

Propiedades de $\mathbf{P}$

$\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \in \Re^{n \times n}$ es la matriz de **proyección sobre el espacio generado por las columnas de $\mathbf{X}$** . Suele llamarse a esta matriz $\mathbf{P}$ por ser de proyección $\mathbf{P}$ o $\mathbf{H}$ de hat matrix.

Matriz simétrica e idempotente, es decir: $\mathbf{P} = \mathbf{P}' = \mathbf{P}^2$. $\mathbf{I} - \mathbf{P}$ también es simétrica es idempotente, es decir también es una matriz de proyección y proyecta sobre el ortogonal de V_r .

Lema:

- i) $\mathbf{P}$ y $\mathbf{I} - \mathbf{P}$ son simétricas e idempotentes
- ii) $\text{rg}(\mathbf{P}) = \text{tr}(\mathbf{P}) = p$ y $\text{rg}(\mathbf{I} - \mathbf{P}) = \text{tr}(\mathbf{I} - \mathbf{P}) = n - p$
- iii) $(\mathbf{I} - \mathbf{P})\mathbf{X} = \mathbf{0}$

Suma de Cuadrados

Tenemos que

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 = \|\mathbf{Y} - \mathbf{PY}\|^2$$

Notemos que obtenemos el Teorema de Pitágoras. En efecto,

$$\begin{aligned}\|\mathbf{Y} - \hat{\mathbf{Y}}\|^2 &= \|\mathbf{Y} - \mathbf{PY}\|^2 &= \|(\mathbf{I} - \mathbf{P})\mathbf{Y}\|^2 \\ &= \mathbf{Y}'(\mathbf{I} - \mathbf{P})'(\mathbf{I} - \mathbf{P})\mathbf{Y} = \mathbf{Y}'(\mathbf{I} - \mathbf{P})\mathbf{Y} \\ &= \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{PY} \\ &= \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{P}'\mathbf{PY} \\ &= \|\mathbf{Y}\|^2 - \|\mathbf{PY}\|^2 \\ &= \|\mathbf{Y}\|^2 - \|\hat{\mathbf{Y}}\|^2 \\ \Rightarrow \quad \|\mathbf{Y}\|^2 &= \|\mathbf{Y} - \hat{\mathbf{Y}}\|^2 + \|\hat{\mathbf{Y}}\|^2\end{aligned}$$

Errores Normales

Si los errores además de ser independientes con media 0 y varianza σ^2 , son normales, es decir

$$\epsilon_i \sim N(0, \sigma^2) \quad i = 1, \dots, n$$

el modelo Ω puede formularse de la siguiente forma:

$$\Omega \quad : \quad \mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\theta}, \sigma^2\mathbf{I})$$

En este caso puede demostrarse que el estimador de máxima verosimilitud de $\boldsymbol{\theta}$ coincide con el de mínimos cuadrados.

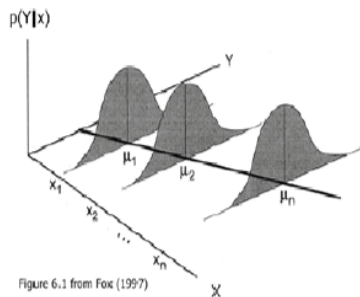


Figure 6.1 from Fox (1997)

En el caso de rango completo, es decir cuando $rg(\mathbf{X}) = p$, obtenemos el siguiente resultado:

Teorema: Supongamos que $\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\theta}, \sigma^2\mathbf{I})$, $rg(\mathbf{X}) = p$, $\boldsymbol{\theta} \in \Re^p$. Entonces,

- i) $\hat{\boldsymbol{\theta}} \sim N_p(\boldsymbol{\theta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$
- ii) $\frac{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'(\mathbf{X}'\mathbf{X})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})}{\sigma^2} \sim \chi_p^2$
- iii) $\hat{\boldsymbol{\theta}}$ y $\frac{(n-p)s^2}{\sigma^2}$ son independientes
- iv) $\frac{(n-p)s^2}{\sigma^2} \sim \chi_{n-p}^2$

Normals Multivariados.

