theta)\,d\theta + \int b(\theta)\,d\theta =$$

$$= \hat{\theta}\left[A(\theta) + h_1(\boldsymbol{x})\right] + B(\theta) + h_2(\boldsymbol{x}) = A(\theta)\hat{\theta} + B(\theta) + h(\boldsymbol{x}),\ \forall \boldsymbol{x} \in \Psi$$

o sea

$$L(\boldsymbol{x},\theta) = \exp\left\{A(\theta)\hat{\theta} + B(\theta) + h(\boldsymbol{x})\right\},\ \forall \boldsymbol{x} \in \Psi$$

□

2.3 Estimadores de máxima verosimilitud

Definition 2.16. Sea $X \sim F(x,\theta), \theta \in \Theta \subset \mathbb{R}^k$ y sea $(X_1, ..., X_n)$ una m.a.s. de X. Dado $(x_1, ..., x_n) \in \Psi$, llamaremos **estimación de máxima verosimilitud de $\boldsymbol{\theta}$** al valor $\hat{\theta}$ que maximiza la función de verosimilitud.

Remark 2.17. La definición anterior nos permite definir una función

$$\begin{array}{rccc} \hat{\theta}: & \Psi & \to & \Theta \\ & \boldsymbol{x} & \mapsto & \hat{\theta}(\boldsymbol{x}) \end{array}$$

donde $\boldsymbol{x} = (x_1, ..., x_n)$, con la propiedad de que $L\left(\boldsymbol{x}, \hat{\theta}(\boldsymbol{x})\right) = \sup_{\theta\in\Theta} L(\boldsymbol{x},\theta)$.

Definition 2.18. A tal $\hat{\theta}(\boldsymbol{X})$ lo llamaremos **estimador de máxima verosimilitud (E.M.V)**.

Theorem 2.19. ***de Zehna***
Sea $X \sim F(x,\theta)$ con $\theta \in \Theta \subset \mathbb{R}^k$. Sea $h: \Theta \to \Lambda \subset \mathbb{R}^p$, con $p \leq k$.
Si $\hat{\theta}$ es E.M.V de θ, entonces $h\left(\hat{\theta}\right)$ es E.M.V de $h(\theta)$.

Proof. Sea $\lambda \in \Lambda$ y calculemos su función de verosimilitud.
Tenemos $L(\boldsymbol{x},\theta)$ y buscamos $M(\boldsymbol{x},\lambda)$, donde $M(\boldsymbol{x},\lambda)$ es la función de verosimilitud inducida por h.
λ tiene como preimágenes a $\Theta_\lambda = \{\theta \in \Theta | h(\theta) = \lambda\}$. Por tanto

$$M(\boldsymbol{x},\lambda) = \sup_{\theta\in\Theta} L(\boldsymbol{x},\theta)$$

Sea ahora $\boldsymbol{x} \in \Psi$ y $\hat{\theta}$ la EMV de θ. Sea $\hat{\lambda} = h\left(\hat{\theta}\right) \in \Lambda$. Entonces

$$M\left(\boldsymbol{x},\hat{\lambda}\right) = \sup_{\theta\in\Theta_{\hat{\lambda}}} L(\boldsymbol{x},\theta) = L\left(\boldsymbol{x},\hat{\theta}\right) \leq \sup_{\lambda\in\Lambda} M(\boldsymbol{x},\lambda) = \sup_{\lambda\in\Lambda}\left\{\sup_{\theta\in\Theta_\lambda} L(\boldsymbol{x},\theta)\right\} = \sup_{\theta\in\Theta} L(\boldsymbol{x},\theta) = L\left(\boldsymbol{x},\hat{\theta}\right)$$

por tanto, la desidualdad es, en realidad, una igualdad, y $M\left(\boldsymbol{x},\hat{\lambda}\right) = \sup_{\lambda\in\Lambda} M(\boldsymbol{x},\lambda)$, de donde $\hat{\lambda}$ es EMV de λ. □

2.4 Problemas

1. Sea $X \sim b(p)$ y $(X_1, ..., X_n)$ una m.a.s. de X.
a) Estudiar las condiciones de regularidad de Cramer-Rao para el estadístico $\hat{\rho} = \sum_{j=1}^{n} \frac{X_j}{n}$
b) Obtener la cantidad de información de Fisher
c) Estudiar si el estadístico $\hat{p}$ es EIMV de p
a) Veamos las 4 condiciones:
i) $\Psi = \{\boldsymbol{x} \in \mathbb{R} | x_i \in \{0,1\}, \forall i = 1, ..., n\}$ no depende de p.
La función de verosimilitud de $(X_1, ..., X_n)$:

$$L(\boldsymbol{x}, p) = P(\boldsymbol{X} = \boldsymbol{x}) = \prod_{j=1}^{n} P(X_j = x_j) = \prod_{j=1}^{n} p^{x_j} (1-p)^{1-x_j} = p^{\sum x_j} (1-p)^{n-\sum x_j}$$

ii) Trivialmente, $\exists \frac{\partial}{\partial p} L(\boldsymbol{x}, p), \forall \boldsymbol{x} \in \Psi$, pues es una función polinómica en p
iii) Para $E[\hat{p}] = \sum_{\boldsymbol{x} \in \Psi} \frac{\sum x_j}{n} p^{\sum x_i} (1-p)^{n-\sum x_j}$, se verifica que la derivada es la suma de las derivadas, pues es una suma finita ($card(\Psi) = 2^n$).
Para $1 = \sum_{\boldsymbol{x} \in \Psi} p^{\sum x_i} (1-p)^{n-\sum x_i}$ igual.
iv) y b)
Calculamos la cantidad de información de Fisher: $I_n(\theta) = nE\left[\left(\frac{\partial}{\partial \theta} \ln f(x, \theta)\right)^2\right]$

- $P(x,p) = p^x (1-p)^{1-x}$
- $\frac{\partial}{\partial p} \ln P(x,p) = \frac{\partial}{\partial p} \ln \left(p^x (1-p)^{1-x}\right) = \frac{1}{p^x (1-p)^{1-x}} \left[p^{x-1} (1-p)^{1-x} x + p^x (1-p)^{-x} (x-1)\right] = xp^{-1} + (x-1)(1-p)^{-1} = \frac{x}{p} + \frac{x-1}{1-p} = \frac{x-xp+xp-p}{p(1-p)} = \frac{x-p}{p(1-p)}$
- $\left(\frac{\partial}{\partial p} \ln P(x,p)\right)^2 = \frac{x^2-2xp+p^2}{p^2(1-p)^2}$
- $E\left[\left(\frac{\partial}{\partial p} \ln P(x,p)\right)^2\right] = \frac{1}{p^2(1-p)^2} \left[E[x^2] - 2pE[x] + p^2\right] = \frac{p-2p^2+p^2}{p^2(1-p)^2} = \frac{1}{p(1-p)}$

Luego

$$I_n(\theta) = \frac{n}{p(1-p)}$$

que es finito.
Así, verifica las condiciones de Cramer-Rao, y la cota es

$$Var(\hat{p}) \geq \frac{p(1-p)}{n}$$

Otra forma de calcular la cantidad de información de Fisher: $I_n(\theta) = -nE\left[\frac{\partial^2}{\partial \theta^2} \ln f(x, \theta)\right]$

- $\frac{\partial^2}{\partial p^2} \ln P(x,p) = \frac{-p(1-p)-(x-p)(1-p-p)}{p^2(1-p)^2} = \frac{-p+p^2-\left(x-2px-p+2p^2\right)}{p^2(1-p)^2} = \frac{-x+2px-p^2}{p^2(1-p)^2}$
- $E\left[\frac{\partial^2}{\partial p^2} \ln P(x,p)\right] = \frac{1}{p^2(1-p)^2} \left[-E[x] + 2pE[x] - p^2\right] = \frac{1}{p^2(1-p)^2} \left[-p + 2p^2 - p^2\right] = \frac{p^2-p}{p^2(1-p)^2} = \frac{-p(1-p)}{p^2(1-p)^2} = \frac{-1}{p(1-p)}$
- $I_n(p) = \frac{n}{p(1-p)}$

c) Como $Var[\hat{p}] = \frac{p(1-p)}{n}$, entonces $\hat{p}$ es EIMV de p.

2. $X \sim b(p)$ y $(X_1, ..., X_n)$ m.a.s. de X. Obtener el estimador de máxima verosimilitud de p.

Remark 2.20. Muchas veces es más sencillo hallar el máximo de $\ln L$ que de L, pues como L es un producto de funciones de densidad, entonces lnL es una suma y es probable que se simplifiquen los cálculos. Como el logaritmo es monótono, mantendrá los máximos.

$$L(\boldsymbol{x}) = \prod_1^n f(x_j, p) = \prod_1^n p^{x_j}(1-p)^{1-x_j} = p^{\sum x_j}(1-p)^{n-\sum x_j}$$

$$\ln L(\boldsymbol{x}) = \sum x_j \ln p + \left(n - \sum x_j\right)\ln(1-p)$$

$$\frac{\partial}{\partial p}\ln L(\boldsymbol{x}) = \frac{\sum x_j}{p} - \frac{n-\sum x_j}{1-p}$$

Dará 0, si, y solo si

$$(1-p)\sum x_j - \left(n - \sum x_j\right)p = 0 \iff p = \frac{\sum x_j}{n}$$

Para ver que es un máximo local:

$$\frac{\partial^2}{\partial p^2}\ln L(\boldsymbol{x}) = -\frac{\sum x_j}{p^2} - \frac{n-\sum x_j}{(1-p)^2} \overset{\textit{ambos términos negativos}}{<} 0$$

Por tanto es un máximo local, y más aún, como se da para todo p, entonces la función es convexa y solo tiene un máximo absoluto.

3. $X \sim Exp(\lambda)$ y $(X_1, ..., X_n)$ m.a.s. de X. Obtener el estimador de máxima verosimilitud de λ.

$$L(\boldsymbol{x}) = \prod \lambda e^{-\lambda x_j} = \lambda^n e^{-\lambda \sum x_j}$$

$$\ln L(\boldsymbol{x}) = n\ln\lambda - \lambda\sum x_j$$

$$\frac{\partial}{\partial \lambda}\ln L(\boldsymbol{x}) = \frac{n}{\lambda} - \sum x_j = 0 \iff \lambda = \frac{n}{\sum x_j}$$

$$\frac{\partial^2}{\partial \lambda^2}\ln L(x) = -\frac{n}{\lambda^2} < 0$$

es cóncava de nuevo y el máximo es absoluto.

4. $X \sim P(\lambda)$ y $(X_1, ..., X_n)$ m.a.s. de X. Obtener el estimador de máxima verosimilitud de λ.

$$L(\boldsymbol{x}) = \prod \frac{e^{-\lambda}\lambda^{x_j}}{x_j!} = \frac{e^{-\lambda n}\lambda^{\sum x_j}}{\prod x_j!}$$

$$\ln L(\boldsymbol{x}) = -\lambda n + \sum x_j \ln\lambda - \sum \ln x_j!$$

$$\frac{\partial}{\partial \lambda}\ln L(\boldsymbol{x}) = -n + \frac{\sum x_j}{\lambda} = 0 \iff \lambda = \frac{\sum x_j}{n}$$

$$\frac{\partial^2}{\partial \lambda^2}\ln L(\boldsymbol{x}) = -\frac{\sum x_j}{\lambda^2} < 0$$

esta también es cóncava y el máximo es absoluto.

5. $X \sim N\left(\mu, \sigma^2\right)$ y $(X_1, ..., X_n)$ m.a.s. de X. Obtener el estimador de máxima verosimilitud de $\left(\mu, \sigma^2\right)$.

$$L(\boldsymbol{x}) = \prod \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\left(x_j-\mu\right)^2}{2\sigma^2}} = \frac{1}{\sqrt{(2\pi\sigma^2)^n}} e^{-\frac{\sum\left(x_j-\mu\right)^2}{2\sigma^2}}$$

$$\ln L(\boldsymbol{x}) = -\frac{\sum (x_j-\mu)^2}{2\sigma^2} - \frac{n}{2}\left(\ln 2\pi + \ln\sigma^2\right)$$

$$\frac{\partial}{\partial\mu}\ln L(\boldsymbol{x}) = \frac{\sum (x_j-\mu)}{\sigma^2} = 0 \iff \sum x_j - n\mu = 0 \iff \mu = \frac{\sum x_j}{n}$$

$$\frac{\partial}{\partial\sigma^2}\ln L(\boldsymbol{x}) = \frac{\sum (x_j-\mu)^2}{2\sigma^4} - \frac{n}{2\sigma^2} = 0 \iff \sum (x_j-\mu)^2 - n\sigma^2 = 0 \iff \sigma^2 = \frac{\sum (x_j-\mu)^2}{n}$$

$$\frac{\partial^2}{\partial\mu^2}\ln L(\boldsymbol{x}) = -\frac{n}{\sigma^2} < 0$$

Es decir, es cóncava respecto de μ. Por tanto, de alcanzar un valor máximo, se debe alcanzar en la recta $\mu = \overline{\mu}$.

$$\frac{\partial^2}{\partial\sigma^2}\ln L(\boldsymbol{x}) = -\frac{\sum (x_j-\mu)^2}{\sigma^6} + \frac{n}{2\sigma^4} < 0 \iff -2\sum (x_j-\mu)^2 + n\sigma^2 < 0$$

$$\iff -2\sum\left(x_j^2 - 2x_j\mu + \mu^2\right) + n\sigma^2 < 0 \iff -2\sum x_j^2 + 4\mu\sum x_j - 2n\mu^2 + n\sigma^2 < 0$$

Y particularizando a la recta mencionada

$$-2\sum x_j^2 + \cancel{2\frac{\sum x_i}{n}\sum x_j} - \cancel{2\frac{(\sum x_j)^2}{n}} + n\sigma^2 < 0 \iff n\sigma^2 < 2\sum x_j^2$$

Y particularizamos al valor obtenido para σ^2

$$\sum (x_j-\mu)^2 < 2\sum x_j^2 \iff \sum\left(x_j^2 - 2\mu x_j + \mu^2\right) < 2\sum x_j^2 \iff \sum\left(-2\mu x_j + \mu^2\right) < \sum x_j^2$$

$$\iff -2\frac{\sum x_i}{n}\sum x_j + \frac{(\sum x_j)^2}{n} < \sum x_j^2 \iff -\frac{(\sum x_j)^2}{n} < \sum x_j^2$$

y esto se verifica. Por tanto, hay un máximo relativo en el punto obtenido. Ahora bien

$$\lim_{\sigma^2\to 0} L(\boldsymbol{x}) = 0$$

$$\lim_{\sigma^2\to\infty} L(\boldsymbol{x}) = 0$$

y entonces el máximo debe ser absoluto.

6. Sea X v.a. con función de densidad

$$f(x,\theta) = \frac{2x}{\theta}\exp\left(-\frac{x^2}{\theta}\right)\chi_{(0,\infty)}(x)$$

Sea $(X_1, ..., X_n)$ m.a.s. de X. Obtener el estimador de máxima verosimilitud de θ.

$$L(\boldsymbol{x}) = \prod \frac{2x_j}{\theta}\exp\left(-\frac{x_j^2}{\theta}\right)\chi_{(0,\infty)}(x_j)\, v = \frac{2^n\prod x_j}{\theta^n}\exp\left(-\frac{\sum x_j^2}{\theta}\right)\chi_{(0,\infty)}\left(\prod x_j\right)$$

$$\ln L(\boldsymbol{x}) = n\ln 2 + \sum \ln x_j - \frac{\sum x_j^2}{\theta} - n\ln\theta$$

$$\frac{\partial}{\partial \theta} \ln L(\boldsymbol{x}) = \frac{\sum x_j^2}{\theta^2} - \frac{n}{\theta} = 0 \iff \sum x_j^2 - n\theta = 0 \iff \theta = \frac{\sum x_j^2}{n}$$

$$\frac{\partial^2}{\partial \theta^2} \ln L(x) = -\frac{2\sum x_j^2}{\theta^3} + \frac{n}{\theta^2}$$

Particularizando

$$-\frac{2\sum x_j^2}{\left(\frac{\sum x_j^2}{n}\right)^3} + \frac{n}{\left(\frac{\sum x_j^2}{n}\right)^2} = -\frac{2n^3}{\left(\sum x_j^2\right)^2} + \frac{n^3}{\left(\sum x_j^2\right)^2} = -\frac{n^3}{\left(\sum x_j^2\right)^2} < 0$$

Por tanto es un máximo relativo.
Además

$$\lim_{\theta \to 0} L = 0$$

$$\lim_{\theta \to \infty} L = 0$$

y por tanto el máximo es absoluto.

7. Sea X v.a. con función de densidad

$$f(x, \theta) = \frac{\theta}{(1+x)^{1+\theta}} \chi_{(0,\infty)}(x)$$

Sea $(X_1, ..., X_n)$ m.a.s. de X. Obtener el estimador de máxima verosimilitud de θ.

$$L(\boldsymbol{x}) = \prod \frac{\theta}{(1+x_j)^{1+\theta}} \chi_{(0,\infty)}(x_j) = \frac{\theta^n}{\left[\prod (1+x_j)\right]^{1+\theta}} \chi_{(0,\infty)}\left(\prod x_j\right)$$

$$\ln L(\boldsymbol{x}) = n \ln \theta - (1+\theta) \sum \ln(1+x_j)$$

$$\frac{\partial}{\partial \theta} \ln L(\boldsymbol{x}) = \frac{n}{\theta} - \sum \ln(1+x_j) = 0 \iff \theta = \frac{n}{\sum \ln(1+x_j)}$$

$$\frac{\partial^2}{\partial \theta^2} \ln L(\boldsymbol{x}) = -\frac{n}{\theta^2} < 0$$

y es cóncava, por lo que es efectivamente un máximo absoluto.

8. $X \sim U(a, 5)$ y $(X_1, ..., X_n)$ m.a.s. de X. Obtener el estimador de máxima verosimilitud de a. Estudiar su distribución en el muestreo.

$$L(\boldsymbol{x}) = \prod \frac{1}{5-a} \chi_{(a,5)}(x_j) = \frac{1}{(5-a)^n} \chi_{(-\infty, X_{1:n})}(a)$$

Y el EMV de a es $\hat{a} = X_{1:n}$.
Básicamente, lo que hacemos es decir: los x_j están entre a y 5 si, y solo si, $a < \min(x_j)$.

3 Estimadores basados en estadísticos suficientes

3.1 Estadísticos suficientes. Teorema de Factorización

Definition 3.1. Sea $X \sim F(x,\theta), \theta \in \Theta \subset \mathbb{R}$. Sea $(X_1,...,X_n)$ m.a.s. de X, el **estadístico $\boldsymbol{T = T(X_1,...,X_n)}$ es suficiente para el parámetro $\boldsymbol{\theta}$**, si la distribución de $(X_1,...,X_n)$ m.a.s. de X condicionada a que $T(X_1,...,X_n) = t$ no depende de θ.
Es decir, si la distribución de

$$(X_1,...,X_n | T(X_1,...,X_n) = t)$$

no depende de θ, para cualquier valor de $t \in Sop(T)$.

Theorem 3.2. ***Teorema de Factorización***
Sea $X \sim F(x,\theta), \theta \in \Theta \subset \mathbb{R}$. Sea $(X_1,...,X_n)$ m.a.s. de X, entonces el estadístico $T = T(X_1,...,X_n)$ es un estadístico suficiente para el parámetro θ si, y solo si, la función de verosimilitud es de la forma

$$L(\boldsymbol{x},\theta) = g(T(\boldsymbol{x}),\theta) \cdot l(\boldsymbol{x})$$

Proof. Lo hacemos para el caso discreto
$[\Longleftarrow]$

$$P(X_1 = x_1,...,X_n = x_n | T(\boldsymbol{X}) = t) = \frac{P(\boldsymbol{X} = \boldsymbol{x}, T(\boldsymbol{X}) = t)}{P(T(\boldsymbol{X}) = t)} = \begin{cases} 0 & T(\boldsymbol{X}) \neq t \\ \frac{g(t,\theta)\cdot l(\boldsymbol{x})}{P(T(\boldsymbol{X}=t))} & T(\boldsymbol{X}) = t \end{cases}$$

$$P(T(\boldsymbol{X}) = t) = \sum_{\boldsymbol{Y}\in\Psi, T(\boldsymbol{Y})=t} P(\boldsymbol{X} = \boldsymbol{Y}) = \sum_{\boldsymbol{Y}\in\Psi, T(\boldsymbol{Y})=t} g(t,\theta)\, l(\boldsymbol{Y}) =$$

Y por tanto

$$P(\boldsymbol{X} = \boldsymbol{x} | T(\boldsymbol{X}) = t) = \begin{cases} 0 & T(\boldsymbol{X}) \neq t \\ \frac{g(t,\theta)\cdot l(\boldsymbol{x})}{g(t,\theta)\sum_{\boldsymbol{Y}\in\Psi, T(\boldsymbol{Y})=t} l(\boldsymbol{Y})} & T(\boldsymbol{X}) = t \end{cases} = \begin{cases} 0 & T(\boldsymbol{X}) \neq t \\ \frac{l(\boldsymbol{x})}{\sum_{\boldsymbol{Y}\in\Psi, T(\boldsymbol{Y})=t} l(\boldsymbol{Y})} & T(\boldsymbol{X}) = t \end{cases}$$

y la distribución de $(X_1,...,X_n | T(X_1,...,X_n) = t)$ no depende de θ, por lo que T es suficiente para θ.
$[\Longrightarrow]$

$$P(\boldsymbol{X} = \boldsymbol{x} | T(\boldsymbol{X}) = T(\boldsymbol{x})) = \frac{P(\boldsymbol{X} = \boldsymbol{x}, T(\boldsymbol{X}) = T(\boldsymbol{x}))}{P(T(\boldsymbol{X}) = T(\boldsymbol{x}))} \overset{\boldsymbol{X}=\boldsymbol{x} \Longrightarrow T(\boldsymbol{X})=T(\boldsymbol{x})}{=} \frac{P(\boldsymbol{X} = \boldsymbol{x})}{P(T(\boldsymbol{X}) = T(\boldsymbol{x}))} =$$

$$= \frac{L(\boldsymbol{x},\theta)}{P(T(\boldsymbol{X}) = T(\boldsymbol{x}))}$$

Luego

$$L(\boldsymbol{x},\theta) = P(T(\boldsymbol{X}) = T(\boldsymbol{x})) \cdot P(\boldsymbol{X} = \boldsymbol{x} | T(\boldsymbol{X}) = T(\boldsymbol{x}))$$

donde $g(T(\boldsymbol{x}),\theta) = P(T(\boldsymbol{X}) = T(\boldsymbol{x}))$ y $l(\boldsymbol{x}) = P(\boldsymbol{X} = \boldsymbol{x} | T(\boldsymbol{X}) = T(\boldsymbol{x}))$ no depende de θ porque T es suficiente para θ. $\square$

Definition 3.3. Sea T un estadístico con soporte S_T y sea $\phi : S_T \to \mathbb{R}$ una función inyectiva. Entonces el estadístico $T' = \phi(T)$ se dice que **está basado en $\boldsymbol{T}$**.

Proposition 3.4. *Sea $X \sim F(x,\theta), \theta \in \Theta \subset \mathbb{R}$ y sea T' un estadístico basado en T, que es suficiente para θ, entonces T' es suficiente para θ.*

3.2 Estadísticos suficientes y EIMV. Teorema de Rao-Blackwell

Repaso de esperanzas condicionadas
Sea (X, Y) una variable aleatoria bidimensional. Se define

$$h(y) = E[X|Y = y], \; y \in Sop(Y)$$

Si h es medible Borel podemos considerar $E[h(Y)] = E[E[X|Y]]$, en el caso de que exista.
Propiedades:

1. $E[h(Y)] = E[X]$ si $E[X]$ existe
2. Si existe $E\left[X^2\right]$ entonces

$$Var[X] = Var[h(Y)] + E[l(Y)]$$

donde $l(y) = Var[X|Y = y]$.

Theorem 3.5. ***Rao-Blackwell***
Sea $X \sim F(x,\theta), \theta \in \Theta \subset \mathbb{R}$. Sea $(X_1, ..., X_n)$ m.a.s. de X y $T = T(X_1, ..., X_n)$ un estadístico suficiente para el parámetro θ y sea $\hat{\theta}$ un estimador insesgado de θ.
Consideremos el estadístico $\widetilde{\theta} = E\left[\hat{\theta}|T\right]$, entonces se verifica:

1. *$\widetilde{\theta}$ no depende de θ y es función de T*
2. *$\widetilde{\theta}$ es estimador insesgado de θ*
3. *$Var\left(\widetilde{\theta}\right) \leq Var\left(\hat{\theta}\right)$ y la desigualdad es estricta si $\hat{\theta}$ no es función de T*

Proof. Veamos cada afirmación:

1. $\tilde{\theta}$ es función de T trivialmente. Por otro lado, como T es suficiente, entonces la distribución de $\left(\hat{\theta}|T = t\right)$ no depende de θ, por lo que su esperanza tampoco depende de θ.
2. $$E\left[\tilde{\theta}\right] = E\left[E\left[\hat{\theta}|T\right]\right] \overset{Propiedad\ 1}{=} E\left[\hat{\theta}\right] = \theta \implies \tilde{\theta} \text{ es estimador insesgado de } \theta$$
3. Se tiene, por la propiedad 2, que

$$Var\left[\hat{\theta}\right] = Var\left[\tilde{\theta}\right] + E[l(T)]$$

con $l(t) = Var\left[\hat{\theta}|T = t\right] \geq 0, \; \forall t \in Sop(T)$.

Por tanto $l(T) \geq 0$ *c.s.* y por tanto $E[l(T)] \geq 0$ y entonces

$$Var\left[\tilde{\theta}\right] \leq Var\left[\hat{\theta}\right]$$

Vamos ahora a ver que si $Var\left[\tilde{\theta}\right] = Var\left[\hat{\theta}\right]$ entonces $\hat{\theta}$ es función de T. Este igualdad se da si, y solo si, $E\left[l\left(T\right)\right] = 0$, por lo que $l\left(T\right) = 0$ $c.s.$ y entonces $l\left(t\right) = 0, \forall t \in Sop\left(T\right)$. Como $l\left(t\right) = Var\left(\hat{\theta}|T = t\right) = 0$, por lo tanto $\left(\hat{\theta}|T = t\right)$ es degenerada en un único punto, que es

$$E\left[\theta|T = t\right] = h\left(t\right),\ c.s.$$

O sea, si $t \in Sop\left(T\right), \exists A_t \subset \Omega|\forall w \in A_t, \left(\hat{\theta}\left(w\right)|T = t\right) = h\left(t\right)$ y $P\left(A_t^c\right) = 0$. Así, $T\left(w\right) = t$ al ser un suceso en el espacio muestral condicionado por el suceso $\{T = t\}$. Podemos entonces escribir que $\forall w \in A_t$, $\left(\hat{\theta}\left(w\right)|T = t\right) = h\left(T\left(w\right)\right) = \tilde{\theta}\left(w\right)$.

Vamos a ver que $\hat{\theta} = h\left(T\right),\ c.s.$

Consideremos $A = \bigcup_{t \in Sop(T)} A_t$ y tomemos $w \in A$. Entonces $\exists t \in Sop\left(T\right)|w \in A_t \implies \hat{\theta}\left(w\right) = h\left(T\left(w\right)\right)$.

Ahora bien

$$0 \leq P\left(A^c\right) \leq P\left(A_t^c\right) \overset{*}{=} 0 \implies 0 = P\left(A^c\right)$$

Y entonces $\hat{\theta} = h\left(T\right)$.

$*$ no es trivial, vamos a verlo en el caso discreto: Consideremos $A'_t = A_t \cap \{T = t\}$ y $A_t^{c'} = A_t^c \cap \{T = t\}$ donde

$$P\left(A_t|T = t\right) = \frac{P\left(A_t \cap \{T = t\}\right)}{P\left(T = t\right)} = \frac{P\left(A'_t\right)}{P\left(T = t\right)} = 1$$

$$P\left(A_t^c|T = t\right) = \frac{P\left(A_t^c \cap \{T = t\}\right)}{P\left(T = t\right)} = \frac{P\left(A_t^{c'}\right)}{P\left(T = t\right)} = 0 \implies P\left(A_t^{c'}\right) = 0$$

Sea $\Omega_T = \{w|T\left(w\right) \in Sop\left(T\right)\}$, entonces $P\left(\Omega_T^c\right) = 0$ porque los sucesos que no forman parte del sopore de la variable tienen probabilidad nula.

Como estamos en el caso discreto, $Sop\left(T\right) = \{t_1, .., t_n, ...\}$ y podemos tomar $T_i = \{w \in \Omega|T\left(w\right) = t_i\}$, de forma que

$$\Omega_T = \bigcup_{t_i \in Sop(T)} T_i = \bigcup_{t_i \in Sop(T)} \left(A'_{t_i} \cup A_t^{c'}\right) = \left(\bigcup_{t_i \in Sop(T)} A'_{t_i}\right) \cup \left(\bigcup_{t_i \in Sop(T)} A_t^{c'}\right)$$

y finalmente tenemos que $1 = P\left(\Omega_T\right) = P\left(A\right) + P\left(A'\right)$ y $P\left(A'\right) = \sum_{t_i \in Sop(T)} P\left(A_t^{c'}\right) = 0$, por lo que $P\left(A\right) = 1$

□

Remark 3.6. Si existe un estadístico suficiente T y un estadístico $\hat{\theta}$ EIMV del parámetro θ, entonces $\hat{\theta}$ debe ser función de T, pues de lo contrario, por el teorema de Rao-Blackwell, podríamos construir un estimador insesgado con varianza menor que la de $\hat{\theta}$, lo cual sería una contradicción.

3.3 Estadísticos suficientes, completos e insesgados de mínima varianza. Teorema de Lehmann-Scheffé

Definition 3.7. $X \sim F_\theta, \theta \in \Theta \subset \mathbb{R}$. Sea $\left(X_1, ..., X_n\right)$ m.a.s. de X. El estadístico $T = T\left(X_1, ..., X_n\right)$ es **completo para el parámetro $\boldsymbol{\theta}$** si para cualquier transformación medible Borel tal que $E\left[g\left(T\right)\right] = 0$ entonces se tiene que $g\left(T\right) = 0,\ c.s.$

Theorem 3.8. *$X \sim F_\theta, \theta \in \Theta \subset \mathbb{R}$. Sea $(X_1, ..., X_n)$ m.a.s. de X, donde X pertenece a la familia exponencial uniparamétrica:*

$$L(\boldsymbol{x}, \theta) = \exp\{A(\theta) T(\boldsymbol{x}) + B(\theta) + h(\boldsymbol{x})\}$$

y sea $\Delta = A(\Theta)$.
Si Δ contiene un intervalo abierto, entonces T es completo.

Remark 3.9. En tal caso, el estadístico también será suficiente. Aplicando el teorema de Factorización con

$$g(T(\boldsymbol{x}), \theta) = \exp\{A(\theta) T(\boldsymbol{x}) + B(\theta)\}$$
$$l(\boldsymbol{x}) = \exp\{h(\boldsymbol{x})\}$$

Theorem 3.10. ***Lehmann-Scheffé***
$X \sim F_\theta, \theta \in \Theta \subset \mathbb{R}$. Sea $(X_1, ..., X_n)$ m.a.s. de X y $T = T(X_1, ..., X_n)$ un estadístico suficiente y completo para θ.
Si existe un estimador insesgado de θ función de T, entonces es el estimador insesgado de mínima varianza de θ.

Proof. ¿$Var\left(\hat{\theta}\right) \leq Var\left(\tilde{\theta}\right), \forall \tilde{\theta}$, *E.I. de* θ?

- Sean $\hat{\theta}_1, \hat{\theta}_2$ estimadores insesgados de θ tales que $\hat{\theta}_1 = h_1(Y), \hat{\theta}_2 = h_2(T)$ y consideremos $\hat{\theta}_1 - \hat{\theta}_2$ y $h = h_1 - h_2$, entonces $E[h(T)] = E\left[\hat{\theta}_1 - \hat{\theta}_2\right] = 0$, pero T es completo, por lo que $\hat{\theta}_1 - \hat{\theta}_2 = h(T) = 0$, *c.s.* y entonces $\hat{\theta}_1 = \hat{\theta}_2$, *c.s.* Así, $\hat{\theta}$ es el único estimador insesgado que es función de T.

- Sea ahora $\tilde{\theta}$ un estimador insesgado de θ que no es función de T, por el teorema de Rao-Blackwell construimos $\overline{\theta} = E\left[\tilde{\theta}|T\right]$ donde $E\left[\overline{\theta}\right] = \theta$ y $Var\left[\overline{\theta}\right] \overset{\tilde{\theta} \text{ no es función de } T}{<} Var\left[\tilde{\theta}\right]$.

  Como $\overline{\theta}$ es E.I. función de T, entonces por lo visto antes es $\overline{\theta} = \hat{\theta}$.

Por tanto, $\hat{\theta}$ es EIMV. □

3.4 Ejercicios

1. $X \sim b(p)$ y $(X_1, ..., X_n)$ m.a.s. de X.
a) Estudiar si $T = \sum_{i=1}^n X_i$ es suficiente para p mediante la definición
Tenemos que ver que $P(\boldsymbol{X} = \boldsymbol{x}|T(\boldsymbol{X}) = t)$ es independiente de p, para todo $t \in Sop(T) = \{0, ..., n\}$

$$P(\boldsymbol{X} = \boldsymbol{x}|T(\boldsymbol{X}) = t) = P(Obtener\ una\ ordenación\ concreta\ entre\ las\ que\ tienen\ t\ apariciones\ favorables) =$$

que no depende de t. Por tanto, T es suficiente.
b) Estudiar la suficiencia de T mediante el teorema de factorización

$$L(\boldsymbol{x}) = p^{\sum x_j}(1-p)^{n-\sum x_j} = p^T(1-p)^{n-T} = \exp\{T \log p + (n-T)\log(1-p)\} =$$

$$= \exp\left\{T\left(\log\frac{p}{1-p}\right) + n\log(1-p)\right\}$$

Por lo que tomando $g(T(\boldsymbol{x}), p) = p^T (1-p)^T$ y $l(\boldsymbol{x}) = 1$ y aplicando el teorema de Factorización, tenemos que T es suficiente para p.
c) Estudiar si T es completo para p
Tomando $A(p) = \log \frac{p}{1-p},\ B(p) = n\log(1-p), h(\boldsymbol{x}) = 0$, tenemos que pertenece a la familia exponencial uniparamétrica. Además

$$\lim_{p\to 0} A(p) = -\infty \qquad \lim_{p\to 1} A(p) = \infty$$

por lo que

$$A(\Theta) = A((0,1)) = \mathbb{R}$$

por lo que contiene un intervalo y T es completo para p.
d) Obtener a partir de T un EIMV para p
Como $E[T] = np$, entonces $\hat{p} = \frac{T}{n}$ es EI de p, y por ser T suficiente y completo, y $\hat{p}$ función de T, entonces el teorema de Lehmann-Scheffé asegura que $\hat{p}$ es EIMV de p.

2. $X \sim P(\lambda)$ y $(X_1, ..., X_n)$ m.a.s. de X
a) Estudiar si $T = \sum_{i=1}^n X_i$ es suficiente para λ
b) Estudiar si T es completo para λ
c) Obtener a partir de T un EIMV para λ

$$L(\boldsymbol{x}) = \frac{e^{-\lambda n}\lambda^{\sum x_j}}{\prod x_j!} = \exp\left\{\log\lambda \cdot \sum x_i - \lambda n - \log\left(\prod x_i!\right)\right\}$$

a) Tomando $g(T(\boldsymbol{x}), \lambda) = e^{-\lambda n}\lambda^{\sum x_j}$ y $l(\boldsymbol{x}) = \frac{1}{\prod x_j!}$ y aplicando el teorema de factorización, tenemos que T es suficiente para λ.
b) Tomando $A(\lambda) = \log\lambda,\ T(\boldsymbol{x}) = \sum x_i,\ B(\lambda) = -\lambda n,\ h(\boldsymbol{x}) = -\log\left(\prod x_i!\right)$, tenemos que pertenece a la familia exponencial uniparamétrica. Además

$$A(\Theta) = \log\left(\mathbb{R}^+\right) = \mathbb{R}$$

por lo que contiene un intervalo abierto, y entonces T es completo para λ.
c) Como $E[T] = n\lambda$, entonces $\hat{X} = \frac{T}{n}$ es estimador insesgado para λ, y por ser T suficiente y completo, y $\hat{X}$ función de T, entonces el teorema de Lehmann-Scheffé asegura que $\hat{X}$ es EIMV de λ.

3. $X \sim Exp\left(\frac{1}{\theta}\right)$ y $(X_1, ..., X_n)$ m.a.s. de X
a) Estudiar si $T = \sum_{i=1}^n X_i$ es suficiente para θ

$$L(\boldsymbol{x}) = \prod \frac{1}{\theta} e^{-\frac{1}{\theta}x_j} = \theta^{-n} e^{-\frac{1}{\theta}\sum x_j} = \exp\left\{-\frac{1}{\theta}\sum x_j - n\log\theta\right\}$$

Tomando $g(T, \theta) = \theta^{-n} e^{-\frac{1}{\theta}T}$ y $l(\boldsymbol{x}) = 1$, por el teorema de Factorización, T es suficiente para θ
b) Estudiar si T es completo para θ
Tomando $A(\theta) = -\frac{1}{\theta}, B(\theta) = -n\log(\theta),\ h(\boldsymbol{x}) = 0$, vemos que pertenece a la familia exponencial uniparamétrica. Además

$$A(\Theta) = A\left(\mathbb{R}^+\right) = \mathbb{R}^-$$

que contiene un intervalo abierto, y entonces T es completo para θ.
c) Obtener a partir de T un EIMV para θ
Puesto que $E(T) = n\theta$, entonces $\hat{\theta} = \frac{T}{n}$ es EI de θ, y por ser T suficiente y completo, y $\hat{\theta}$ función de T, entonces el teorema de Lehmann-Scheffé asegura que $\hat{\theta}$ es EIMV de λ.
d) Obtener la cantidad de información de Fisher para θ

$$I_n(\theta) = nE\left[\left(\frac{\partial}{\partial\theta}\log\left(\frac{1}{\theta}e^{-\frac{1}{\theta}x}\right)\right)^2\right] = nE\left[\left(\frac{\partial}{\partial\theta}\left(\frac{x}{\theta}\log\theta\right)\right)^2\right] = nE\left[\left(-\frac{x}{\theta^2}\log\theta + \frac{x}{\theta^2}\right)^2\right] =$$

$$= nE\left[\frac{x^2}{\theta^4}(1-\log\theta)^2\right] = \frac{n(1-\log\theta)^2}{\theta^4}E\left(x^2\right) = \frac{n(1-\log\theta)^2}{\theta^4}\left[Var(x) + E(x)^2\right] =$$

$$= \frac{n(1-\log\theta)^2}{\theta^4}\left[2\theta^2\right] = \frac{2n(1-\log\theta)^2}{\theta^2}$$

4. $X \sim N\left(\mu, \sigma^2\right)$, σ^2 conocido y $(X_1, ..., X_n)$ m.a.s. de X
a) Estudiar si $T = \sum_{i=1}^n X_i$ es suficiente para μ

$$L(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi\sigma^2)^n}}\exp\left\{-\frac{\sum(x_j-\mu)^2}{2\sigma^2}\right\} = \frac{1}{\sqrt{(2\pi\sigma^2)^n}}\exp\left\{-\frac{\sum\left(x_j^2 - 2\mu x_j + \mu^2\right)}{2\sigma^2}\right\} =$$

$$= \frac{1}{\sqrt{(2\pi\sigma^2)^n}}\exp\left\{-\frac{\sum x_j^2}{2\sigma^2}\right\}\exp\left\{\frac{\mu}{\sigma^2}\sum x_j - \frac{n\mu^2}{2\sigma^2}\right\} = \exp\left\{\frac{\mu}{\sigma^2}\sum x_j - \frac{n\mu^2}{2\sigma^2} - \frac{\sum x_j^2}{2\sigma^2} - \frac{n}{2}\log\left(2\pi\sigma^2\right)\right\}$$

Tomando $g(T,\mu) = \exp\left\{\frac{\mu}{\sigma^2}\sum x_j - \frac{n\mu^2}{2\sigma^2}\right\}$, $l(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi\sigma^2)^n}}\exp\left\{-\frac{\sum x_j^2}{2\sigma^2}\right\}$ y aplicando el teorema de Factorización, tenemos que T es suficiente para μ.
b) Estudiar si T es completo para μ
Tomando $A(\mu) = \frac{\mu}{\sigma^2}, B(\mu) = -\frac{n\mu^2}{2\sigma^2}$, $h(\boldsymbol{x}) = -\frac{\sum x_j^2}{2\sigma^2} - \frac{n}{2}\log\left(2\pi\sigma^2\right)$, tenemos que pertenece a la familia exponencial uniparamétrica. Además

$$A(\Theta) = A(\mathbb{R}) = \mathbb{R}$$

que contiene un intervalo, por lo que T es completo para μ.
c) Obtener a partir de T un EIMV para μ
Como $E(T) = n\mu$, entonces $\hat{\mu} = \frac{T}{n}$ es EI de μ, y como T es suficiente y completo, y $\hat{\mu}$ función de T, el teorema de Lehmann-Scheffé asegura que es EIMV de μ.
d) Obtener la cantidad de información de Fisher para μ

$$I_n(\theta) = nE\left[\left(\frac{\partial}{\partial\mu}\log\left(\exp\left\{\frac{\mu x}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - \frac{x^2}{2\sigma^2} - \frac{1}{2}\log\left(2\pi\sigma^2\right)\right\}\right)\right)^2\right] =$$

$$= nE\left[\left(\frac{\partial}{\partial\mu}\left(\frac{\mu x}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - \frac{x^2}{2\sigma^2} - \frac{1}{2}\log\left(2\pi\sigma^2\right)\right)\right)^2\right] = nE\left[\left(\frac{x}{\sigma^2} - \frac{\mu}{\sigma^2}\right)^2\right] =$$

$$= \frac{n}{\sigma^4}E\left[x^2 - 2x\mu + \mu^2\right] = \frac{n}{\sigma^4}\left(E\left(x^2\right) - 2\mu^2 + \mu^2\right) = \frac{n}{\sigma^4}\left(Var(x) + E(x)^2 - \mu^2\right) = \frac{n\sigma^2}{\sigma^4} = \frac{n}{\sigma^2}$$

5. Sea X v.a. con función de densidad

$$f(x,\theta) = \frac{2x}{\theta}\exp\left(-\frac{x^2}{\theta}\right)\chi_{(0,\infty)}(x)$$

Sea $(X_1, ..., X_n)$ m.a.s. de X.