Estadística (II)

2019

A. Prieto.

Decimos que $x \in \mathbb{R}^n$ tiene distribución normal multivariada si su función de densidad es:

$$f_x(x) = \left(\frac{1}{2\pi}\right)^{n/2} \frac{1}{|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)' \Sigma^{-1}(x-\mu)\right)$$

siendo $\mu \in \mathbb{R}^n$, $\Sigma \in \mathbb{R}^{n \times n}$ definido positiva

$$x \sim N_n(\mu, \Sigma)$$

• Si $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_m^2)$ resulta

$$f_X(x) = \left(\frac{1}{2\pi}\right)^{n/2} \left(\frac{1}{\prod_{i=1}^n \sigma_i^2}\right)^{1/2} \cdot \exp\left(-\frac{1}{2} \sum_{i=1}^n \frac{(x_i - \mu_i)^2}{\sigma_i^2}\right)$$

A si $f_X(x)$ puede factorizarse:

$$f_X(x) = \prod_{i=1}^n \left(\frac{1}{2\pi \sigma_i^2}\right)^{1/2} \cdot \exp\left(-\frac{1}{2} \left(\frac{x_i - \mu_i}{\sigma_i}\right)^2\right)$$

$\Rightarrow x_i \sim N(\mu_i, \sigma_i^2)$ independientes

• Amor lo mismo al caso univariado, vale

$$\text{Si } x \sim N_m(\mu, \Sigma) \Rightarrow \Sigma^{-1/2}(x - \mu) \sim N_m(0, I)$$

• Esto es consecuencia de un resultado más general: Si $x \sim N(\mu, \Sigma)$, $A \in \mathbb{R}^{u \times u}$ de rango completo, $b \in \mathbb{R}^{u \times 1} \Rightarrow$

$$y = Ax + b \sim N_m(A\mu + b, A\Sigma A')$$

Tengamos en cuenta las siguientes definiciones:

- Dados dos vectores aleatorios x e y de $m \times 1$ y $n \times 1$ el operador de covarianza generalizado es:

$$(Cov(x, y))_{ij} = Cov(x_i, y_j)$$

- luego
$$Cov(x, y) = E \left((x - \underbrace{\mu_x}_{E(x)}) (\underbrace{y - \mu_y}_{E(y)})' \right)$$

- Em estas condições, se $A \in \mathbb{R}^{l \times m}$ e $B \in \mathbb{R}^{p \times n}$ são duas matrizes constantes, então

$$\text{Cov}(AX, BY) = A \text{Cov}(X, Y) B'$$

- Función Característica de un v.a.

Dado x un s.a de $n \times 1$ de mínimos
para todo $t \in \mathbb{R}^n$

$$\varphi_x(t) = E(e^{it'x}) = E(e^{i \sum_{j=1}^n t_j x_j})$$

- Como en el caso univariado φ caracteriza
a la distribución : $F_x \leftrightarrow \varphi_x$

$$\therefore F_x = F_y \iff \varphi_x = \varphi_y$$

• Vejamos que si $z \sim N(0, I)$, entonces

$$\varphi_z(t) = e^{-\frac{1}{2} t' t} \quad \forall t \in \mathbb{R}^n.$$

$$\varphi_z(t) = E(e^{i t' z}) = E\left(e^{i \sum_{j=1}^n t_j z_j}\right) = E\left(\prod_{j=1}^n e^{i t_j z_j}\right)$$

↓ independencia

$$= \prod_{j=1}^n E(e^{i t_j z_j}) = \prod_{j=1}^n e^{-\frac{1}{2} t_j^2} = e^{-\frac{1}{2} \sum_{j=1}^n t_j^2}$$

↑ φ de $N(0, 1)$

$$= e^{-\frac{1}{2} t' t}$$

- Consideremos un vector $x \sim N(\mu, \Sigma)$
 donde Σ es definido positiva. Veamos
 que entonces

$$\varphi_x(t) = e^{it'\mu - \frac{1}{2}t'\Sigma t}$$

Basta descomponer $\Sigma = U \Delta U' = B B'$
 y considerar $E(e^{it' B B' (x - \mu + \mu)})$
 donde B es no singular

• Esto permite una definición más general:
 x es normal multivariado, $N_m(\mu, \Sigma)$,
donde Σ es semidefinida positiva o si
su función característica es de la forma

$$\varphi_x(t) = e^{it'\mu - \frac{1}{2}t'\Sigma t} \quad \forall t \in \mathbb{R}$$

- Como consecuencia de esta caracterización es fácil probar que:

Si $x \sim N_m(\mu, \Sigma)$ y $A \in \mathbb{R}^{p \times m}$, entonces

$$y = Ax \sim N(A\mu, A\Sigma A')$$

Em efecto:

Si $w = Ax$, luego

$$\begin{aligned}\varphi_w(t) &= \varphi_{Aw}(t) = E(e^{it'w}) = \\ &= E(e^{it'A x}) = \varphi_x(A't)\end{aligned}$$

$$= e^{it'A\mu - \frac{1}{2}t'A\Sigma A't}$$

que corresponde a la φ de una $N(A\mu, A\Sigma A')$

• Una propiedad interesante, a que le pasa a dos subvectores de una partición de un v.a. normal: Sea $X \sim N_m(\mu, \Sigma)$

y dividamos a $X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ y de manera

en forma

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \quad \text{donde}$$

$$\Sigma_{21} = \Sigma_{12}'$$

Tenemos que:

$$\psi_x(t) = e^{it'\mu - \frac{1}{2} t' \Sigma t}$$

, luego el
exponente puede escribirse como:

$$it'\mu - \frac{1}{2} t' \Sigma t = it'_1 \mu_1 + it'_2 \mu_2 - \frac{1}{2} t'_1 \Sigma_{11} t_1 \\ - \frac{1}{2} t'_2 \Sigma_{22} t_2 - t'_1 \Sigma_{12} t_2$$