a) Estudiar si $T=\sum_{i=1}^{n}X_i^2$ es suficiente para θ

$$L(\boldsymbol{x})=\frac{2^n\prod x_j}{\theta^n}\exp\left(-\frac{\sum x_j^2}{\theta}\right)=\exp\left(-\frac{\sum x_j^2}{\theta}-n\log\theta+n\log 2+\sum\log x_j\right)$$

Tomando $g(T,\theta)=\frac{1}{\theta^2}\exp\left(-\frac{\sum x_j^2}{\theta}\right)$, $l(\boldsymbol{x})=2^n\prod x_j$, por el teorema de factorización tenemos que T es suficiente para θ.
b) Estudiar si T es completo para θ
Tomando $A(\theta)=-\frac{1}{\theta}, B(\theta)=-n\log\theta,\ h(\boldsymbol{x})=n\log 2+\sum\log x_j$, vemos que pertenece a la familia exponencial uniparamétrica. Además

$$A(\Theta)=A\left(\mathbb{R}^+\right)=\mathbb{R}^-$$

que contiene un intervalo, por lo que T es completo para θ.
c) Obtener a partir de T un EIMV para θ
Como $E(T)=\sum E\left(x^2\right)=nE\left(x^2\right)$,

$$E\left(x^2\right)=\int_0^\infty\frac{2x^3}{\theta}e^{-\frac{x^2}{\theta}}dx=\frac{2}{\theta}\int_0^\infty x^3e^{-\frac{x^2}{\theta}}dx=\left[\begin{array}{c}x^2=t\\2xdx=dt\end{array}\right]=\frac{1}{\theta}\int_0^\infty te^{-\frac{t}{\theta}}dt=$$

$$=\left[\begin{array}{c}-\frac{t}{\theta}=z\\-\frac{1}{\theta}dt=dz\\t=\infty\implies z=-\infty\\t=0\implies z=0\end{array}\right]=-\int_0^{-\infty}-\theta ze^zdz=-\theta\int_{-\infty}^0 ze^zdz=\left[\begin{array}{c}u=z\implies du=dz\\dv=e^zdz\implies v=e^z\end{array}\right]$$

$$=-\theta ze^z|_{-\infty}^0+\theta\int_{-\infty}^0e^zdz=0+\theta e^z|_{-\infty}^0=\theta$$

Es decir, $E(T)=n\theta$, por lo que $\hat{\theta}=\frac{T}{n}$ es EI de θ y como T es suficiente y completo, y $\hat{\theta}$ función de T, por el teorema de Lehmann-Scheffé $\hat{\theta}$ es EIMV de θ.
d) Obtener la cantidad de información de Fisher para θ

$$I_n(\theta)=nE\left[\left(\frac{\partial}{\partial\theta}\log f(x)\right)^2\right]=nE\left[\left(\frac{\partial}{\partial\theta}\left(\log(2x)-\log\theta-\frac{x^2}{\theta}\right)\right)^2\right]=nE\left[\left(-\frac{1}{\theta}+\frac{x^2}{\theta^2}\right)^2\right]=$$

$$=nE\left[\left(\frac{x^2-\theta}{\theta^2}\right)^2\right]=\frac{n}{\theta^4}E\left[x^4-2x^2\theta+\theta^2\right]=\frac{n}{\theta^4}\left(E\left(x^4\right)-2\theta^2+\theta^2\right)=$$

$$=\frac{n}{\theta^4}\left(2\theta^2-\theta^2\right)=\frac{n}{\theta^2}$$

6. Sea $X\sim U(0,\theta)$ y $(X_1,...,X_n)$ m.a.s. de X
a) Estudiar si $X_{n:n}=\max(X_1,...,X_n)$ es suficiente para θ

$$f(x)=\frac{1}{\theta}\chi_{[0,\theta]}$$

$$L(\boldsymbol{x},\theta)=\prod\frac{1}{\theta}\chi_{[0,\theta]}(x_j)=\prod\frac{1}{\theta}\chi_{[X_{n:n},\infty)}(\theta)=\frac{1}{\theta^n}\chi_{[X_{n:n},\infty)}(\theta)$$

y tomando $g(X_{n:n},\theta)=\frac{1}{\theta^n}\chi_{[X,\infty)}(\theta)$ y $l(\boldsymbol{x})=1$, por el teorema de factorización tenemos que $X_{n:n}$ es suficiente para θ.

b) Estudiar si $X_{n:n}$ es completo para θ
Queremos ver que si $E\left[g\left(X_{n:n}\right)\right]=0$, entonces $g\left(X_{n:n}\right)=0$ c.s.

$$E\left[g\left(X_{n:n}\right)\right]=\int_0^\theta g\left(x\right)dF_{n:n}\left(x\right)=0$$

con $F_{n:n}$ la función de distribución del máximo, $F_{n:n}\left(x\right)=\left(F\left(x\right)\right)^n=\left(\frac{x}{\theta}\right)^n, \forall x\in Sop\left(X\right)=\left(0,\theta\right)$.

$$E\left[g\left(X_{n:n}\right)\right]=\int_0^\theta g\left(x\right)\frac{nx^{n-1}}{\theta^n}dx=\frac{n}{\theta^n}\int_0^\theta g\left(x\right)x^{n-1}dx=0,\forall\theta\in\mathbb{R}^+$$

$$\iff h\left(\theta\right)=\int_0^\theta g\left(x\right)x^{n-1}dx=0,\forall\theta\in\mathbb{R}^+\implies h'\left(\theta\right)=g\left(\theta\right)\theta^{n-1}=0\overset{\theta>0}{\iff}g\left(\theta\right)=0,\forall\theta>0$$

Y es completo.
c) Obtener a partir de $X_{n:n}$ un EIMV para θ

$$E\left[X_{n:n}\right]=\int_0^\theta xdF_{n:n}\left(x\right)=\int_0^\theta\frac{x^{n+1}}{\theta^n}dx=\frac{x^{n+2}}{\theta^n\left(n+2\right)}|_0^\theta=\frac{\theta^2}{n+2}$$

Y entonces

$$T=\sqrt{X_{n:n}\cdot\left(n+2\right)}$$

es EI de θ.
Como $X_{n:n}$ es suficiente y completo, y T es EI y función de $X_{n:n}$, entonces es EIMV de θ por el teorema de Lehmann-Scheffé.

7. Sea X_1,X_2 m.a.s. de $X\sim N\left(\mu,1\right)$
a) Calcular $E\left[\frac{X_1+X_2}{2}|X_1=x\right]$
$\hat{\theta}=\frac{X_1+X_2}{2}$ es EI de μ y $X_1=T$

$$\left(\hat{\theta}|T=t\right)=\left(\frac{X_1+X_2}{2}|X_1=x\right)=\frac{x+X_2}{2}\sim N\left(\frac{x+\mu}{2},\frac{1}{4}\right)$$

$$E\left[\frac{x+X_2}{2}\right]=\frac{1}{2}\left(x+E\left[X_2\right]\right)$$

$$Var\left[\frac{x+X_2}{2}\right]=\frac{1}{4}$$

Y entonces la distribución no es independiente de x, por lo que X_1 no es suficiente para μ.
b) Sea $\hat{\theta}=E\left[\frac{X_1+X_2}{2}|X_1\right]$. Calcular $E\left[\hat{\theta}\right]$ y $Var\left[\hat{\theta}\right]$. Comparar $Var\left[\hat{\theta}\right]$ y $Var\left[\frac{X_1+X_2}{2}\right]$.

$$\tilde{\theta}=h\left(t\right)=E\left[\frac{X_1+X_2}{2}|X_1=t\right]=\frac{t+\mu}{2}$$

por lo que $\tilde{\theta}$ depende de μ y por tanto no sirve para estimarlo. Seguimos con el ejercicio

$$E\left[\tilde{\theta}\right]=\mu$$

$$Var\left[\tilde{\theta}\right]=\frac{1}{4},\ Var\left[\frac{X_1+X_2}{2}\right]=\frac{1}{2}$$

Por tanto

$$Var\left(\tilde{\theta}\right)<Var\left(\frac{X_1++X_2}{2}\right)$$

4 Estimación mediante intervalos de confianza

4.1 Introducción

Definition 4.1. $X \sim F_\theta, \theta \in \Theta \subset \mathbb{R}$, sea $\boldsymbol{X} = (X_1, ..., X_n)$ m.a.s. de X, se dice que el intervalo $(i(\boldsymbol{X}), s(\boldsymbol{X}))$ es un **intervalo de confianza al nivel $\mathbf{1 - \alpha}$ para el parámetro $\boldsymbol{\theta}$** si

$$P(i(\boldsymbol{X}) \leq \theta \leq s(\boldsymbol{X})) \geq 1 - \alpha$$

Se llama **nivel de confianza** a

$$\inf_{\theta \in \Theta} P(i(\boldsymbol{X}) \leq \theta \leq s(\boldsymbol{X})) = 1 - \alpha$$

Remark 4.2. En el caso continuo, pueden conseguirse normalmente intervalos con igualdad estricta. Esto no sucede así con variables discretas, caso en el que deberemos encontrar un intervalo de nivel de confianza $1 - \alpha' \geq 1 - \alpha$

4.2 Método de la función pivote

Definition 4.3. $X \sim F_\theta, \theta \in \Theta \subset \mathbb{R}$ y $\boldsymbol{X} = (X_1, ..., X_n)$ m.a.s. de X. Sea $T : \Psi \times \Theta \to \mathbb{R}$, $\Psi = sop(\boldsymbol{X})$. Se dice que T es una **función pivote para $\boldsymbol{\theta}$** si

1. $\forall \boldsymbol{x} \in \Psi, T(\boldsymbol{x}, \theta)$ es una función estrictamente monótona en $\theta \in \Theta$
2. La distribución de $T(\boldsymbol{X}, \theta)$ es independiente de θ

Proposition 4.4. *Sea $X \sim F_\theta, \theta \in \Theta \subset \mathbb{R}$ y $\boldsymbol{X} = (X_1, ..., X_n)$ m.a.s. de X. Sea $T(\boldsymbol{X}, \theta)$ una función pivote para θ y $\Lambda = sop(T)$.*
Supongamos que $\forall \boldsymbol{x} \in \Psi, \lambda \in \Lambda$, la ecuación $\lambda = T(\boldsymbol{x}, \theta)$ puede resolverse en θ.
Entonces se puede construir un intervalo de confianza al nivel $1 - \alpha$ para θ.

Proof. Como $T(\boldsymbol{x}, \theta)$ no depende en su distribución de θ, podemos encontrar dos valores $\lambda_1(\alpha), \lambda_2(\alpha) \in \Lambda$ tales que

$$P(\lambda_1(\alpha) \leq T(\boldsymbol{x}, \theta) \leq \lambda_2(\alpha)) \geq 1 - \alpha$$

donde λ_1, λ_2 no dependen de θ.
Las ecuaciones $\lambda_i(\alpha) = T(\boldsymbol{x}, \theta), i = 1, 2$ tienen solución para todo $\boldsymbol{x} \in \Psi$, en particular $\exists \theta_1(\boldsymbol{x}, \alpha), \theta_2(\boldsymbol{x}, \alpha)$ tales que

$$\begin{cases} T(\boldsymbol{x}, \theta_1(\boldsymbol{x}, \alpha)) = \lambda_1(\alpha) \\ T(\boldsymbol{x}, \theta_2(\boldsymbol{x}, \alpha)) = \lambda_2(\alpha) \end{cases}$$

Si $T(\boldsymbol{x}, \theta)$ es creciente en θ, entonces $\theta_1(\boldsymbol{x}, \alpha) \leq \theta_2(\boldsymbol{x}, \alpha)$.
En este caso

$$P(\theta_1(\boldsymbol{x}, \alpha) \leq \theta \leq \theta_2(\boldsymbol{x}, \alpha)) = P(T(\boldsymbol{x}, \theta_1(\boldsymbol{x}, \alpha)) \leq T(\boldsymbol{x}, \theta) \leq T(\boldsymbol{x}, \theta_2(\boldsymbol{x}, \alpha))) \geq 1 - \alpha$$

□

4.3 Método de Neyman

Proposition 4.5. *$X \sim F_\theta, \theta \in \Theta \subset \mathbb{R}$, $\boldsymbol{X} = (X_1, ..., X_n)$ m.a.s. de X y $T(\boldsymbol{X})$ un estadístico. Supongamos que dado $0 < \alpha < 1$, existen $c_1(\alpha, \theta)$ y $c_2(\alpha, \theta)$ tales que*

$$P(c_1(\alpha, \theta) \leq T(\boldsymbol{X}) \leq c_2(\alpha, \theta)) \geq 1 - \alpha, \ \forall \theta \in \Theta$$

Entonces, si $c_1(\alpha, \theta)$ y $c_2(\alpha, \theta)$ son funciones estrictamente monótonas en θ (en el mismo sentido) y continuas, se puede construir un intervalo de confianza al nivel $1 - \alpha$ para θ.

Proof. Supongamos que c_1, c_2 son crecientes en θ.
Sea $t \in Sop(t)$, entonces $\exists \theta_1(t), \theta_2(t) \mid c_1(\alpha, \theta_1(t)) = t, c_2(\alpha, \theta_2(t)) = t$.
Vamos a ver que

$$c_1(\alpha, \theta) \leq t \leq c_2(\alpha, \theta) \iff \theta_2(t) \leq \theta \leq \theta_1(t), \forall \theta \in \Theta$$

$[\implies]$

$$c_1(\alpha, \theta) \leq t = c_1(\alpha, \theta_1(t)) \overset{c_1 \text{ crec}}{\implies} \theta \leq \theta_1(t)$$

$$c_2(\alpha, \theta_2(t)) = t \leq c_2(\alpha, \theta) \overset{c_2 \text{ crec}}{\implies} \theta_2(t) \leq \theta$$

$[\impliedby]$

$$c_1(\alpha, \theta) \leq c_1(\alpha, \theta_1(t)) = t = c_2(\alpha, \theta_2(t)) \leq c_2(\alpha, \theta)$$

Y entonces

$$P(\theta_2(T) \leq \theta \leq \theta_1(T)) = P(c_1(\alpha, \theta) \leq T \leq c_2(\alpha, \theta)) \geq 1 - \alpha$$

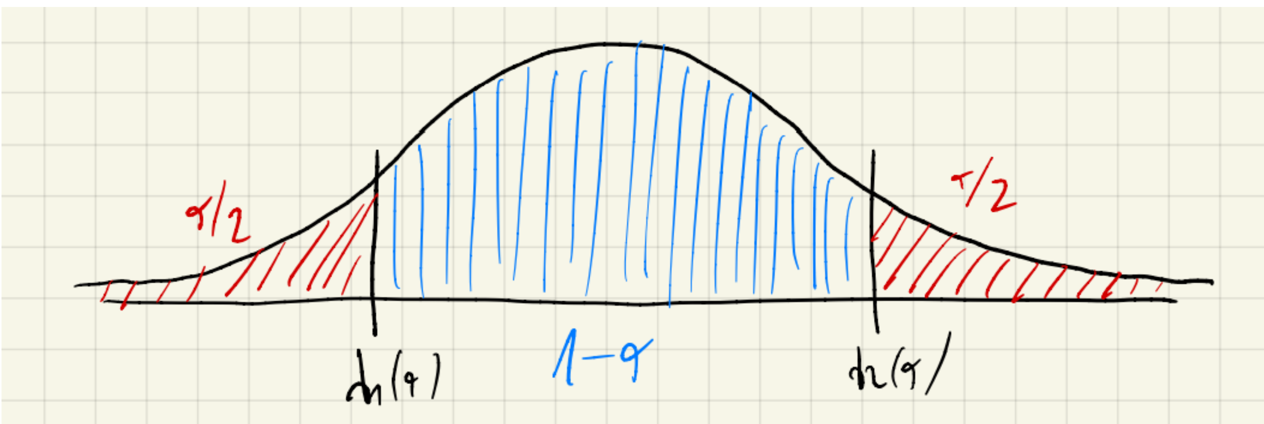
□

4.4 Ejercicios

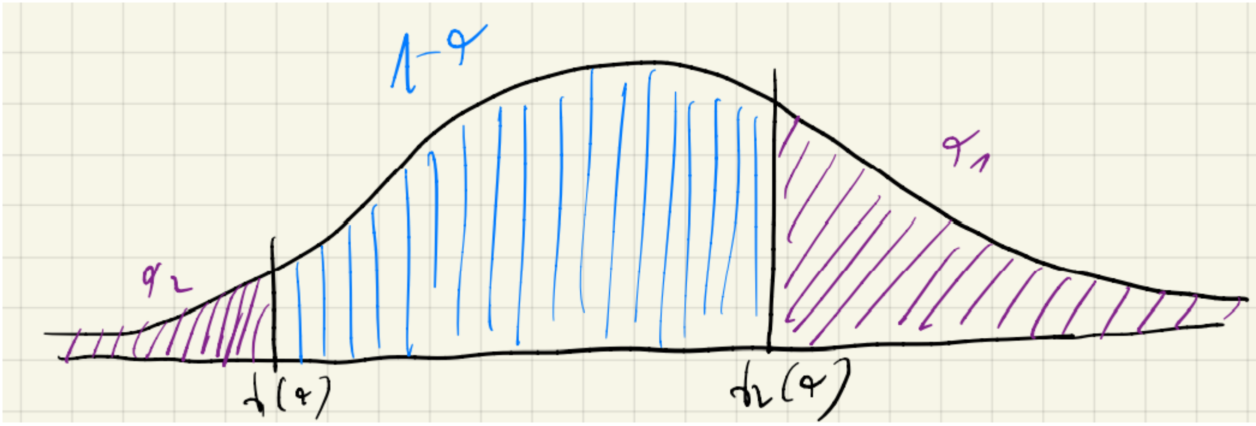
1. Sean $X \sim N(\mu, \sigma^2)$ y $(X_1, ..., X_n)$ m.a.s. de X.
a) Suponiendo σ^2 conocida obtener el intervalo de confianza óptimo para μ con nivel de confianza $1 - \alpha$
Tomamos la función pivote $\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$ y entonces elegimos $\lambda_1(\alpha), \lambda_2(\alpha)$ tales que

$$P\left(\lambda_1(\alpha) \leq \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq \lambda_2(\alpha)\right) = 1 - \alpha$$

Podemos hacer



o tomar $\alpha = \alpha_1 + \alpha_2$ con $\alpha_1 > \frac{\alpha}{2} > \alpha_2 \geq 0$:



La longitud del intervalo es

$$\frac{\sigma}{\sqrt{n}}\left(\lambda_2-\lambda_1\right)$$

por lo que debemos minimizar $\left(\lambda_2-\lambda_1\right)$. Para ello, usamos multiplicadores de Lagrange en

$$\begin{cases}\min & \left(\lambda_2-\lambda_1\right)\\ s.a. & \int_{\lambda_1}^{\lambda_2} f_Z(x)\,dx=1-\alpha\end{cases}$$

Entonces

$$L_g\left(\lambda,\lambda_1,\lambda_2\right)=\lambda\left(\int_{\lambda_1}^{\lambda_2} f_Z(x)\,dx-(1-\alpha)\right)+\left(\lambda_2-\lambda_1\right)$$

de donde

$$\left\{\begin{array}{l}\frac{\partial}{\partial\lambda_1}=-\lambda f_Z\left(\lambda_1\right)-1=0\implies\lambda=-\frac{1}{f_Z\left(\lambda_2\right)}\\ \frac{\partial}{\partial\lambda_2}=\lambda f_Z\left(\lambda_2\right)+1=0\implies\lambda=-\frac{1}{f_Z\left(\lambda_1\right)}\end{array}\right\}\implies f_Z\left(\lambda_1\right)=f_Z\left(\lambda_2\right)\overset{f_Z\ sim\ resp\ 0}{\Longrightarrow}\lambda_1=-\lambda_2$$

Por tanto, la elección que minimizará la longitud del intervalo es la que toma $\alpha_1=\alpha_2=\frac{\alpha}{2}$, o sea

$$\left(\overline{X}-Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}},\overline{X}+Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$$

b) Suponiendo σ^2 desconocida, obtener el intervalo de confianza óptimo para μ con nivel de confianza $1-\alpha$

Usamos S para estimar σ^2, de forma que por las consecuencias del teorema de Fisher:

$$\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}\sim t_{n-1}$$

y entonces el intervalo es

$$\left(\overline{X}-t_{n-1,\frac{\alpha}{2}}\frac{S}{\sqrt{n}},\overline{X}+t_{n-1,1-\frac{\alpha}{2}}\right)$$

óptimo por ser la distribución simétrica respecto del 0.

2. Sean $X\sim N\left(\mu_1,\sigma_1^2\right)$ e $Y\sim N\left(\mu_2,\sigma_2^2\right)$ independientes. Sean $\left(X_1,...,X_{n_1}\right)$ m.a.s. de X y $\left(Y_1,...,Y_{n_2}\right)$ m.a.s. de Y.

a) Suponiendo σ_1^2 y σ_2^2 conocidas obtener el intervalo de confianza óptimo para $\mu_1-\mu_2$ con nivel de confianza $1-\alpha$

b) Suponiendo σ_1^2 y σ_2^2 desconocidas obtener el intervalo de confianza óptimo para $\mu_1 - \mu_2$ con nivel de confianza $1-\alpha$
Este se hace como el anterior, usando la otra consecuencia del teorema de Fisher.

3. Sea $X \sim N\left(\mu, \sigma^2\right)$ y $(X_1, ..., X_n)$ m.a.s. de X. Suponiendo μ desconocida, obtener un intervalo de confianza para σ^2 con nivel de confianza $1-\alpha$

$$\lambda_1 \leq \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq \lambda_2 \iff \sigma\lambda_1 \leq \sqrt{n}\left(\overline{X}-\mu\right) \leq \sigma\lambda_2 \iff \begin{cases} \sigma \leq \frac{\sqrt{n}(\overline{X}-\mu)}{\lambda_1} \\ \sigma \geq \frac{\sqrt{n}(\overline{X}-\mu)}{\lambda_2} \end{cases} \implies \left(\frac{\overline{X}-\mu}{Z_{1-\frac{\alpha}{2}}}\sqrt{n}, \frac{\overline{X}-\mu}{Z_{\frac{\alpha}{2}}}\sqrt{n}\right)$$

que es óptimo, de nuevo, por la simetría de la distribución.

4. Sea X una población con función de densidad

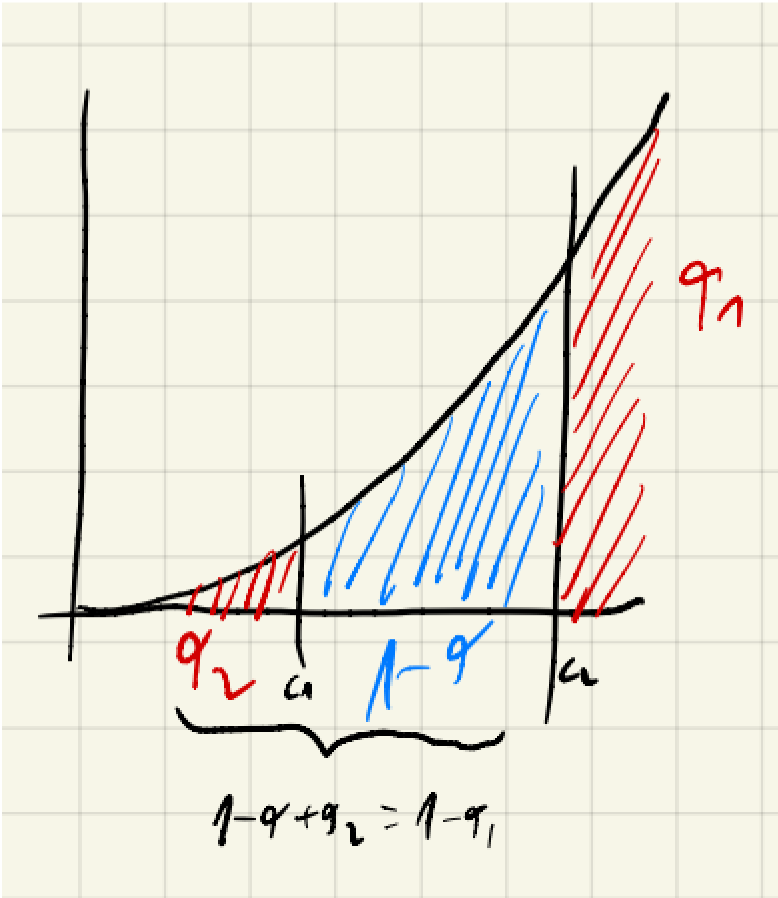
$$f(x,\theta) = \frac{2x}{\theta^2},\ x \in (0,\theta)$$

siendo θ un parámetro positivo desconocido. Sea $(X_1, ..., X_n)$ m.a.s. de X. Obtener el intervalo de confianza óptimo para θ con nivel de confianza $1-\alpha$.
Primero vamos a obtener el estimador de máxima verosimilitud de θ:

$$L(\boldsymbol{x}) = \frac{2^n \prod x_j}{\theta^{2n}} \chi_{(0,\theta)}(X_{n:n}) = \frac{2^n \prod x_j}{\theta^{2n}} \chi_{(X_{n:n},\infty)}(\theta)$$

El EMV es $X_{n:n}$, con distribución $P(X_{n:n} \leq x) = (F(x))^n = \left(\frac{x^2}{\theta^2}\right)^n = \left(\frac{x}{\theta}\right)^{2n}$.

$$1-\alpha = P(c_1(\alpha,\theta) \leq X_{n:n} \leq c_2(\alpha,\theta)) = P(X_{n:n} \leq c_2(\alpha,\theta)) - P(X_{n:n} \leq c_1(\alpha,\theta)) = \left(\frac{c_2(\alpha,\theta)}{\theta}\right)^{2n} - \left(\frac{c_1(\alpha,\theta)}{\theta}\right)^{2n}$$



Hacemos

$$\left(\frac{c_2\left(\alpha,\theta\right)}{\theta}\right)^{2n}=1-\alpha_1$$

$$\left(\frac{c_1\left(\alpha,\theta\right)}{\theta}\right)^{2n}=\alpha_2$$

de donde

$$c_2\left(\alpha,\theta\right)=\left(1-\alpha_1\right)^{\frac{1}{2n}}\theta$$

$$c_1\left(\alpha,\theta\right)=\alpha_2^{\frac{1}{2n}}\theta$$

Por último,

$$c_2\left(\alpha,\theta\right)=t=\left(1-\alpha_1\right)^{\frac{1}{2n}}\theta_1 \iff \theta_1\left(t\right)=\frac{t}{\left(1-\alpha_1\right)^{\frac{1}{2n}}}$$

$$c_1\left(\alpha,\theta\right)=t=\alpha_2^{\frac{1}{2n}}\theta_2 \iff \theta_2\left(t\right)=\frac{t}{\alpha_2^{\frac{1}{2n}}}$$

Y el intervalo de confianza es

$$I=\left(\frac{X_{n:n}}{\alpha_2^{\frac{1}{2n}}},\frac{X_{n:n}}{\left(1-\alpha_1\right)^{\frac{1}{2n}}}\right)$$

Que tiene longitud

$$X_{n:n}\left(\frac{1}{\left(1-\alpha_1\right)^{\frac{1}{2n}}}-\frac{1}{\alpha_2^{\frac{1}{2n}}}\right)=X_{n:n}\left(\frac{1}{\left(1-\alpha_1\right)^{\frac{1}{2n}}}-\frac{1}{\left(\alpha-\alpha_1\right)^{\frac{1}{2n}}}\right)=g\left(\alpha_1\right)$$

y minimizamos g.

5. Sea X una población con función de densidad

$$f(x,\theta) = \exp\left(-\left(x-\theta\right)\right), \ x \in (0,\infty)$$

siendo $\theta > 0$ desconocido. Sea $(X_1, ..., X_n)$ m.a.s. de X. Obtener el intervalo de confianza óptimo para θ con nivel de vonfianza $1-\alpha$.

5 Contrastes de Hipótesis Paramétricas

5.1 Introducción. Tests no aleatorizados. Errores. Tests óptimos

Formulación general del problema de contraste de hipótesis.

Sea $X \sim F(x, \theta)$ con $\theta \in \Theta \subset \mathbb{R}^p$ y sea $\boldsymbol{X}$ una m.a.s. de X. Sean $\Theta_0, \Theta_1 \subset \Theta$ con $\Theta = \Theta_0 \cup \Theta_1$ y $\Theta_0 \cap \Theta_1 = \emptyset$.

Definition 5.1. Un **test** o contraste es una regla basada en $\boldsymbol{X}$ para decidir entre las dos hipótesis:

$$\begin{array}{ll} H_0 : & \theta \in \Theta_0 \\ H_1 : & \theta \in \Theta_1 \end{array}$$

H_0 se denomina **hipótesis nula** y H_1 se denomina **hipótesis** alternativa**.**
Se llama **test no aleatorizado** a una función $\delta : sop(\boldsymbol{X}) \to \{0, 1\}$ tal que si $\delta(\boldsymbol{x}) = 0$ se acepta la hipótesis nula H_0 y si $\delta(\boldsymbol{x}) = 1$ se recahza la hipótesis H_0.
Se representa por T el conjunto de test definidos sobre $sop(\boldsymbol{X})$.

Las hipótesis pueden ser **simples** cuando el espacio paramétrico asociado a la hipótesis es unipuntual o **compuestas**, cuando no lo es. Así, los tests pueden presentar sus hipótesis nula y alternativa en cualquier combinación de estos dos tipos.

Definition 5.2. Se define la **región de aceptación** como las muestras que nos proporcionan la aceptación de la hipótesis nula:

$$S_0 = \{\boldsymbol{x} \in sop(\boldsymbol{X}) : \delta(\boldsymbol{x}) = 0\}$$

y la **región de rechazo** como aquellas muestras en las que se rechaza:

$$S_1 = \{\boldsymbol{x} \in sop(\boldsymbol{X}) : \delta(\boldsymbol{x}) = 1\}$$

Errores

Definition 5.3. Dado un test $\delta \in T$, se llama **error de tipo I** al que se comete al rechazar la hipótesis nula cuando es verdadera.
Se denomina **error de tipo II** al que se comete al aceptar la hipótesis nula cuando es falsa.
Se definen, de forma obvia atendiendo a estas definiciones, las probabilidades de error:
La **probabilidad de cometer un error de tipo I** es

$$P_I(\theta) = P(\boldsymbol{X} \in S_1 | \theta \in \Theta_0)$$

Y la **probabilidad de error de tipo II**:

$$P_{II}(\theta) = P(\boldsymbol{X} \in S_0 | \theta \in \Theta_1)$$

Así, un test ideal sería aquel en que ambos errores son nulos: no fallamos. Pero esto solo es posible si conocemos la población, y no solo una muestra. Por tanto, esto no tiene interés y debemos estudiar un tipo de idealidad u optimalidad restringida a un dominio más útil.

Definition 5.4. Un test $\delta^* \in T$ se denomina **óptimo** si verifica:

1. $P_{I,\delta^*}(\theta) \leq P_{I,\delta}(\theta), \forall\delta \in T, \forall\theta \in \Theta_0$
2. $P_{II,\delta^*}(\theta) \leq P_{II,\delta}(\theta), \forall\delta \in T, \forall\theta \in \Theta_1$

Remark 5.5. De forma general no existe tal óptimo, ya que normalmente, la disminución del error de primer tipo conlleva un aumento en el error de segundo tipo, y viceversa.

Remark 5.6. **Importancia de las hipótesis:** la hipótesis nula y la hipótesis alternativa no desempeñan, en general, un papel simétrico, ya que suele tomarse como hipótesis nula la que representa una situación mantenida anteriormente, o una situación importante. De este modo, el error de tipo I es el más importante de los dos y tiene sentido actuar de modo que se busque acotar la probabilidad de error de tipo I por un valor prefijado y dentro de esta clase restringida de tests, intentar minimizar la probabilidad de error de tipo II. Este razonamiento constituye la base filosófica de la teoría de Neyman y Pearson sobre optimización en test de hipótesis.

5.2 Teoría de Neyman-Pearson

Vamos a denotar por T_α al conjunto de tests que tienen probabilidad de error de tipo I acotado superiormente por α:

$$T_\alpha = \{\delta \in T | P_I(\theta) \leq \alpha, \forall\theta \in \Theta_0\}$$

A α se le llama **nivel de significación del test**.

Definition 5.7. Dado $\delta^* \in T_\alpha$, se dirá que es **óptimo** si

$$P_{II,\delta^*}(\theta) \leq P_{II,\delta}(\theta), \forall\delta \in T_\alpha \forall\theta \in \Theta_1$$

Se llama **tamaño** o **extensión del test** al número dado por

$$\sup_{\theta\in\Theta_0} P_I(\theta)$$

Los test, generalmente, vienen dados como funciones de un estadístico cuya distribución es conocida cuando H_0 es cierta. Lo indicaremos como **estadístico del test** y sirve para medir la diferencia entre los datos y lo que se espera de ellos bajo la hipótesis H_0.

Definition 5.8. Se llama **función potencia de un test $\boldsymbol{\delta}$** a la probabilidad de rechazar la hipótesis nula:

$$\pi_\delta(\theta) = P(\boldsymbol{X} \in S_1, \theta \in \Theta)$$

Remark 5.9. Si $\theta \in \Theta_0$, entonces $\pi_\delta(\theta) = P_{I,\delta}(\theta)$
Si $\theta \in \Theta_1$ entonces

$$\pi_\delta(\theta) = P(\boldsymbol{X} \in S_1, \theta \in \Theta_1) = 1 - P(\boldsymbol{X} \in S_0, \theta \in \Theta_1) = 1 - P_{II,\delta}(\theta)$$

Definition 5.10. Sea $\delta^* \in T_\alpha$. Diremos que es el **test uniforme de máxima potencia para el contraste**

$$\begin{array}{ll} H_0: & \theta \in \Theta_0 \\ H_1: & \theta \in \Theta_1 \end{array}$$

si se verifica

$$\pi_{\delta^*}(\theta) \geq \pi_\delta(\theta), \forall \delta \in T_\alpha, \forall \theta \in \Theta_1$$

Remark 5.11. Se pueden encontrar problemas al fijar el nivel de significación cuando la distribución es discreta.

Definition 5.12. Una vez obtenida la muestra concreta del problema de estudio, se observa el nivel de significación más pequeño para el cual la muestra obligaría a rechazar la hipótesis nula. Este valor se denomina **p-valor**. Puede verse como la probabilidad de rechazar la hipótesis nula para la muestra de la que disponemos, siendo esta cierta.

5.3 Tests de hipótesis nula y alternativa simples

Vamos a tratar el caso en que los subconjuntos del espacio paramétrico son unitarios: $\Theta_0 = \{\theta_0\}$ y $\Theta_1 = \{\theta_1\}$, por lo que los errores de ambos tipos serán cantidades fijas.

Theorem 5.13. ***Teorema de Neyman-Pearson***
Sea $X \sim F(x, \theta), \Theta = \{\theta_0, \theta_1\}, \boldsymbol{X} = (X_1, ..., X_n)$ una m.a.s. de X. Sea $\alpha \in (0,1)$ fijo. Para contrastar la hipótesis nula $H_0: \theta = \theta_0$ frente a la hipótesis alternativa $H_1: \theta = \theta_1$, el test de región crítica:

$$S_1 = \left\{ \boldsymbol{x} \in sop(\boldsymbol{X}) : \frac{L(\boldsymbol{x}, \theta_1)}{L(\boldsymbol{x}, \theta_0)} \geq k \right\}$$

con $k > 0$, que tenga tamaño α, o sea que

$$\alpha = P(\boldsymbol{X} \in S_1 | \theta = \theta_0)$$

cumple que es el test de máxima potencia en la clase de los tests T_α.

Proof. $\delta^* \in T_\alpha$ es el test de máxima potencia si verifica que

$$\pi_{\delta^*}(\theta_1) \geq \pi_\delta(\theta_1), \forall \delta \in T_\alpha$$

Esto es equivalente a que, si S_1 es la región crítica de δ^* y S la de δ, se verifique

$$P(\boldsymbol{X} \in S_1 | \theta = \theta_1) \geq P(\boldsymbol{X} \in S | \theta = \theta_1)$$

$$P(\boldsymbol{X} \in S | \theta = \theta_0) \leq \alpha$$

Sea

$$D = P(\boldsymbol{X} \in S_1 | \theta = \theta_1) - P(\boldsymbol{X} \in S | \theta = \theta_1)$$

Queremos que este valor sea no negativo, para obtener la desigualdad mencionada antes.
Ahora bien:

$$P(\boldsymbol{X} \in S_1 | \theta = \theta_1) = P(\boldsymbol{X} \in S_1 \setminus S | \theta = \theta_1) + P(\boldsymbol{X} \in S_1 \cap S | \theta = \theta_1)$$

$$P(\boldsymbol{X} \in S | \theta = \theta_1) = P(\boldsymbol{X} \in S \setminus S_1 | \theta = \theta_1) + P(\boldsymbol{X} \in S_1 \cap S | \theta = \theta_1)$$

Por tanto
$$D = P\left(\boldsymbol{X} \in S_1 \setminus S | \theta = \theta_1\right) - P\left(\boldsymbol{X} \in S \setminus S_1 | \theta = \theta_1\right)$$
Y es
$$P\left(\boldsymbol{X} \in S_1 \setminus S | \theta = \theta_1\right) = \int_{S_1 \setminus S} L\left(\boldsymbol{x}, \theta_1\right) dx_1 ... dx_n \overset{hipótesis: \frac{L(\boldsymbol{x},\theta_1)}{L(\boldsymbol{x},\theta_0)} \geq k \ en \ S_1}{\geq} k \int_{S_1 \setminus S} L\left(\boldsymbol{x}, \theta_0\right) dx_1 ... dx_n$$
$$P\left(\boldsymbol{X} \in S \setminus S_1 | \theta = \theta_1\right) = \int_{S \setminus S_1} L\left(\boldsymbol{x}, \theta_1\right) dx_1 ... dx_n \overset{\frac{L(\boldsymbol{x},\theta_1)}{L(\boldsymbol{x},\theta_0)} < k \ fuera \ de \ S_1}{<} k \int_{S \setminus S_1} L\left(\boldsymbol{x}, \theta_0\right) dx_1 ... dx_n$$
Juntando estas dos desigualdades, tenemos
$$D \geq k \left[\int_{S_1 \setminus S} L\left(\boldsymbol{x}, \theta_0\right) d\boldsymbol{x} - \int_{S \setminus S_1} L\left(\boldsymbol{x}, \theta_0\right) d\boldsymbol{x}\right]$$
Y entonces
$$D \geq k\left[P\left(\boldsymbol{X} \in S_1 \setminus S | \theta = \theta_0\right) - P\left(\boldsymbol{X} \in S \setminus S_1 | \theta = \theta_0\right)\right] =$$
$$= k\left[P\left(\boldsymbol{X} \in S_1 \setminus S | \theta = \theta_0\right) - P\left(\boldsymbol{X} \in S \setminus S_1 | \theta = \theta_0\right) \pm P\left(\boldsymbol{X} \in S_1 \cap S | \theta = \theta_0\right)\right] =$$
$$= k\left[P\left(\boldsymbol{X} \in S_1 | \theta = \theta_0\right) - P\left(\boldsymbol{X} \in S | \theta = \theta_0\right)\right] = k\left[P_{I,\delta^*}(\theta) - P_{I,\delta}(\theta)\right] \overset{hipótesis}{=} k\left[\alpha - P_{I,\delta}(\theta)\right] \overset{hipótesis: \delta \in T_\alpha}{\geq} 0$$
□

Remark 5.14. La región de aceptación sería
$$S_0 = \left\{\boldsymbol{x} \in sop\left(\boldsymbol{X}\right) | \frac{L\left(\boldsymbol{x}, \theta_1\right)}{L\left(\boldsymbol{x}, \theta_0\right)} < k\right\}$$

Remark 5.15. Si la distribución es de tipo continuo siempre se cumple que el test alcanza el tamaño, obteniéndose el test de extensión α.
Si la distribución es de tipo discreto, en general, el test de máxima potencia será de extensión menor que α. En este caso para obtener el test de extensión α deberíamos considerar test aleatorizados.

Proposition 5.16. *Sea* $X \sim F(x,\theta), \Theta = \{\theta_0, \theta_1\}, \boldsymbol{X}$ *una m.a.s. de* X, $\alpha \in (0,1)$ *fijo. Bajo las condiciones del teorema de Neyman-Pearson, si* $X \sim \varepsilon(1)$, *el test de máxima potencia y extensión* α *para el contraste*

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

viene dado por:
Si $A(\theta)$ *es monótona creciente (decreciente)*

1. *En el caso* $\theta_0 < \theta_1$:
$$\delta(\boldsymbol{x}) = \begin{cases} 0 & T(\boldsymbol{x}) < (>)\, c \\ 1 & T(\boldsymbol{x}) \geq (\leq)\, c \end{cases}$$
2. *En el caso* $\theta_0 > \theta_1$:
$$\delta(\boldsymbol{x}) = \begin{cases} 0 & T(\boldsymbol{x}) > (<)\, c' \\ 1 & T(\boldsymbol{x}) \leq (\geq)\, c' \end{cases}$$

Y según el caso, la constante se obtiene mediante

$$P(T(\boldsymbol{X}) \geq (\leq)\, c | \theta = \theta_0) = \alpha$$

$$P\left(T(\boldsymbol{X}) \leq (\geq)\, c' | \theta = \theta_0\right) = \alpha$$

Proof. Si $X \in \varepsilon(1)$ es porque $L(\boldsymbol{x},\theta) = c(\theta) K(\boldsymbol{x}) \exp\{A(\theta) T(\boldsymbol{x})\}$ con $c(\theta) = \exp\{B(\theta)\}, K(\boldsymbol{x}) = \exp\{h(\boldsymbol{x})\}$. Por el teorema de Neyman-Pearson, el test de máxima potencia y extensión α tiene región crítica

$$S_1 = \left\{\boldsymbol{x} \in sop(\boldsymbol{X}) : \frac{L(\boldsymbol{x},\theta_1)}{L(\boldsymbol{x},\theta_0)} \geq k\right\}$$

con

$$P(\boldsymbol{x} \in S_1 | \theta = \theta_0) = \alpha$$

Por ser $X \in \varepsilon(1)$, se tiene que

$$\frac{L(\boldsymbol{x},\theta_1)}{L(\boldsymbol{x},\theta_0)} = \exp\{[A(\theta_1) - A(\theta_0)]\, T(\boldsymbol{x})\} \exp\{B(\theta_1) - B(\theta_0)\}$$

Luego

$$S_1 = \{\boldsymbol{x} \in sop(\boldsymbol{X}) : \exp\{[A(\theta_1) - A(\theta_0)]\, T(\boldsymbol{x})\} \exp\{B(\theta_1) - B(\theta_0)\} \geq k\}$$

Esto puede reescribirse mediante

$$S_1 = \{\boldsymbol{x} \in sop(\boldsymbol{X}) : \exp\{[A(\theta_1) - A(\theta_0)]\, T(\boldsymbol{x})\} \geq k_1\}$$

con

$$k_1 = \frac{k}{\exp\{B(\theta_1) - B(\theta_0)\}}$$

que es positivo por serlo k. A su vez, podemos seguir simplificando la región crítica:

$$S_1 = \{\boldsymbol{x} \in sop(\boldsymbol{X}) : [A(\theta_1) - A(\theta_0)]\, T(\boldsymbol{x}) \geq k_2\}$$

$$k_2 = \log k_1$$

Hagamos el caso en que $A(\theta)$ es monótona creciente, pero el otro caso es análogo.