Note mos que si $\Sigma_{12} = 0$, entonces

ψ_x se puede factorizar como:

$$\begin{aligned}\psi_x(t) &= e^{it_1'\mu_1 - \frac{1}{2}t_1'\Sigma_1t_1} \cdot e^{it_2'\mu_2 - \frac{1}{2}t_2'\Sigma_2t_2} \\ &= \psi_{x_1}(t_1) \cdot \psi_{x_2}(t_2)\end{aligned}$$

donde $x_1 \sim N(\mu_1, \Sigma_1)$

$$x_2 \sim N(\mu_2, \Sigma_2)$$

$$\forall t_1, t_2$$

$$\psi_x(t) = \psi_{x_1}(t_1) \cdot \psi_{x_2}(t_2) \quad \forall t = (t_1, t_2)$$

Como φ caracteriza lo dióti bncin, se dir

$$F_{x_1, x_2} \Leftrightarrow \varphi_{x_1, x_2}$$

tenemos que

$$\varphi_{F_1, F_2} = \varphi_{F_1} \cdot \varphi_{F_2} \Leftrightarrow F_{x_1, x_2} = F_{x_1} \cdot F_{x_2}$$

$\Rightarrow x_1$ y x_2 son independientes

• Además

Si X_1 y X_2 son independientes dados t_1 y t_2 :

$$\varphi_{\overset{\uparrow}{(X_1, X_2)}}(\overset{t}{t_1, t_2}) = E(e^{i\overset{t}{t_1}X_1 + i t_2 X_2}) = E(e^{i t_1 X_1} e^{i t_2 X_2}) \quad \forall t_1, t_2$$
$$= \varphi_{X_1}(t_1) \cdot \varphi_{X_2}(t_2)$$

$\Rightarrow$ para que valga $\otimes \quad \forall t_1, t_2 \Rightarrow \Sigma_{12} = 0$.

• Algo más sobre independencia

Si $X \sim N_m(\mu, \Sigma)$, $U = AX$, $W = BX$ entonces

U y W son independientes si

$$\text{Cov}(U, W) = A\Sigma B' = 0.$$

Em efecto $\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} AX \\ BX \end{pmatrix} = \underbrace{\begin{pmatrix} A \\ B \end{pmatrix}}_C X$

Notemos que CX es normal con media

$\begin{pmatrix} A \\ B \end{pmatrix} \mu$ y matriz de covarianza:

$$C \Sigma C' = \begin{pmatrix} A \\ B \end{pmatrix} \Sigma \begin{pmatrix} A' & B' \end{pmatrix} = \begin{pmatrix} A \Sigma A' & A \Sigma B' \\ B' \Sigma A & B \Sigma B' \end{pmatrix}$$

y por lo anterior son independientes U y V
 oii $A \Sigma B' = 0$.

• Show $x \sim N_m(\mu, \Sigma) \Rightarrow$

$$(x - \mu)' \Sigma^{-1} (x - \mu) \sim \chi^2_m$$

• $x \sim N_n(\theta, \sigma^2 I)$ y P es una matriz
idempotente y simétrica de rango r ,

entonces

$$\frac{(x - \theta)' P (x - \theta)}{\sigma^2} \sim \chi_r^2$$

Estos resultados nos permiten deducir intervalos de confianza o tests para cada uno de los coeficientes del modelo lineal:

Como $\hat{\boldsymbol{\theta}} \sim N_p(\boldsymbol{\theta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$, entonces

$$\hat{\theta}_i = \mathbf{e}_i' \hat{\boldsymbol{\theta}} \sim N(\theta_i, \sigma^2 \mathbf{e}_i' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{e}_i)$$

Si llamamos $\Sigma_{\hat{\boldsymbol{\theta}}} = \sigma^2 \mathbf{D}$

$$\hat{\theta}_i \sim N(\theta_i, \sigma^2 d_{ii})$$

siendo d_{ii} el i -ésimo elemento diagonal de $\mathbf{D}$.

Si para un i fijo quisiéramos testear

$$H_o : \theta_i = 0 \quad \text{vs.} \quad H_1 : \theta_i \neq 0$$

tenemos que bajo H_o

$$\frac{\hat{\theta}_i}{s\sqrt{d_{ii}}} \sim t_{n-p}$$

Por lo tanto, rechazaremos H_o con nivel α si

$$\left| \frac{\hat{\theta}_i}{s\sqrt{d_{ii}}} \right| > t_{n-p, \frac{\alpha}{2}}$$

Vayamos al ejemplo...