1. Si $\theta_0 < \theta_1$, entonces $A(\theta_0) < A(\theta_1) \implies A(\theta_1) - A(\theta_0) > 0$, y la región crítica queda

$$S_1 = \{\boldsymbol{x} \in sop(\boldsymbol{X}) : T(\boldsymbol{x}) \geq c\}$$

$$c = \frac{k_2}{A(\theta_1) - A(\theta_0)}$$

y para calcular c debemos resolver la ecuación

$$P(T(\boldsymbol{X}) \geq c | \theta = \theta_0) = \alpha$$

2. Si $\theta_0 > \theta_1 \implies A(\theta_0) > A(\theta_1) \implies A(\theta_1) - A(\theta_0) < 0$ y queda

$$S_1 = \{\boldsymbol{x} \in sop(\boldsymbol{X}) : T(\boldsymbol{x}) \leq c_1\}$$

$$c_1 = \frac{k_2}{A(\theta_1) - A(\theta_0)}$$

y para obtener c_1 resolvemos

$$P(T(\boldsymbol{X}) \leq c_1 | \theta = \theta_0) = \alpha$$

□

Definition 5.17. Se llama **test aleatorizado** a una función $\delta : sop(\boldsymbol{X}) \to [0,1]$ tal que $\delta(\boldsymbol{x})$ representa la probabilidad de rechazar la hipótesis nula, cuando se observa la muestra $\boldsymbol{x}$ de $sop(\boldsymbol{X})$.
La estructura de los test aleatorizados de interés real son de la forma:

$$\delta(\boldsymbol{x}) = \begin{cases} 0 & \boldsymbol{x} \in S_0 \\ \gamma & \boldsymbol{x} \in S_1' \\ 1 & \boldsymbol{x} \in S_1 = sop(\boldsymbol{X}) \setminus S_0 \setminus S_1' \end{cases}$$

con $0 \leq \gamma \leq 1$.
Si $\sup_{\theta \in \Theta_0} P(\boldsymbol{X} \in S_1) < \alpha$ y $\sup_{\theta \in \Theta_0} P(\boldsymbol{X} \in S_1 \cup S_1') > \alpha$, se obtiene γ con la condición de que la extensión del test es α:

$$\sup_{\theta \in \Theta_0} \left\{ P(\boldsymbol{X} \in S_1) + \gamma P\left(\boldsymbol{X} \in S_1'\right) \right\} = \alpha$$

Proposition 5.18. *Util para los ejercicios.*
Si $X \sim \gamma(a,p)$ *entonces* $\psi = 2aX \sim \chi^2_{2p}$

6 Contrastes de Hipótesis Compuestas

6.1 Introducción

En las aplicaciones a problemas reales no suelen plantearse contrastes de hipótesis simple y alternativa simple, pues las hipótesis alternativas no suelen ser tan precisas para que estén definidas por un único valor del parámetro. Vamos a ver los contrastes:

1. $\begin{array}{ll} H_0: & \theta = \theta_0 \\ H_1: & \theta > \theta_0 \end{array}$

2. $\begin{array}{ll} H_0: & \theta = \theta_0 \\ H_1: & \theta < \theta_0 \end{array}$

3. $\begin{array}{ll} H_0: & \theta \leq \theta_0 \\ H_1: & \theta > \theta_0 \end{array}$

4. $\begin{array}{ll} H_0: & \theta \geq \theta_0 \\ H_1: & \theta < \theta_0 \end{array}$

5. $\begin{array}{ll} H_0: & \theta = \theta_0 \\ H_1: & \theta \neq \theta_0 \end{array}$

6. $\begin{array}{ll} H_0: & \theta \in [\theta_1, \theta_2] \\ H_1: & \theta \notin [\theta_1, \theta_2] \end{array}$

6.2 Familias con cociente de verosimilitud monótono. Contrastes unilaterales

En los casos 1-4 anteriores es posible obtener tests uniformes de máxima potencia, siempre que nuestro modelo tenga la propiedad de 'cociente de verosimilitud monótono':

Definition 6.1. Sea $X \sim F(x, \theta), \theta \in \Theta \subset \mathbb{R}$ y $\boldsymbol{X}$ una m.a.s. de X. Decimos que X o F tienen la **propiedad de cociente de verosimilitud monótono en el estadístico** $R = R(\boldsymbol{X})$ si para todo $\theta_0, \theta_1 \in \Theta$ con $\theta_0 < \theta_1$ se verifica que el cociente de verosimilitudes

$$\frac{L(\boldsymbol{x}, \theta_1)}{L(\boldsymbol{x}, \theta_0)}$$

es creciente en $R(\boldsymbol{x})$.

Proposition 6.2. *Sea $X \sim F(x, \theta)$. Si $X \in \varepsilon(1)$ para el estadístico $T = T(\boldsymbol{X})$ con $A(\theta)$ monótona, entonces se verifica:*

1. *Si $A(\theta)$ es monótona creciente en θ, X tiene cociente de verosimilitud monótono en el estadístico T*

2. *Si $A(\theta)$ es monótona decreciente en θ, X tiene cociente de verosimilitud monótono en el estadístico $-T$*

Proof.

$$\frac{L(\boldsymbol{x}, \theta_1)}{L(\boldsymbol{x}, \theta_0)} = \exp\left\{\left[A(\theta_1) - A(\theta_0) T(\boldsymbol{x})\right]\right\} \exp\left\{B(\theta_1) - B(\theta_0)\right\} = g(T)$$

1. Si A es monótona creciente, entonces $\theta_0 < \theta_1 \implies A(\theta_1) - A(\theta_0) = d > 0 \implies g(T) = c\exp\{dT\} \implies g'(T) = cd\exp\{dT\} > 0$ por lo que g es creciente y X tiene CVM en el estadístico T

2. Si A es monótona decreciente, entonces $\theta_0 < \theta_1 \implies d < 0$ y la derivada sale negativa. Por tanto, $h(T) = g(-T)$ es creciente.

□

Theorem 6.3. *Sea $X \sim F(x, \theta), \theta \in \Theta \subset \mathbb{R}$ y $\boldsymbol{X}$ una m.a.s. de X. Si X tiene cociente de verosimilitud monótono en el estadístico R para contrastar*

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

se sigue que el test

$$\delta^*(\boldsymbol{x}) = \begin{cases} 0 & R(\boldsymbol{x}) \leq c \\ 1 & R(\boldsymbol{x}) > c \end{cases}$$

$$P_{I,\delta}(\theta_0) = \alpha$$

es el test uniforme de máxima potencia en la clase de tests $T_\alpha = \{\delta \in T | P_I(\theta_0) \leq \alpha\}$ con $\alpha \in (0, 1)$.

Proof. δ^* es el TUMP para el contraste anterior si

$$\pi_{\delta^*}(\theta) \geq \pi_\delta(\theta), \forall \delta \in T_\alpha, \theta > \theta_0$$

Vamos a basarnos en el test de hipótesis simple y alternativa simple.
Consideremos $\theta_1 > \theta_0$ y el contraste

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

que, por el Teorema de Neyman Pearson, el test de máxima potencia tiene región crítica

$$S_1 = \left\{ \boldsymbol{x} \in sop(\boldsymbol{X}) : \frac{L(\boldsymbol{x}, \theta_1)}{L(\boldsymbol{x}, \theta_0)} > k \right\}$$

Como X tiene cociente de verosimilitud monótono en el estadístico R, entonces será equivalente a

$$S_1' = \{\boldsymbol{x} \in sop(\boldsymbol{X}) : R(\boldsymbol{x}) > c\}$$

$$P(R(\boldsymbol{x}) > c | \theta = \theta_0) = \alpha$$

y como esto es independiente de θ_1, el test sirve para todo $\theta > \theta_0$ y será el TUMP para el contraste unilateral del enunciado. □

Theorem 6.4. *Sea* $X \sim F(x,\theta), \theta \in \Theta \subset \mathbb{R}$ *y* $\boldsymbol{X}$ *una m.a.s. de* X*. Si* X *tiene cociente de verosimilitud monótono en el estadístico* R *para contrastar*

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

se sigue que el test

$$\delta^*(\boldsymbol{x}) = \begin{cases} 0 & R(\boldsymbol{x}) \geq c \\ 1 & R(\boldsymbol{x}) < c \end{cases}$$

$$P_{I,\delta}(\theta_0) = \alpha$$

es el test uniforme de máxima potencia en la clase de tests $T_\alpha = \{\delta \in T | P_I(\theta_0) \leq \alpha\}$ *con* $\alpha \in (0,1)$.

6.3 Test de la razón de verosimilitudes generalizado

Definition 6.5. Se denomina **razón de verosimilitudes generalizado** para contrastar

$$\begin{aligned} H_0: &\quad \theta \in \Theta_0 \\ H_1: &\quad \theta \in \Theta_1 \end{aligned}$$

al cociente

$$\lambda(\boldsymbol{x}) = \frac{\sup_{\theta \in \Theta_0} L(\boldsymbol{x},\theta)}{\sup_{\theta \in \Theta} L(\boldsymbol{x},\theta)}$$

Un **test de razón de verosimilitudes generalizado** para el contraste anterior es aquel que tiene región crítica de la forma

$$S_1 = \{\boldsymbol{x} \in sop(\boldsymbol{X}) : \lambda(\boldsymbol{x}) < k\}$$

lo que indica que rechazamos H_0 cuando el cociente toma valores a un valor $k \in (0,1)$.
Dicha constante k se determina con la condición de que el test tenga extensión α:

$$\sup_{\theta \in \Theta_0} P(\boldsymbol{X} \in S_1 | \theta) = \alpha$$

Remark 6.6. La principal dificultad en la construcción de este test es la obtención de la distribución del estadístico λ.
La idea del método se basa en que para cada muestra fija, la función de verosimilitud es un indicador de lo bien que explica el valor del parámetro los resultados obtenidos.

Remark 6.7. Para el contraste

$$\begin{aligned} H_0: &\quad \theta \in \Theta_0 \\ H_1: &\quad \theta \in \Theta_1 \end{aligned}$$

si $dim(\Theta) = q$ y $dim(\Theta_0) = q'$, se verifica, bajo ciertas condiciones, que el estadístico $-2\log\lambda(\boldsymbol{X})$ converge en distribución a una $\chi^2_{q-q'}$.

Definition 6.8. Un test es un **test insesgado** si la probabilidad de rechazar H_0 cuando es verdadera es menor que la probabilidad de rechazarla cuando es falsa:

$$P(\boldsymbol{X} \in S_1 | \theta \in \Theta_0) \leq P(\boldsymbol{X} \in S_1 | \theta \in \Theta_1)$$

6.4 Ejercicios de los capítulos 5 y 6 (6 y 7 de la asignatura)

1. Sea X una población con función de densidad

$$f(x,\theta) = \frac{2\theta}{1-\theta} x^{\frac{3\theta-1}{1-\theta}} \chi_{(0,1)}(x)$$

donde $\theta \in (0,1)$ es un parámetro desconocido. Sea $X_1, ..., X_n$ una muestra aleatoria simple de X, se pide:
a) Obtener el test de máxima potencia y extensión α para el contraste $H_0 : \theta = \theta_0$ frente a $H_1 : \theta = \theta_1$, con $\theta_0 > \theta_1$
b) Si $n = 20, \alpha = 0.05, \theta_0 = \frac{1}{2}$ y $\sum_{i=1}^{20} \log(x_i) = -6.21$, obtener el p-valor para dicha muestra. ¿Qué criterio de decisión se tomaría?
c) Si para el contraste anterior, $\theta_0 = \frac{1}{2}$, $\theta_1 = \frac{1}{4}$ y las probabilidades de error de primer y segundo tipo son iguales a 0.05, obtener el tamaño muestral y la región crítica de dicho test.
a) Por el **teorema de Neymann-Pearson**, la región crítica del contraste es

$$S_1 = \left\{ \boldsymbol{X} \in (0,1)^n : \frac{L(\boldsymbol{X}, \theta_1)}{L(\boldsymbol{X}, \theta_0)} > k \right\}$$

con $k > 0$ y

$$P(\boldsymbol{X} \in S_1 | \theta = \theta_0) = \alpha$$

Ahora bien,

$$L(\boldsymbol{X}, \theta) = \prod_1^n f(x_i, \theta) = \frac{2^n \theta^n}{(1-\theta)^n} \prod_1^n x_j^{\frac{3\theta-1}{1-\theta}}, \ 0 < x_j < 1, \forall j = 1, ..., n$$

Así

$$\frac{L(\boldsymbol{X}, \theta_1)}{L(\boldsymbol{X}, \theta_0)} = \left(\frac{\theta_1}{1-\theta_1} \frac{1-\theta_0}{\theta_0} \right)^n \prod_1^n x_j^{\frac{3\theta_1-1}{1-\theta_1} - \frac{3\theta_0-1}{1-\theta_0}}, \ 0 < x_j < 1, \forall j$$

El exponente da:

$$\frac{(3\theta_1-1)(1-\theta_0) - (3\theta_0-1)(1-\theta_1)}{(1-\theta_1)(1-\theta_0)} = \frac{3\theta_1 - 3\theta_1\theta_0 - 1 + \theta_0 - 3\theta_0 + 3\theta_1\theta_0 + 1 - \theta_1}{(1-\theta_0)(1-\theta_1)} = \frac{2(\theta_1-\theta_0)}{(1-\theta_1)(1-\theta_0)}$$

Además, como $\theta_0 > \theta_1$ y $\theta \in (0,1)$, se tiene que $A = \left(\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)} \right)^n > 0$. Por tanto, la región crítica anterior es equivalente a

$$S_1 = \left\{ \boldsymbol{X} \in (0,1)^n : \prod_1^n x_j^{\frac{2(\theta_1-\theta_0)}{(1-\theta_0)(1-\theta_1)}} > k_1 \right\}, \ k_1 = \frac{k}{A} > 0$$

Tomando logaritmos:

$$S_1 = \left\{ \boldsymbol{X} \in (0,1)^n : \frac{2(\theta_1-\theta_0)}{(1-\theta_1)(1-\theta_0)} \sum_1^n \log x_j > k_2 \right\}, \ k_2 = \log k_1$$

Y el término $B = \frac{2(\theta_1-\theta_0)}{(1-\theta_1)(1-\theta_0)}$ es negativo, puesto que $\theta_0 > \theta_1$. Así, queda

$$S_1 = \left\{ \boldsymbol{X} \in (0,1)^n : \sum_1^n \log x_j < k_3 \right\},\ k_2 = \frac{k_2}{B}$$

Y debe ser

$$P\left(\sum_1^n \log x_j < k_3 | \theta = \theta_0 \right) = \alpha$$

Y estamos interesados ahora en la distribución de esta suma de logaritmos. Si definimos la variable aleatoria $Y = -\sum_1^n \log X_j$ y $c = -k_3$, entonces

$$S_1 = \{ \boldsymbol{X} \in (0,1)^n : Y > c \}$$

$$P\left(Y > c | \theta = \theta_0\right) = \alpha \iff P\left(Y < c | \theta = \theta_0\right) = 1 - \alpha \iff H_Y\left(c, \theta = \theta_0\right) = 1 - \alpha$$

con H_Y la función de distribución de Y.
Ahora bien, $Y_j = -\log X_j$, de forma que $sop\left(Y_j\right) = (0, \infty)$, $X_j = e^{-Y_j}$ y entonces

$$h_{Y_j}\left(y\right) = f\left(x\left(y\right)\right)\left|x'\left(y\right)\right|\chi_{(0,\infty)}\left(y\right) = \frac{2\theta}{1-\theta}\left(e^{-y}\right)^{\frac{3\theta-1}{1-\theta}} e^{-y}\chi_{(0,\infty)}\left(y\right) = \frac{2\theta}{1-\theta}\left(e^{-y}\right)^{\frac{2\theta}{1-\theta}}\chi_{(0,\infty)}\left(y\right) \sim Exp\left(\frac{2\theta}{1-\theta}\right)$$

Y, usando el método de la función característica:

$$\varphi_{Y_j}\left(t\right) = \left(1 - \frac{it}{\frac{2\theta}{1-\theta}}\right)^{-1} \implies \varphi_Y\left(t\right) = \prod_1^n \varphi_{Y_j}\left(t\right) = \left(1 - \frac{it}{\frac{2\theta}{1-\theta}}\right)^{-n} \sim \gamma\left(\frac{2\theta}{1-\theta}, n\right)$$

Y entonces

$$h_Y\left(y\right) = \left(\frac{2\theta}{1-\theta}\right)^n \frac{1}{\Gamma\left(n\right)} y^{n-1} e^{-\frac{2\theta}{1-\theta}}\chi_{(0,\infty)}\left(y\right)$$

Por lo que

$$H_Y\left(c, \theta = \theta_0\right) = 1 - \alpha \iff \int_0^c h_Y\left(y, \theta = \theta_0\right) dy = 1 - \alpha$$

Usamos ahora la proposición 5.18, definiendo

$$U = 2\frac{2\theta}{1-\theta}Y = \frac{4\theta}{1-\theta}Y \sim \chi^2_{2n}$$

de forma que

$$P\left(Y < c | \theta = \theta_0\right) = 1 - \alpha \iff P\left(U < \frac{4\theta_0}{1-\theta_0}c\right) = 1 - \alpha \iff \frac{4\theta_0}{1-\theta_0} = \chi^2_{2n,1-\alpha}$$

Por lo que

$$c = \frac{\left(1-\theta_0\right)\chi^2_{2n,1-\alpha}}{4\theta_0}$$

Obtenemos, entonces, el **criterio de decisión** siguiente:
Si tomamos una muestra $x_1, ..., x_n$, calculamos el valor experimental $Y_{exp} = -\sum_1^n \log x_j$ y:

$$\begin{cases} si\ Y_{exp} < c & \implies Aceptamos\ H_0 \\ si\ Y_{exp} > c & \implies Rechazamos\ H_0 \end{cases}$$

b)

$$p = P\left(Y > 6,21|\theta = \frac{1}{2}\right) \overset{U=\frac{4\frac{1}{2}y}{1-\frac{1}{2}}=4y}{=} P\left(U > 4 \cdot 6.21\right) = 1 - P\left(U \leq 24,84\right) \overset{tabla\ \chi^2_{40}}{\sim} 1 - 0.025 = 0.975$$

Como $p > \alpha$, aceptamos H_0.
Otra forma de tomar la decisión es calcular el c y compararlo con el valor de Y_{exp}.
c)

$$P_{I\delta^*} = 0.05 = P\left(Y > c|\theta = \frac{1}{2}\right) = P\left(U > 4c\right) \iff P\left(U \leq 4c\right) = 0.95 \iff 4c = \chi^2_{2n,0.95}$$

$$P_{II\delta^*} = 0.05 = P\left(Y < c|\theta = \frac{1}{4}\right) = P\left(U < \frac{4}{3}c\right) \iff \frac{4}{3}c = \chi^2_{2n,0.05} \iff 4c = 3\chi^2_{2n,0.05}$$

Y buscamos en la tabla de la chi cuadrado, valores de n tales que

$$\chi^2_{2n,0.95} = 3\chi^2_{2n,0.05}$$

Sale que $n = 10\ u\ 11$, luego $c = \frac{\chi^2_{21,0.954}}{4}$ y

$$S_1 = \left\{\boldsymbol{x} \in (0,1)^n : -\sum_1^n x_j > c\right\}$$

2. Sea X una población con función de densidad

$$f\left(x,\theta\right) = \frac{2}{\left(x-\theta\right)^3}\chi_{(\theta+1,\infty)}\left(x\right)$$

Se considera una m.a.s. de tamaño 1. Obtener el test de máxima potencia y nivel de significación α para el contraste $H_0 : \theta = \theta_0$ frente a $H_1 : \theta = \theta_1$ con $\theta_0 < \theta_1$.
Calculemos el cociente de verosimilitud:

$$\frac{L\left(\boldsymbol{x},\theta_1\right)}{L\left(\boldsymbol{x},\theta_0\right)} = \frac{f\left(x,\theta_1\right)}{f\left(x,\theta_0\right)} = \left(\frac{x-\theta_0}{x-\theta_1}\right)^3 \frac{\chi_{(\theta_1+1,\infty)}\left(x\right)}{\chi_{(\theta_0+1,\infty)}\left(x\right)}$$

El cociente es indeterminado para $x < \theta_0+1$, es 0 si $x \in \left(\theta_0+1, \theta_1+1\right)$ y $\left(\frac{x-\theta_0}{x-\theta_1}\right)^3 > 0$ para $x > \theta_1+1$. El signo positivo se debe a que $x > \theta_1 + 1 > \theta_1 > \theta_0$. Así, podemos escribir el cociente como

$$\frac{f\left(x,\theta_1\right)}{f\left(x,\theta_0\right)} = \left(\frac{x-\theta_0}{x-\theta_1}\right)^3 \chi_{(\theta_1+1,\infty)}\left(x\right)$$

Y, por el teorema de Neyman-Pearson, tenemos la región crítica

$$S_1 = \left\{x > \theta_1 + 1 : \left(\frac{x-\theta_1}{x-\theta_0}\right)^3 > c\right\} = \left\{x > \theta_1 + 1 : \frac{x-\theta_0}{x-\theta_1} > c_1\right\},\ c_1 = \sqrt[3]{c} > 0$$

con

$$P\left(\frac{x-\theta_0}{x-\theta_1} > c_1|\theta = \theta_0\right) = \alpha$$

Sea la función

$$g\left(x\right) = \frac{x-\theta_0}{x-\theta_1},\ x > \theta_1 + 1$$

entonces

$$g'(x) = \frac{x-\theta_1 - x + \theta_0}{(x-\theta_0)^2} = \frac{\theta_0 - \theta_1}{(x-\theta_0)^2} < 0 \implies g \; decreciente$$

Es decir, que $g(a) < g(b) \iff a < b$, y entonces

$$S_1 = \{x > \theta_1 + 1 : x < c_2\}, c_2 = g^{-1}(c_1)$$

con

$$P(x < c_2 | \theta = \theta_0) = \alpha$$

Pero

$$P(x < c_2|\theta = \theta_0) = \int_{-\infty}^{c_2} f(t,\theta_0)\,dt = \int_{\theta_1+1}^{c_2} f(t,\theta_0)\,dt = \int_{\theta_1+1}^{c_2} \frac{2}{(t-\theta_0)^3} dt = 2\int_{\theta_1+1}^{c_2} (t-\theta_0)^{-3}\,dt =$$

$$= 2\frac{(t-\theta_0)}{-2}|_{\theta_1+1}^{c_2} = -\frac{1}{(c_2-\theta_0)^2} + \frac{1}{(\theta_1+1-\theta_0)^2} = \alpha$$

$$\iff \frac{1}{(c_2-\theta_0)^2} = \frac{1}{(\theta_1+1-\theta_0)^2} - \alpha \iff c_2 = \theta_0 + \frac{1}{\sqrt{\frac{1}{(\theta_1+1-\theta_0)^2} - \alpha}}$$

Y el **criterio de decisión** es el siguiente:
Si tenemos una m.a.s. x:

$$\begin{cases} x < c_2 & \implies Rechazo \; H_0 \\ x > c_2 & \implies Acepto \; H_0 \end{cases}$$

3. Sea X una población con función de densidad gamma:

$$f(x,\theta) = \theta^2 x \exp(-\theta x)$$

con $x > 0,\ \theta > 0$.
a) Estudiar si el modelo anterior tiene cociente de verosimilitud monótono en algún estadístico
b) En caso afirmativo, obtener el test uniforme de máxima potencia y extensión α para el contraste $H_0 : \theta = \theta_0$ frente a $H_1 : \theta > \theta_0$
c) Si $n = 20,\ \alpha = 0.05,\ \theta_0 = 1,\ \overline{X} = 1.2$, ¿qué criterio de decisión se tomaría? Obtener el p-valor para dicha muestra.
Si n es suficientemente grande,
d) Obtener el test asintótico del apartado b), mediante el teorema central del límite
e) Obtener el test de razón de verosimilitudes generalizado y extensión α, para el contraste $H_0 : \theta = 2$ frente a la hipótesis $H_1 : \theta \neq 2$
a) Si vemos que está en la familia exponencial uniparamétrica para algún estadístico y la $A(\theta)$ es monótona, tendremos una respuesta afirmativa a la pregunta, por la proposición 6.2:

$$L(\boldsymbol{x},\theta) = \theta^n \prod_1^n x_j \cdot \exp\left(-\theta \sum_1^n x_j\right),\ x_j > 0, \forall j$$

y claramente está en la familia exponencial uniparamétrica, con

$$A(\theta) = -\theta \qquad T(\boldsymbol{X}) = \sum_1^n X_j \qquad C(\theta) = \theta^n \qquad k(\boldsymbol{X}) = \prod_1^n x_j$$

y A es monóton decreciente, por lo que X tiene cociente de verosimilitud monótono en el estadístico $R(\boldsymbol{X}) = -T(\boldsymbol{X})$.

b) Por el teorema 6.3, el test uniforme de máxima potencia tiene región crítica

$$S_1 = \left\{ \boldsymbol{x} \in \mathbb{R}^n : -\sum_1^n x_j > c \right\} = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \sum_1^n x_j < k \right\}, \; k = -c$$

con

$$P\left(\sum_1^n X_j < k | \theta = \theta_0 \right) = \alpha$$

Como $X \sim \gamma(\theta, 2) \implies T = \sum_1^n X_j \sim \gamma(\theta, 2n) \implies U = 2\theta T \sim \chi^2_{4n}$, y entonces la probabilidad anterior es equivalente a

$$P(U < 2\theta_0 k) = \alpha \implies 2\theta_0 k = \chi^2_{4n,\alpha} \implies k = \frac{\chi^2_{4n,\alpha}}{2\theta_0}$$

Y el **criterio de decisión** es:
m.a.s. $x_1, ..., x_n$, calculamos $T_{exp} = \sum x_j$:

$$\begin{cases} T_{exp} > \frac{\chi^2_{4n,\alpha}}{2\theta_0} & \textit{Acepto } H_0 \\ T_{exp} < \frac{\chi^2_{4n,\alpha}}{2\theta_0} & \textit{Rechazo } H_0 \end{cases}$$

c) $T_{exp} = 1.2 \cdot 20 = 24$

$$\frac{\chi^2_{4n,\alpha}}{2\theta_0} = \frac{\chi^2_{80,0.05}}{2} \simeq \frac{60.3915}{2} = 30.1935705$$

Por lo que rechazamos H_0.
d) Como X es una gamma, tiene media $\frac{2}{\theta}$ y varianza $\frac{2}{\theta^2}$, así, aplicando el teorema central del límite, tenemos que

$$Y_n = \frac{\sum_1^n X_j - n\frac{2}{\theta}}{\sqrt{\frac{2}{\theta^2} n}} \xrightarrow{d} N(0,1)$$

Y tendremos

$$S_1 = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \sum x_j < l \right\}$$

con

$$P\left(\sum X_j < k | \theta = \theta_0 \right) = \alpha \iff P\left(Y_n < \frac{k - n\frac{2}{\theta_0}}{\sqrt{\frac{2}{\theta_0^2} n}} \right) = \alpha \iff \frac{k - n\frac{2}{\theta_0}}{\sqrt{\frac{2}{\theta_0^2} n}} = Z_\alpha \iff k = \frac{\sqrt{2n}}{\theta_0} Z_\alpha + \frac{2n}{\theta_0}$$

Obteniendo el **criterio de decisión**:

$$\begin{cases} \sum x_j > k & \textit{Acepto } H_0 \\ \sum x_j < k & \textit{Rechazo } H_0 \end{cases}$$

d) Según la definición del test de razón de verosimilitudes generalizado con extensión α, tenemos que obtener

$$\lambda(\boldsymbol{x}) = \frac{\sup_{\theta \in \Theta_0} L(\boldsymbol{x}, \theta)}{\sup_{\theta \in \Theta} L(\boldsymbol{x}, \theta)} = \frac{L(\boldsymbol{x}, \theta_0)}{L\left(\boldsymbol{x}, \hat{\theta}\right)}$$

ya que $\Theta_0 = \{\theta_0\}$ y siendo $\hat{\theta}$ EMV de θ.
Veamos cuál es el EMV:

$$\log L(\boldsymbol{x}, \theta) = 2n \log \theta + \sum \log x_j - \theta \sum x_j$$

derivando

$$\frac{\partial}{\partial\theta}\log L = \frac{2n}{\theta} - \sum x_j = 0 \iff \hat{\theta} = \hat{\theta}\left(\boldsymbol{x}\right) = \frac{2n}{\sum x_j}$$

O sea

$$\lambda\left(\boldsymbol{x}\right) = \frac{\theta_0^{2n}\exp\left(-\theta_0\sum x_j\right)}{\left(\frac{2n}{\sum x_j}\right)^{2n}\exp\left(-\frac{2n}{\sum x_j}\sum x_j\right)} = \frac{\theta_0^{2n}\exp\left(2n\right)\left(\sum x_j\right)^{2n}\exp\left(-\theta_0\sum x_j\right)}{\left(2n\right)^{2n}}$$

La región crítica del contraste es la del test de razón de verosimilitudes generalizado:

$$S_1 = \left\{\boldsymbol{x}\in\mathbb{R}^{+^n} : \lambda\left(\boldsymbol{x}\right) < c\right\},\ c\in(0,1)$$

$$P\left(\lambda\left(\boldsymbol{x}\right) < c|\theta=\theta_0\right) = \alpha$$

Ahora bien

$$\lambda < c \iff \left(\sum x_j\right)^{2n}\exp\left(-\theta_0\sum x_j\right) < c_1,\ c_1 = \frac{c\left(2n\right)^{2n}}{\theta_0^{2n}\exp\left(2n\right)} > 0$$

$$\iff \sum x_j\exp\left(-\frac{\theta_0}{2n}\sum x_j\right) < c_2,\ c_2 = c_1^{\frac{1}{2n}}$$

Tomamos el estadístico $T = \sum_1^n X_j$ y consideramos

$$g\left(t\right) = t\exp\left(-\frac{\theta_0}{2n}t\right)$$

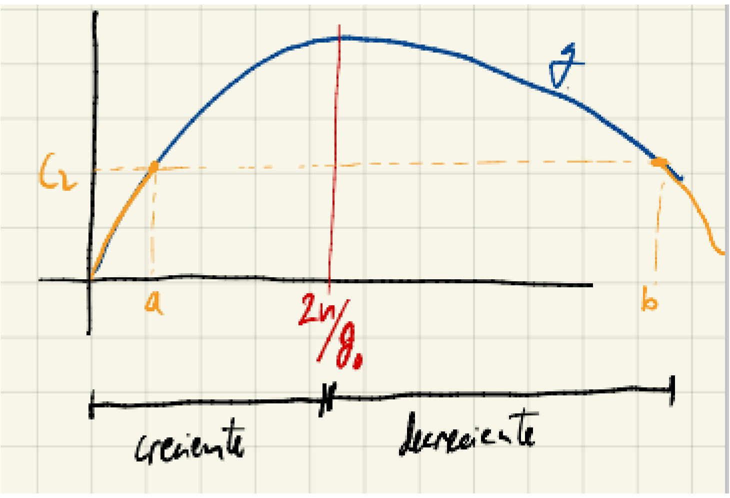
$$g'\left(t\right) = \exp\left(-\frac{\theta_0}{2n}t\right) - \frac{\theta_0}{2n}t\exp\left(-\frac{\theta_0}{2n}t\right) = \exp\left(-\frac{\theta_0}{2n}t\right)\left[1-\frac{\theta_0}{2n}t\right]$$

$$g'\left(t\right) = 0 \iff t = \frac{2n}{\theta_0}$$

y este valor es un máximo. Es decir, g presenta un máximo en ese valor. Así, la región crítica queda como:

$$S_1 = \left\{\boldsymbol{x}\in\mathbb{R}^{+^n} : g\left(T\left(\boldsymbol{x}\right)\right) < c_2\right\} = \left\{\boldsymbol{x}\in\mathbb{R}^{+^n} : \left(0 < T < a\right)\cup\left(t > b\right)\right\}$$

donde a y b son los valores de t para los que g alcanza c_2:



Y es

$$P(0<T<a|\theta=\theta_0)+P(T>b|\theta=\theta_0)=\alpha \overset{U=2\theta T\sim\chi^2_{4n}}{\Longrightarrow} P(0<U<2\theta_0 a)+P(U>2\theta_0 b)=\alpha$$

Tomamos $\alpha_1+\alpha_2=\alpha$ y $P(0<U<2\theta_0 a)=\alpha_1$, $P(U>2\theta_0 b)=\alpha_2$, por lo que es

$$2\theta_0 a=\chi^2_{4n,\alpha_1} \implies a=\frac{\chi^2_{4n,\alpha_1}}{2\theta_0}$$

$$2\theta_0 b=\chi^2_{4n,1-\alpha_2} \implies b=\frac{\chi^2_{4n,1-\alpha+\alpha_1}}{2\theta_0}$$

y faltaría usar estos datos para resolver $g(a)=g(b)$.
4. Sea X una población con función de densidad

$$f(x,\theta)=\frac{1}{\theta(1-x)^2}\exp\left(-\frac{x}{\theta(1-x)}\right)\chi_{(0,1)}(x)$$

donde θ es un parámetro desconocido, estrictamente positivo. Sea $X_1,...,X_n$ una m.a.s. de X.
a) Estudiar si X tiene cociente de verosimilitud monótono en algún estadístico
b) Si X tiene cociente de verosimilitud monótono, obtener el test uniforme de máxima potencia y extensión α, para el contraste $H_0:\theta\geq\theta_0$ frente a $H_1:\theta<\theta_0$
c) Si consideramos una muestra de tamaño $n=25$ y $\theta_0=0.2,\alpha=0.05$ y $\sum_1^{25}\frac{x_i}{1-x_i}=6.23$, obtener el p-valor de dicha muestra. Explica de dos formas diferentes qué criterio de decisión se tomaría en este caso.
d) Urilizando el test de la razón de verosimilitudes generalizado, obtener la región crítica del contraste $H_0:\theta=\frac{1}{2}$ frente a $H_1:\theta\neq\frac{1}{2}$.
a) Sí, pues tomando $T(\boldsymbol{X})=\sum_1^n\frac{x_i}{1-x_i}$, $X\in\varepsilon(1)$.
b) Como tiene cociente de verosimilitud monótono para T, el TUMP tiene región crítica

$$S_1=\{\boldsymbol{X}\in Sop(\boldsymbol{X}):T(\boldsymbol{X})<c\}$$

con

$$\sup P(T(\boldsymbol{X})<c|\theta\geq\theta_0)=\alpha$$

Definimos, para $\theta\geq\theta_0$:

$$g(\theta)=P(T(\boldsymbol{X})<c|\theta)$$

Como $T_j=\frac{X_j}{1-X_j}$, si hacemos la cuenta sale que $T_j\sim Exp\left(\frac{1}{\theta}\right)$, y entonces $T=\sum T_j\sim\gamma\left(\frac{1}{\theta},n\right)$, luego

$$P(T(\boldsymbol{X})<c|\theta)=\int_0^c\frac{\frac{1}{\theta^n}}{\Gamma(n)}x^{n-1}e^{-\frac{x}{\theta}}dx=\frac{1}{\theta^n(n-1)!}\int_0^c x^{n-1}e^{-\frac{x}{\theta}}dx$$

Derivamos respecto de θ:

$$g'(\theta)=\frac{-1}{\theta^{n+1}(n-1)!}\int_0^c x^{n-1}e^{-\frac{x}{\theta}}dx-\frac{x}{\theta^{n+2}(n-1)!}\int_0^c x^{n-1}e^{-\frac{x}{\theta}}dx<0$$

por lo que es decreciente en θ, y entonces el máximo se alcanza en θ_0. Por tanto, queremos que

$$P(T(\boldsymbol{X})<c|\theta=\theta_0)=\alpha$$

Ahora bien, como $T\sim\gamma\left(\frac{1}{\theta},n\right)$, entonces $U=\frac{2}{\theta}T\sim\chi^2_{2n}$, y obtenemos $c=\frac{\theta_0\cdot\chi^2_{2n,\alpha}}{2}$.

c)

$$T_{exp} = 6.23 > 3.476 = c$$

por tanto $T_{exp} \in S_0$, y aceptamos H_0.
Otra forma: usando el p-valor

$$p = P\left(T < 6.23 | \theta = 0.2\right) = P\left(U < \frac{6.23 \cdot 2}{0.2}\right) = P\left(U < 62.3\right) \implies p \overset{tabla\ \chi^2}{>} 0.75 > \alpha \implies Aceptamos\ H_0$$

d)

$$\lambda\left(\boldsymbol{x}\right) = \frac{L\left(\boldsymbol{x}, \frac{1}{2}\right)}{\sup_{\theta\in\Theta} L\left(\boldsymbol{x}, \theta\right)} = \frac{4\prod \frac{1}{(1-x_i)^2} \exp\left\{-\frac{1}{2}\sum \frac{x_i}{1-x_i}\right\}}{\frac{1}{\hat{\theta}^2}\prod \frac{1}{(1-x_i)^2} \exp\left\{-\frac{1}{\hat{\theta}}\sum \frac{x_i}{1-x_i}\right\}} = 4\hat{\theta}^2 \exp\left\{\left(\frac{1}{\hat{\theta}} - \frac{1}{2}\right) T\left(\boldsymbol{x}\right)\right\}$$

Donde $\hat{\theta}$ es el EMV. Calculémoslo:

$$\log L = -2\log\theta - 2\sum \log\left(1 - x_i\right) - \frac{1}{\theta}\sum \frac{x_i}{1 - x_i}$$

$$\frac{\partial}{\partial\theta}\log L = -\frac{2}{\theta} + \frac{1}{\theta^2}\sum \frac{x_i}{1 - x_i} = 0 \iff \frac{1}{\theta}\sum \frac{x_i}{1 - x_i} = 2 \iff \hat{\theta} = \frac{1}{2}\sum \frac{x_i}{1 - x_i}$$

luego

$$\lambda\left(\boldsymbol{x}\right) = T\left(\boldsymbol{x}\right)^2 \exp\left\{\left(\frac{2}{T\left(\boldsymbol{x}\right)} - \frac{1}{2}\right) T\left(\boldsymbol{x}\right)\right\} = T\left(\boldsymbol{x}\right)^2 \exp\left\{\frac{4 - 2T\left(\boldsymbol{x}\right)}{2T\left(\boldsymbol{x}\right)} T\left(\boldsymbol{x}\right)\right\} = T\left(\boldsymbol{x}\right)^2 \exp\left\{2 - T\left(\boldsymbol{x}\right)\right\}$$

y la región crítica del contraste es

$$S_1 = \left\{\boldsymbol{x} \in sop\left(\boldsymbol{X}\right) : \lambda\left(\boldsymbol{x}\right) < k\right\} = \left\{\boldsymbol{x} \in sop\left(\boldsymbol{X}\right) : T\left(\boldsymbol{x}\right)^2 \exp\left\{2 - T\left(\boldsymbol{x}\right)\right\} < k\right\}$$

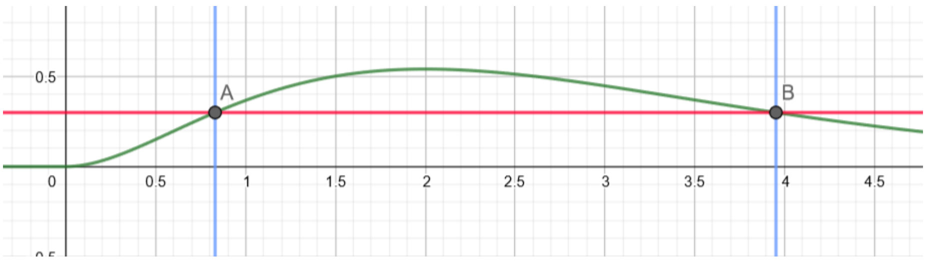
equivalente a

$$S_1 = \left\{\boldsymbol{x} \in sop\left(\boldsymbol{X}\right) : T\left(\boldsymbol{x}\right)^2 \exp\left\{-T\left(\boldsymbol{x}\right)\right\} < k_1\right\},\ k_1 = e^{-2}k$$

Tenemos

$$g\left(t\right) = t^2 e^{-t}$$

$$g'\left(t\right) = 2te^{-t} - t^2 e^{-t} = 0 \iff 2te^{-t} = t^2 e^{-t} \iff 2 = t$$



Aquí la curva verde es g, la línea roja es k_1 y las azules tienen coordenada x a y b:

$$S_1 = \left\{\boldsymbol{x} \in sop\left(\boldsymbol{X}\right) : g\left(T\left(\boldsymbol{x}\right)\right) < k_1\right\} = \left\{\boldsymbol{x} \in sop\left(\boldsymbol{X}\right) : 0 < T\left(\boldsymbol{x}\right) < a\ o\ T\left(\boldsymbol{x}\right) > b\right\}$$

con

$$P\left(0 < T\left(\boldsymbol{x}\right) < a | \theta = \theta_0\right) \cup P\left(T\left(\boldsymbol{x}\right) > b | \theta = \theta_0\right) = \alpha$$

Si llamamos

$$P\left(T\left(\boldsymbol{x}\right) < a | \theta = \theta_0\right) = \alpha_1$$

$$P\left(T\left(\boldsymbol{x}\right)>b|\theta=\theta_0\right)=\alpha_2=\alpha-\alpha_1$$

Esto es lo mismo que

$$P\left(U<\frac{2}{\frac{1}{2}}a\right)=P\left(U<4a\right)=\alpha_1 \iff 4a=\chi^2_{2n,\alpha_1} \iff a=\frac{\chi^2_{2n,\alpha_1}}{4}$$

$$1-P\left(T\left(\boldsymbol{x}\right)<b|\theta=\theta_0\right)=1-P\left(U<4b\right)=1-\alpha+\alpha_1 \iff b=\frac{\chi^2_{2n,1-\alpha+\alpha_1}}{4}$$

y faltaría resolver $g\left(a\right)=g\left(b\right)$.

5. Sea X una población con función de densidad

$$f\left(x,\theta\right)=\frac{2}{\theta x\left(2-x\right)}\left(\frac{x}{2-x}\right)^{-\frac{1}{\theta}}\chi_{(1,2)}\left(x\right)$$

donde θ es un parámetro desconocido positivo. Sea $X_1,...,X_n$ una m.a.s. de X. Se pide:

a) Obtener el test de máxima potencia y extensión α para el contraste $H_0:\theta=\theta_0$ frente a $H_1:\theta=\theta_1$ con $\theta_0<\theta_1$

b) Si $n=20,\theta_0=1,\alpha=0.05$ y $\sum_1^{20}\log\left(\frac{x_i}{1-x_i}\right)=29.7$,obtener el p-valor para dicha muestra. ¿Qué criterio de decisión se tomaría?

a)

$$L\left(\boldsymbol{x},\theta\right)=\frac{2^n}{\theta^n\prod x_j\left(x-x_j\right)}\prod\left(\frac{x_j}{2-x_j}\right)^{-\frac{1}{\theta}}=\frac{2^n}{\theta^n\prod x_j\left(1-x_j\right)}e^{-\frac{1}{\theta}\sum\log\frac{x_j}{1-x_j}}$$

Luego $X\in\varepsilon\left(1\right)$ para el estadístico $T\left(\boldsymbol{X}\right)=\sum\log\frac{X_j}{2-X_j}$ y $A\left(\theta\right)=-\frac{1}{\theta}$ que es monótona creciente. Por tanto, la región crítica es

$$S_1=\left\{\boldsymbol{x}\in\left(1,2\right)^n:T\left(\boldsymbol{x}\right)>c\right\}$$

con

$$P\left(T\left(\boldsymbol{X}\right)>c|\theta=\theta_0\right)=\alpha \iff P\left(T\left(\boldsymbol{X}\right)<c|\theta=\theta_0\right)=1-\alpha$$

Calculemos la distribución de T. Definimos $T_j=\log\frac{X_j}{2-X_j}$ que tiene $sop\left(T_j\right)=\left(0,\infty\right)$ y es $X_j=\frac{e^{T_j}}{e^{T_j}+}$, $X_j'=\frac{2e^{T_j}}{\left(e^{T_j}+1\right)^2}$:

$$f_{T_j}\left(t\right)=f\left(x\left(t\right)\right)\left|x'\left(t\right)\right|=\frac{1}{\theta}e^{-\frac{t}{\theta}}\chi_{(0,\infty)}\left(t\right)\sim Exp\left(\frac{1}{\theta}\right)$$

y $T=\sum T_j\sim\gamma\left(\frac{1}{\theta},n\right)$. Y entonces $U=\frac{2}{\theta}T\sim\chi^2_{2n}$. Por tanto, la probabilidad anteriores es equivalente a

$$P\left(U<\frac{2}{\theta_0}c\right)=1-\alpha \iff \frac{2}{\theta_0}c=\chi_{2n,1-\alpha} \iff c=\frac{\theta_0}{2}\chi_{2n,1-\alpha}$$

Y tenemos el siguiente **criterio de decisión**: dada una m.a.s. x de X, calculamos $T_{exp}=\sum\log\frac{x_j}{2-x_j}$ y:

- si $T_{exp}<c$, entonces aceptamos H_0
- si $T_{exp}>c$, entonces rechazamos H_0

b)

$$p = P(T > 29,70|\theta = 1) = P(U > 2 \cdot 29.7) = P(U > 59.4) = 1 - P(U < 59.4) \overset{tabla\ \chi^2_{40}}{\sim} 1 - 0.975 = 0.025$$

Como $p \sim 0.025 < 0.05 = \alpha$, rechazamos H_0.
6. Sea X una población con función de densidad

$$f(x,\theta) = \frac{1}{2}\exp\left(\frac{\theta - x}{2}\right)\chi_{(\theta,\infty)}(x)$$

siendo θ un parámetro desconocido positivo. Sea $X_1, ..., X_n$ una m.a.s. de X. Se pide:
a) Estudiar si X tiene cociente de verosimilitud monótono en algún estadístico
b) En caso afirmativo, obtener el test uniforme de máxima potencia y extensión α para el contraste $H_0 : \theta = \theta_0$ frente a $H_1 : \theta > \theta_0$
c) Si $n = 6, \theta_0 = 1, \alpha = 0.05$ y $x_1 = 1.5, x_2 = 2, x_3 = 1.3, x_4 = 3, x_5 = 2.5, x_6 = 1.2$, obtener el p-valor para dichar muestra. ¿Qué decisión tomarías en base al p-valor? ¿Podrías mantener dicho criterio por otro procedimiento?
a)

$$L(\boldsymbol{x},\theta) = \left(\frac{1}{2}\right)^n \exp\left\{\frac{n\theta}{2}\right\}\exp\left\{-\frac{\sum x_j}{2}\right\}\chi_{(\theta,\infty)}(X_{1:n})$$

$$\frac{L(\boldsymbol{x},\theta_1)}{L(\boldsymbol{x},\theta_0)} = \exp\left\{\frac{n}{2}(\theta_1 - \theta_0)\right\}\frac{\chi_{(\theta_1,\infty)}(X_{1:n})}{\chi_{(\theta_0,\infty)}(X_{1:n})} = h(R)$$

con $R(\boldsymbol{X}) = X_{1:n}$, y definimos la constante $A = \exp\left\{\frac{n}{2}(\theta_1 - \theta_0)\right\} > 0$.

- Si $X_{1:n} \in (\theta_0, \theta_1) \implies h(R) = 0$
- Si $X_{1:n} > \theta_1 \implies h(R) = A > 0$

Por lo que X tiene cociente de verosimilitud monótono en el estadístico $X_{1:n}$.
b) La región crítica del contraste es

$$S_1 = \{\boldsymbol{x} \in (0,\infty)^n : X_{1:n} > c\}$$

con

$$P(X_{1:n} > c|\theta = \theta_0) = \alpha \iff 1 - P(X_{1:n} < c) = \alpha \iff 1 - G_R(c, \theta = \theta_0) = \alpha$$

donde G es la función de distribución de $X_{1:n}$. Es decir

$$G(r) = 1 - (1 - F(r))^n$$

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = 0 \ si\ x < \theta$$

$$F(x) = \int_{\theta}^{x}\frac{1}{2}\exp\left(\frac{\theta - t}{2}\right)dt = 1 - e^{\frac{\theta - x}{2}},\ x \geq \theta$$

Por tanto

$$G(r) = \begin{cases} 0 & r < \theta \\ 1 - \left(1 - 1 + e^{\frac{\theta - r}{2}}\right)^n = 1 - e^{\frac{n}{2}(\theta - r)} & r \geq \theta \end{cases}$$

Por lo que la probabilidad anterior es

$$1 - G_R(c, \theta = \theta_0) = \alpha \iff \cancel{1} - \cancel{1} + e^{\frac{n}{2}(\theta_0 - c)} = \alpha \iff \frac{n}{2}(\theta_0 - c) = \log\alpha \iff c = \theta_0 - \frac{2}{n}\log\alpha$$

Y tenemos el **criterio de decisión**: dada una m.a.s. x de X, calculamos $R_{exp} = \min(x_1, ..., x_n)$ y:

- Si $R_{exp} < \theta_0 - \frac{2}{n}\log\alpha$, aceptamos H_0
- Si $R_{exp} > \theta_0 - \frac{2}{n}\log\alpha$, rechazamos H_0

c)

$$p = P\left(X_{1:n} > 1.2|\theta = 1\right) = 1 - G_R\left(1.2, \theta = 1\right) = 1 - \left(1 - e^{\frac{6}{n}(1-1.2)}\right) \sim 0.55 > \alpha$$

por lo que aceptamos H_0.

7. Sea X una población con función puntual de probabilidad

$$P\left(X = x\right) = \left(\frac{\theta}{2}\right)^{|x|}\left(1-\theta\right)^{1-|x|},\ x = -1, 0, 1$$

donde $0 < \theta < 1$.
Se considera una m.a.s. de tamaño n suficientemente grande.
a) Obtener el TUMP y extensión α para el contraste $H_0 : \theta \leq \theta_0$ frente a $H_1 : \theta > \theta_0$
b) Si $\theta_0 = \frac{1}{3}, \alpha = 0.05$, determinar el tamaño muestral para que la probabilidad de error de tipo II para $\theta = \frac{2}{3}$ no sea mayor que 0.0001.

8. Sea X una población con función de densidad

$$f\left(x, \theta\right) = \frac{\theta}{\left(1+x\right)^{\theta+1}}\chi_{(0,\infty)}\left(x\right)$$

con $\theta > 0$. Se considera una m.a.s. de X de tamaño n.
a) Obtener el TMP y extensión α para el contraste $H_0 : \theta = \theta_0$ frente a $H_1 : \theta = \theta_1$, con $\theta_0 < \theta_1$
b) Si $\alpha = 0.05$ y la probabilidad de error de tipo II es 0.2, obtener el diseño del test

7 Modelo Lineal General. Caso de Rango Completo

7.1 Introducción

7.1.1 Análisis de Regresión Lineal

El estudio de la relación entre una variable Y y una variable X, en la forma $Y = f(X)$ es un problema, en general, complicado. Por ello se comienza estudiando el caso más sencillo en que la relación sea lineal:

$$Y = a + bx$$

El **análisis de regresión lineal simple** es el estudio de esta relación lineal que nos da el valor de la variable Y o **variable dependiente**, en términos de la X o **variable independiente**, en el caso de que exista esa relación lineal.
Para hacer este estudio obtendremos un conjunto de pares de observaciones (x_i, y_i) de la variable bidimensional (X, Y). Si existiese una relación lineal exacta entre ambas variables, esta quedaría de manifiesto simplemente dibujando en el plano los pares (x_i, y_i), dando lugar a lo que se conoce como **diagrama de dispersión** o **nube de puntos**. Si la relación fuese lineal los puntos quedarían alineados sobre una recta. Sin embargo esta es una situación extrema que no se suele dar en la práctica. Los puntos normalmente no se hallarán alineados sobre una recta, aunque pueden presentar una disposición en cierta forma alineada.

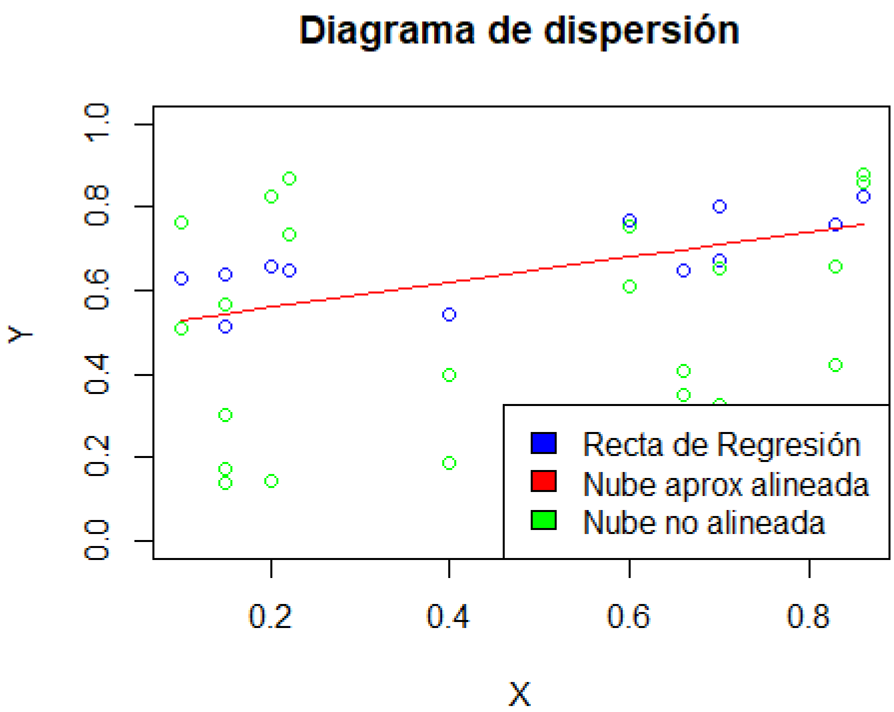


Figure 1: Nube de puntos y recta de regresión

En la figura pueden verse los puntos azules, que no están perfectamente alineados pero están cerca de la recta roja. Esto se debe a diversos factores, como pueden ser errores de medición, presencia de factores experimentales,... que son incontrolables y que hacen que la relación lineal no sea exacta. Así, un modelo más razonable para los pares (x_i, y_i) es:

$$y_i = a + bx_i + \varepsilon_i$$

donde ε_i son valores aleatorios que estropean la linealidad y se denominan **residuos**.
Esto lo podemos modelizar probabilísticamente considerando que las observaciones $y_1, ..., y_n$ provienen de un vector aleatorio n-dimensional, que en forma matricial podemos expresar:

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \boldsymbol{X}\beta + \varepsilon$$

donde

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \ \beta = \begin{pmatrix} a \\ b \end{pmatrix}, \ \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

y los valores $x_1, ..., x_n$ han sido fijados previamente.

7.1.2 Análisis de la varianza simple

Otra situación donde nos encontramos con un caso particular del modelo lineal general se da cuando queremos estudiar, y comparar, la media de dos poblaciones.
En tal caso se disponen de observaciones $y_{i,1}, ..., y_{i,n}$ de variables Y_i con $i = 1, ..., k$ con media μ_i. Como antes, podemos expresarlo de forma matricial:

$$\boldsymbol{Y} = \begin{pmatrix} Y_{1,1} \\ \vdots \\ Y_{1,n_1} \\ \vdots \\ Y_{k,1} \\ \vdots \\ Y_{k,n_k} \end{pmatrix} = \boldsymbol{X}\beta + \varepsilon$$

con

$$\boldsymbol{X} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & 1 \\ \vdots & & \vdots & \vdots \\ 0 & \dots & 0 & 1 \end{pmatrix}_{N \times k}, \ \beta = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_k \end{pmatrix}, \ \varepsilon = \begin{pmatrix} \varepsilon_{1,1} \\ \vdots \\ \varepsilon_{1,n_1} \\ \vdots \\ \varepsilon_{k,1} \\ \vdots \\ \varepsilon_{k,n_k} \end{pmatrix}$$

donde $y_1, ..., y_{1,n_1}, ..., y_{k,1}, ..., y_{k,n_k}$ son observaciones del vector aleatorio de dimensión $N = \sum_{i=1}^{k} n_i$ y $\varepsilon_{i,j} = Y_{i,j} - \mu_i$.
En ambos casos las observaciones de la muestra provienen de una variable aleatoria n-dimensional $\boldsymbol{Y}$ que se puede escribir, en general, según el modelo

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \varepsilon$$

donde $\boldsymbol{X}$ es una matriz con valores conocidos de dimensión $n \times k$, β es un vector de parámetros desconocidos y de dimensión $k \times 1$ y ε es un vector aleatorio de dimensión n, el **vector de residuos**. En este caso

$$y_i = \sum_{j=1}^{k} x_{ij}\beta_j + \varepsilon_i$$

Este modelo probabilístico se conoce como el **modelo lineal general** y cualquier resultado de inferencia sobre este modelo podrá ser aplicado a los dos casos particulares anteriores.

7.2 Inferencia bajo la suposición de linealidad e incorrelación

Supondremos que el modelo lineal general $\boldsymbol{Y} = \boldsymbol{X}\beta + \varepsilon$ verifica:

1. $\boldsymbol{Y}$ es un vector aleatorio n-dimensional

2. $\boldsymbol{X}$ es una matriz de valores conocidos de dimensión $n \times k$, con $k \leq n$ y $r(\boldsymbol{X}) = k$ (es de **rango completo**)

3. β es un vector de valores desconocidos de dimensión $k \times 1$

4. ε es un vector aleatorio n-dimensional que verifica:

   (a) $E[\varepsilon] = 0$
   (b) $cov[\varepsilon] = \sigma^2 I_n$

   Es decir, es un vector de componentes **incorreladas** y con igual varianza (**homocedásticas**).

Estas suposiciones se pueden escribir en términos del vector $\boldsymbol{Y}$ diciendo que $\boldsymbol{Y}$ es un vector aleatorio de dimensión n de variables incorreladas y homocedásticas con vector de medias $\boldsymbol{X}\beta$.
Observemos que

$$E[\boldsymbol{Y}] = \boldsymbol{X}\beta$$

$$E[Y_i] = \sum_{j=1}^{k} x_{ij}\beta_j$$

$$cov[\boldsymbol{Y}] = cov[\varepsilon] = \sigma^2 I_n$$

El primer problema que nos planteamos es el de estimar el vector de parámetros desconocidos β. Legendre propuesto como estimadores aquellos valores $\widehat{\beta}_j$ que minimizaran las diferencias al cuadrado entre los valores observados y_i y la expresión $\sum_{j=1}^{k} x_{ij}\beta_j$. Es decir, aquellos que minimicen:

$$\sum_{i=}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{k} x_{ij}\beta_j \right)^2 = (\boldsymbol{y} - \boldsymbol{X}\beta)^T (\boldsymbol{y} - \boldsymbol{X}\beta) = \varepsilon^T \varepsilon$$

y se denominan **estimadores de mínimos cuadrados**.

Theorem 7.1. *Sea $\boldsymbol{Y} = \boldsymbol{X}\beta + \varepsilon$ verificando 1., 2., 3. y 4., entonces el estimador de mínimos cuadrados del parámetro β viene dado por*

$$\widehat{\beta} = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{y}$$

Y podemos ahora estudiar $\widehat{\beta}$ como estadístico. Veamos su vector de medias y su matriz de covarianzas:

Theorem 7.2. *Sea* $\boldsymbol{Y} = \boldsymbol{X}\beta + \varepsilon$ *verificando 1.,2.,3. y 4., entonces*

1. $E\left[\widehat{\beta}\right] = \beta$

2. $cov\left[\widehat{\beta}\right] = \sigma^2\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}$

Proof. Veamos ambos resultados:

1.
$$E\left[\widehat{\beta}\right] = E\left[\left(X^TX\right)^{-1}X^TY\right] = \left(X^TX\right)^{-1}X^TE\left[Y\right] = \left(X^TX\right)^{-1}X^T\left(X\beta\right) = \cancel{\left(X^TX\right)^{-1}}\cancel{\left(X^TX\right)}\beta = \beta$$

2.
$$cov\left[\widehat{\beta}\right] = cov\left[\left(X^TX\right)^{-1}X^TY\right] = \left(X^TX\right)^{-1}X^T\cdot cov\left[Y\right]\cdot X\left(X^TX\right)^{-1} =$$
$$= \left(X^TX\right)^{-1}X^T\cdot\sigma^2I_n\cdot X\left(X^TX\right)^{-1} = \sigma^2\cancel{\left(X^TX\right)^{-1}\left(X^TX\right)}\left(X^TX\right)^{-1} = \sigma^2\left(X^TX\right)^{-1}$$

□

Observamos, por tanto, que $\widehat{\beta}$ es estimador insesgado de β.

Definition 7.3. Se denomina **estimador lineal** a aquel que se puede poner como combinación lineal de los valores de la muestra.

Remark 7.4. $\widehat{\beta}$ es un estimador lineal.

Theorem 7.5. ***Teorema de Gauss-Markov***
Sea $\boldsymbol{Y} = \boldsymbol{X}\beta + \varepsilon$ *verificando las condiciones 1.,2.,3. y 4. y sea* $\theta = \lambda^T\beta$ *con* $\lambda^T = (\lambda_1, ..., \lambda_k)$*. Entonces, dentro de la clase de estimadores lineales, el estadístico*

$$\widehat{\theta} = \lambda^T\widehat{\beta}$$

es el único estimadore insesgado de mínima varinaza de θ*.*

Proof. Veamos que es **lineal**
$$\hat{\theta} = \lambda^T\hat{\beta} = \lambda^T\left(X^TX\right)^{-1}X^TY$$

y es una combinación lineal, ya que λ^T es $1\times k$, $\left(X^TX\right)^{-1}$ es $k\times k$, X^T es $k\times n$ e Y es $n\times 1$, de forma que el resultado es un escalar suma de los y_i.
Y veamos que es de **mínima varianza**. Sea $\overline{\theta} = a^TY$ un estimador insesgado de $\theta = \lambda^T\beta$ y $b^T = a^T - \lambda^T\left(X^TX\right)^{-1}X^T$. Entonces

$$\lambda^T\beta = E\left[\overline{\theta}\right] = E\left[a^TY\right] = E\left[\left(b^T + \lambda^T\left(X^TX\right)^{-1}X^T\right)Y\right] = E\left[b^TY\right] + E\left[\lambda^T\left(X^TX\right)^{-1}X^TY\right] =$$

$$= b^TX\beta + \lambda^T\cancel{\left(X^TX\right)^{-1}X^TX}\beta = b^YX\beta + \lambda^T\beta$$

para todo $\beta\in\mathbb{R}^k$. Restando $\lambda^T\beta$ a ambos lados de la igualdad, obtenemos que $b^TX\beta = 0, \forall\beta\in\mathbb{R}^k \implies b^TX = 0$.

Por otro lado

$$Var\left[\overline{\theta}\right] = Var\left[a^T Y\right] = a^T cov\left[Y\right] a = a^T \sigma^2 I_n a = \sigma^2 a^T a = \sigma^2 \left(b^T + \lambda^T \left(X^T X\right)^{-1} X^T\right)\left(b + X\left(X^T X\right)^{-1}\lambda\right) =$$

$$= \sigma^2 \left[b^T b + \cancelto{0}{b^T X}\left(X^T X\right)^{-1}\lambda + \lambda^T\left(X^T X\right)^{-1}\cancelto{0}{X^T b} + \lambda^T \cancel{\left(X^T X\right)^{-1} X^T X}\left(X^T X\right)^{-1}\lambda\right] =$$

$$= \sigma^2\left[b^T b + \lambda^T\left(X^T X\right)^{-1}\lambda\right] = \sigma^2 b^T b + \lambda^T\left(X^T X\right)^{-1}\sigma^2\lambda = \sigma^2 b^T b + \lambda^T cov\left[\hat{\beta}\right]\lambda =$$

$$= \sigma^2 b^T b + Var\left[\lambda\hat{\beta}\right] = \sigma^2 b^T b + Var\left[\hat{\theta}\right]$$

y como $\sigma^2 b^T b \geq 0$, entonces

$$Var\left[\overline{\theta}\right] \geq Var\left[\hat{\theta}\right]$$

y entonces es estimador lineal insesgado de mínima varianza.
Además

$$Var\left[\overline{\theta}\right] = Var\left[\hat{\theta}\right] \iff \sigma^2 b^T b = 0 \iff b = 0 \iff a^T = \lambda^T\left(X^T X\right)^{-1} X^T \iff \overline{\theta} = \hat{\theta}$$

□

Corollary 7.6. *$\widehat{\beta_i}$ es estimador lineal insesgado de mínima varianza (ELIMV) de β_i.*

Quedaría dar un estimador de la varianza del modelo, que es, en general, otro parámetro desconocido. Para obtener dicho estimador, necesitamos el siguiente resultado técnico:

Lemma 7.7. *Sea $P = X\left(X^T X\right)^{-1} X^T$, entonces se verifica:*

1. *$P,\ I_n - P$ son simétricas e idempotentes*
2. *$r\left(I_n - P\right) = tr\left(I_n - P\right) = n - k$*
3. *$\left(I_n - P\right) X = 0$*

Y podemos establecer el siguiente resultado:

Theorem 7.8. *Sea $\boldsymbol{Y} = \boldsymbol{X}\beta + \varepsilon$ verificando las condiciones 1.,2.,3. y 4., entonces el estadístico*

$$S^2 = \frac{\left(Y - X\widehat{\beta}\right)^T\left(Y - X\widehat{\beta}\right)}{n-k}$$

es un estimador insesgado de σ^2.

Proof.

$$Y - X\widehat{\beta} = Y - X\left(X^T X\right)^{-1} X^T Y = Y - PY = \left(I_n - P\right) Y$$

Si tomamos

$$\left(n-k\right) S^2 = \left(Y - X\widehat{\beta}\right)^T\left(Y - X\widehat{\beta}\right) = Y^T\left(I_n - P\right)^T\left(I_n - P\right) Y \overset{I_n - P \text{ simétrica}}{=} Y^T\left(I_n - P\right)^2 Y =$$

$$\overset{I_n - P \text{ idempotente}}{=} Y^T\left(I_n - P\right) Y$$

y tomando la esperanza

$$E\left[(n-k)S^2\right] = E\left[Y^T(I_n-P)Y\right] = E\left[(Y-E[Y]+E[Y])^T(I_n-P)(Y-E[Y]+E[Y])\right] =$$

$$= E\left[(Y-E[Y])^T(I_n-P)(Y-E[Y]) + E[Y]^T(I_n-P)E[Y]\right] =$$

$$= E\left[(Y-E[Y])^T(I_n-P)(Y-E[Y])\right] + E[Y]^T E[I_n-P]E[Y] =$$

$$= E\left[(Y-E[Y])^T(I_n-P)(Y-E[Y])\right] + (X\beta)^T(I_n-P)X\beta \overset{(I_n-P)X=0}{=}$$

$$= E\left[(Y-E[Y])^T(I_n-P)(Y-E[Y])\right] = E\left[\sum_i\sum_j (Y_i-E[Y_i])\, m_{ij}\,(Y_j-E[Y_j])\right] = (*)$$

donde $M = I_n - P = (m_{ij})$, y entonces

$$(*) = \sum_i\sum_j m_{ij}E\left[(Y_i-E[Y_i])(Y_j-E[Y_j])\right] \overset{Y_i,Y_j \text{ incorreladas } \forall i\neq j}{=} \sum_i m_{ii}\sigma^2 = \sigma^2 tr(I_n-P) = \sigma^2(n-k)$$

y tenemos el resultado. □

7.3 Inferencia bajo la suposición de normalidad

A las hipótesis 1.,2.,3. y 4. añadimos:

5. ε_i sigue una distribución normal, para $i = 1, ..., n$

Por tanto, las hipótesis 4. y 5. son equivalentes a la hipótesis

$$\varepsilon \sim N_n\left(0, \sigma^2 I_n\right)$$

Recordemos algunas propiedades de la distribución normal n-dimensional:

a. Si $U \sim N_n(\mu, V)$ e $Y = AU + B$, donde A, B son matrices con dimensiones tales que el producto y la suma tienen sentido, se verifica que $Y \sim N_n(\eta, W)$ donde $\eta = A\mu + B, W = AVA^T$.

   En nuestro caso, para el vector Y tendremos que es

$$Y \sim N_n\left(X\beta, \sigma^2 I_n\right)$$

b. Cualquier distribución marginal de una distribución normal n-dimensional es también normal. En nuestro caso:

$$Y_i \sim N\left(\sum_{j=1}^{k} x_{ij}\beta_j, \sigma^2\right)$$

c. En el modelo normal n-dimensional, si la matriz de covarianzas es diagonal, las componentes del vector son incorreladas e independientes. En nuestro caso, como la matriz de covarianzas de Y es diagonal, las componentes Y_i, Y_j son independientes, para $i \neq j$.

7.3.1 Estimación de Máxima Verosimilitud de los parámetros

Podemos ahora, supuesta la normalidad, estudiar los estimadores máximo verosímiles de los parámetros $\left(\beta_1, .., \beta_k, \sigma^2\right)$. Para ello tenemos que maximizar la función de verosimilitud:

$$L\left(y_1, ..., y_n, \beta_1, ..., \beta_k, \sigma^2\right) = \frac{1}{\left(\sigma^2 2\pi\right)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} \frac{\left(y_i - \sum_{j=1}^{k} x_{ij}\beta_j\right)^2}{\sigma^2}\right\} =$$

$$= \frac{1}{\left(\sigma^2 2\pi\right)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2}\frac{\left(Y - X\beta\right)^T\left(Y - X\beta\right)}{\sigma^2}\right\}$$

Tomando logaritmos, hay que maximizar:

$$-\frac{n}{2}\log\left(\sigma^2 2\pi\right) - \frac{1}{2}\sum_{i=1}^{n} \frac{\left(y_i - \sum_{j=1}^{k} x_{ij}\beta_j\right)^2}{\sigma^2}$$

Fijando σ^2, se trata de minimizar

$$\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{k} x_{ij}\beta_j\right)^2$$

Esto ya lo hemos hecho, y obtuvimos $\widehat{\beta} = \left(X^TX\right)^{-1} X^TY$, coincidiendo con el estimador de mínimos cuadrados de β.
Conociendo el estimador $\widehat{\beta}$ de β y puesto que este es independiente de σ^2, basta ahora buscar el valor de σ^2 para el que se alcanza el máximo, fijando $\beta = \widehat{\beta}$. El estimador resultante es

$$\widehat{\sigma}^2 = \frac{\left(Y - X\widehat{\beta}\right)^T\left(Y - X\widehat{\beta}\right)}{n}$$

Este estimador de máxima verosimilitud de σ^2 no coincide con el estimador insesgado S^2 de σ^2.
Y ahora nuestro interés se encuentra en estudiar las distribuciones en el muestreo de estos estadísticos.

Lemma 7.9. *Sea $Q = Q_1 - Q_2$ donde $Q_i \sim \chi^2_{n_i}, i = 1, 2$ y $n = n_1 - n_2 > 0$, siendo Q independiente de Q_2. Entonces*

$$Q \sim \chi^2_n$$

Theorem 7.10. *Sea $Y = X\beta + \varepsilon$ verificando 1.,2.,3.,4. y 5., entonces*

1. $\widehat{\beta} \sim N_k\left(\beta, \sigma^2\left(X^TX\right)^{-1}\right)$
2. $\frac{\left(\widehat{\beta}-\beta\right)^T\left(X^TX\right)\left(\widehat{\beta}-\beta\right)}{\sigma^2} \sim \chi^2_k$
3. *$\widehat{\beta}$ es independiente de S^2*
4. $\frac{(n-k)}{\sigma^2}S^2 \sim \chi^2_{n-k}$

Proof. Veamos cada afirmación:

1. $\widehat{\beta}=\left(X^TX\right)^{-1}X^TY$ donde $Y\sim N_n\left(X\beta,\sigma^2I_n\right)$, luego, por la propiedad a. y dado que $\left(X^TX\right)^{-1}X^T\left(X\beta\right)=\beta$ y $\left(X^TX\right)^{-1}X^T\left(\sigma^2I_n\right)X\left(X^TX\right)^{-1}=\sigma^2\left(X^TX\right)^{-1}$, se tiene que

$$\widehat{\beta}\sim N_k\left(\beta,\sigma^2\left(X^TX\right)^{-1}\right)$$

2.

$$\frac{\left(\widehat{\beta}-\beta\right)^T\left(X^TX\right)\left(\widehat{\beta}-\beta\right)}{\sigma^2}=\left(\widehat{\beta}-\beta\right)^T cov\left(\widehat{\beta}\right)^{-1}\left(\widehat{\beta}-\beta\right)\overset{(*)}{=}\left(\widehat{\beta}-\beta\right)^T C^TC\cdot cov\left(\widehat{\beta}\right)^{-1}\cdot C^TC\left(\widehat{\beta}-\beta\right)$$

donde $(*)$ se debe a que existe C matriz ortogonal, tal que $Z=C\left(\widehat{\beta}-\beta\right)\sim N_k\left(0,\Lambda\right)$ con $\Lambda=diag\left(\lambda_1,...,\lambda_k\right)$. Así, lo anterior es igual a

$$Z^TC\cdot cov\left(\widehat{\beta}\right)^{-1}\cdot C^TZ\overset{(**)}{=}Z^Tcov\left(Z\right)^{-1}Z=\sum_{i=1}^{k}\left(\frac{Z_i}{\sqrt{\lambda_i}}\right)^2\overset{Z_i\sim N(0,\lambda_i),\ indeps}{\sim}\chi_k^2$$

y $(**)$ es porque $cov\left(Z\right)=cov\left(C\left(\widehat{\beta}-\beta\right)\right)=Ccov\left(\widehat{\beta}-\beta\right)C^T\overset{\beta\ cte}{=}Ccov\left(\widehat{\beta}\right)C^T$

3.

$$S^2=\frac{\left(Y-X\widehat{\beta}\right)^T\left(Y-X\widehat{\beta}\right)}{n-k}=h\left(Y-X\widehat{\beta}\right)$$

donde h es medible Borel, y basta demostrar que $\widehat{\beta}$ e $Y-X\widehat{\beta}$ son independientes. Vamos a ver la independencia entre $\widehat{\beta}-\beta$ e $Y-X\widehat{\beta}$, que es equivalente a la anterior.
Por la normalidad de $\widehat{\beta}-\beta$ e $Y-X\widehat{\beta}$, basta ver la incorrelación de los dos vectores. Así, tomamos la variable $Z=\left(\widehat{\beta}-\beta,Y-X\widehat{\beta}\right)$ y calculamos su matriz de covarianza:

$$V_{i,j}=cov\left(Z_i,Z_j\right)$$

pero de aquí solo nos interesa ver tiene esta forma

$$V=\begin{pmatrix}V_1&0\\0&V_2\end{pmatrix}$$

donde V_1 es la matriz de covarianza de $\widehat{\beta}-\beta$ y V_2 la de $Y-X\widehat{\beta}$. Así, queremos ver que

$$E\left[\left(\widehat{\beta}-\beta\right)_i\left(Y-X\widehat{\beta}\right)_j\right]=0,\forall i,j$$

Esto es equivalente a que

$$E\left[\left(\widehat{\beta}-\beta\right)\left(Y-X\widehat{\beta}\right)^T\right]=0_{k\times n}$$

Veamos que es así:

$$E\left[\left(\widehat{\beta}-\beta\right)\left(Y-X\widehat{\beta}\right)^T\right]\overset{Y-X\widehat{\beta}=(I_n-P)Y}{=}E\left[\left(\left(X^TX\right)^{-1}X^TY-\beta\right)Y^T\left(I_n-P\right)\right]=$$

$$=E\left[\left(X^TX\right)^{-1}X^TYY^T\left(I_n-P\right)\right]-\beta E\left[Y^T\left(I_n-P\right)\right]=$$

$$=\left(X^TX\right)^{-1}X^TE\left[YY^T\right]\left(I_n-P\right)-\beta\beta^TX^T\left(I_n-P\right)=$$

$$\overset{X^T(I_n-P)=0}{=} \left(X^TX\right)^{-1} X^T \left[cov\left(Y\right) + E\left[Y\right] E\left[Y\right]^T\right] \left(I_n - P\right) =$$

$$= \left(X^TX\right)^{-1} X^T \left[cov\left(Y\right) + X\beta\beta^T X\right] \left(I_n - P\right) = \left(X^TX\right)^{-1} X^T cov\left(Y\right) \left(I_n - P\right)$$

$$\overset{cov(Y)=\sigma^2 I_n}{=} \sigma^2 \left(X^TX\right) X^T \left(I_n - P\right) \overset{X^T(I_n-P)}{=} 0$$

como queríamos.

4.

$$\frac{(n-k)\, S^2}{\sigma^2} = \frac{\left(Y - X\widehat{\beta}\right)^T \left(Y - X\widehat{\beta}\right)}{\sigma^2} = \frac{\left(Y - X\beta\right)^T \left(Y - X\beta\right)}{\sigma^2} - \frac{\left(\widehat{\beta} - \beta\right)^T X^T X \left(\widehat{\beta} - \beta\right)}{\sigma^2}$$

Ahora bien:

- $\frac{(Y-X\beta)^T(Y-X\beta)}{\sigma^2}$ es suma de variables aleatorias al cuadrado con distribución $N\left(0,1\right)$ e independientes entre sí, luego su distribución es χ^2_n
- Por 2., $\frac{\left(\widehat{\beta}-\beta\right)^T X^T X\left(\widehat{\beta}-\beta\right)}{\sigma^2}$ es χ^2_k
- Por 3., $\frac{(n-k)S^2}{\sigma^2}$ es independiente de $\frac{\left(\widehat{\beta}-\beta\right)^T X^T X\left(\widehat{\beta}-\beta\right)}{\sigma^2}$

Aplicando el lema, tenemos que $\frac{(n-k)S^2}{\sigma^2} \sim \chi^2_{n-k}$

□

7.3.2 Contrastes de hipótesis sobre los parámetros

Planteamos ahora el problema de contrastar las hipótesis:

$$H_0: \ A\beta = c$$

frente a

$$H_1: \ A\beta \neq c$$

Este tipo de contrastes puede darse en varias situaciones:
Contrastes sobre los parámetros β_i:

$$H_0: \ \beta_i = d$$

frente a

$$H_1: \ \beta_i \neq d$$

puede ponerse tomando $A = (0, ...0, 1, 0, ..., 0)$ y $c = d$.
Análisis de regresión lineal
El contraste de las hipótesis

$$H_0: \ \beta_1 = \beta_2 = ... = \beta_p = 0, \ p \leq k$$

frente a

$$H_1: \ \exists i | \beta_i \neq 0$$

se puede poner como

$$A = I_p, \ c = 0_p$$

Para hacer estos contrastes, consideramos el test de cociente de verosimilitudes generalizado. La región crítica es

$$S_1 = \left\{ (y_1, ..., y_n) \in \mathbb{R}^n | \lambda (y_1, ..., y_n) = \frac{\sup_{(\beta,\sigma^2), A\beta = c} L(y_1, ..., y_n, \beta, \sigma^2)}{\sup_{(\beta,\sigma^2)} L(y_1, ..., y_n, \beta, \sigma^2)} < l \right\}$$

con $0 < l < 1$. Por ser la extensión del test α, el valor de l se obtiene de

$$P(\lambda(\boldsymbol{y}) < l | A\beta = c) = \alpha$$

Haciendo algunos cálculos, se puede llegar a que la región crítica es

$$S_1 = \left\{ \boldsymbol{y} \in \mathbb{R}^n | \frac{n-k}{q} \frac{\hat{\sigma_A}^2 - \hat{\sigma}^2}{\hat{\sigma}^2} > l' \right\}$$

con $l' > 0$ y donde

$$\hat{\sigma_A}^2 = \frac{\left(\boldsymbol{y} - X\hat{\beta_A}\right)^T \left(\boldsymbol{y} - X\hat{\beta_A}\right)}{n}$$

$$\hat{\beta_A} = \hat{\beta} + \left(X^T X\right)^{-1} A^T \left(A \left(X^T X\right)^{-1} A^T\right)^{-1} \left(c - A\hat{\beta}\right)$$

Theorem 7.11. *Sea $Y = X\beta + \varepsilon$ verificando las condiciones 1.,2.,3.,4. y 5.. Si $A\beta = c$ entonces el estadístico*

$$F = \frac{n-k}{q} \frac{\hat{\sigma_A}^2 - \hat{\sigma}^2}{\hat{\sigma}^2} \sim \mathcal{F}_{q,n-k}$$

Como el test es de extensión α si $A\beta = c$, se verifica $P(F < l') = 1 - \alpha$ y obtenemos que

$$l' = \mathcal{F}_{q,n-k,1-\alpha}$$

Por tanto, aceptamos H_0 si $F_{exp} < l'$ y rechazamos H_0 en caso contrario.

Proof.

$$F = \frac{\frac{n}{\sigma^2} \frac{\hat{\sigma_A}^2 - \hat{\sigma}^2}{q}}{\frac{n}{\sigma^2} \frac{\hat{\sigma}^2}{n-k}}$$

Y queremos probar que

$$\frac{n}{\sigma^2} \left(\hat{\sigma_A}^2 - \hat{\sigma}^2\right) \sim \chi^2_q$$

$$\frac{n}{\sigma^2} \hat{\sigma}^2 \sim \chi^2_{n-k}$$

y que son independientes, para tener el resultado.
Por un lado:

$$\frac{n}{\sigma^2} \hat{\sigma}^2 = \frac{\cancel{n}}{\sigma^2} \frac{\left(\boldsymbol{y} - X\hat{\beta}\right)^T \left(\boldsymbol{y} - X\hat{\beta}\right)}{\cancel{n}} = \frac{n-k}{\sigma^2} S^2 \overset{7.10}{\sim} \chi^2_{n-k}$$

Por otro:

$$\frac{n}{\sigma^2} \left(\hat{\sigma_A}^2 - \hat{\sigma}^2\right) = \frac{\cancel{n}}{\sigma^2} \left(\frac{\left(\boldsymbol{y} - X\hat{\beta_A}\right)^T \left(\boldsymbol{y} - X\hat{\beta_A}\right)}{\cancel{n}} - \frac{\left(\boldsymbol{y} - X\hat{\beta}\right)^T \left(\boldsymbol{y} - X\hat{\beta}\right)}{\cancel{n}} \right) =$$

$$= \frac{1}{\sigma^2}\left[\left(X\hat{\beta} - X\hat{\beta_A}\right)^T\left(X\hat{\beta} - X\hat{\beta_A}\right)\right] = \frac{1}{\sigma^2}\left[\left(\hat{\beta} - \hat{\beta_A}\right)^T X^T X\left(\hat{\beta} - \hat{\beta_A}\right)\right] = (*)$$

Desarrollando ahora la expresión de $\hat{\beta_A}$ obtenemos

$$(*) = \frac{1}{\sigma^2}\left[\left(A\hat{\beta} - c\right)^T\left(A\left(X^TX\right)^{-1}A^T\right)^{-1}A\left(X^TX\right)^{-1}\cancel{X^TX\left(X^TX\right)^{-1}}A^T\left(A\left(X^TX\right)^{-1}A^T\right)^{-1}\left(A\hat{\beta} - c\right)\right] =$$

$$= \frac{1}{\sigma^2}\left[\left(A\hat{\beta} - c\right)^T\cancel{\left(A\left(X^TX\right)^{-1}A^T\right)^{-1}A\left(X^TX\right)^{-1}A^T}\left(A\left(X^TX\right)^{-1}A^T\right)^{-1}\left(A\hat{\beta} - c\right)\right] =$$

$$= \frac{1}{\sigma^2}\left[\left(A\hat{\beta} - c\right)^T\left(A\left(X^TX\right)^{-1}A^T\right)^{-1}\left(A\hat{\beta} - c\right)\right] = (*)$$

Observamos aquí que $A\hat{\beta} \sim N_q\left(A\beta, \sigma^2 A\left(X^TX\right)^{-1}A^T\right)$ y que estamos trabajando suponiendo $H_0 : A\beta = c$, por tanto

$$(*) = \left(A\hat{\beta} - E\left[A\hat{\beta}\right]\right)^T\left(cov\left[A\hat{\beta}\right]\right)^{-1}\left(A\hat{\beta} - E\left[A\hat{\beta}\right]\right) \overset{como\ b\ de\ 7.10}{\sim} \chi^2_q$$

Y solo falta ver la independencia. Pero

$$\hat{\sigma}^2 = g\left(S^2\right)\ y\ \hat{\sigma_A}^2 - \hat{\sigma}^2 = h\left(\hat{\beta}\right)$$

que son independientes, por el teorema 7.10. □