Funciones Lineales de los Parámetros

En muchas aplicaciones estamos más interesados en estimar funciones lineales de θ que en estimar θ en sí mismo:

$$c'\theta \text{ o } C\theta$$

Funciones Lineales de los Parámetros

En muchas aplicaciones estamos más interesados en estimar funciones lineales de θ que en estimar θ en sí mismo:

$$c'\theta \text{ o } C\theta$$

Un caso de especial interés:

En un futuro valor x_o , ¿qué esperamos observar? $\mathbb{E}(y_o) = ?$

Funciones Lineales de los Parámetros

En muchas aplicaciones estamos más interesados en estimar funciones lineales de θ que en estimar θ en sí mismo:

$$c'\theta \text{ o } C\theta$$

Un caso de especial interés:

En un futuro valor x_o , ¿qué esperamos observar? $\mathbb{E}(y_o) = ?$

$$\mathbb{E}(y_o) = x_o'\theta \Rightarrow \widehat{\mathbb{E}(y_o)} = x_o'\hat{\theta}$$

Notemos: $\hat{\theta}$ es insesgado entonces el estimador propuesto sería insesgado de $\mathbb{E}(y_o)$.

Podría haber muchos estimadores de una función lineal $c'\theta$ o $C\theta$, nos enfocaremos en los **estimadores lineales**: funciones lineales de $y_1, \dots, y_n$.

Funciones Lineales de los Parámetros

En muchas aplicaciones estamos más interesados en estimar funciones lineales de θ que en estimar θ en sí mismo:

$$c'\theta \text{ o } C\theta$$

Un caso de especial interés:

En un futuro valor x_o , ¿qué esperamos observar? $\mathbb{E}(y_o) = ?$

$$\mathbb{E}(y_o) = x_o'\theta \Rightarrow \widehat{\mathbb{E}(y_o)} = x_o'\hat{\theta}$$

Notemos: $\hat{\theta}$ es insesgado entonces el estimador propuesto sería insesgado de $\mathbb{E}(y_o)$.

Podría haber muchos estimadores de una función lineal $c'\theta$ o $C\theta$, nos enfocaremos en los **estimadores lineales**: funciones lineales de $y_1, \dots, y_n$.

Primero veremos cuando una función paramétrica es **estimable**.

Funciones Estimables

Estimadores Lineales Insesgados

Definición: Una **función paramétrica** ψ se dice que es una **función lineal** de los parámetros $\{\theta_1, \dots, \theta_p\}$ si existen $\{c_1, \dots, c_p\}$ constantes conocidas tal que

$$\psi = \mathbf{c}'\boldsymbol{\theta} = \sum_{j=1}^p c_j \theta_j$$

donde $\mathbf{c} = (c_1, \dots, c_p)'$.

Funciones Estimables

Estimadores Lineales Insesgados

Definición: Una **función paramétrica** ψ se dice que es una **función lineal** de los parámetros $\{\theta_1, \dots, \theta_p\}$ si existen $\{c_1, \dots, c_p\}$ constantes conocidas tal que

$$\psi = \mathbf{c}'\boldsymbol{\theta} = \sum_{j=1}^p c_j \theta_j$$

donde $\mathbf{c} = (c_1, \dots, c_p)'$.

Definición: Decimos que una función paramétrica $\psi = \mathbf{c}'\boldsymbol{\theta}$ es **estimable** si tiene un estimador lineal (en $\mathbf{Y}$) insesgado, es decir si existe $\mathbf{a} \in \Re^n$ tal que

$$E(\mathbf{a}'\mathbf{Y}) = \psi = \mathbf{c}'\boldsymbol{\theta} \quad \forall \boldsymbol{\theta} \in \Re^p$$

¿Hay funciones que no son estimables?

Funciones Estimables: Caracterización

El siguiente resultado caracteriza las funciones paramétricas estimables suponiendo el modelo

$$\Omega : E(\mathbf{Y}) = \mathbf{X}\theta \quad \Sigma_{\mathbf{Y}} = \sigma^2\mathbf{I}$$

Teorema: La función paramétrica $\psi = \mathbf{c}'\boldsymbol{\theta}$ es estimable si y sólo si $\mathbf{c}$ es una combinación lineal de las filas de $\mathbf{X}$, o sea si existe $\mathbf{a} \in \Re^n$ tal que

$$\mathbf{c}' = \mathbf{a}'\mathbf{X}$$

Funciones Estimables: Caracterización

El siguiente resultado caracteriza las funciones paramétricas estimables suponiendo el modelo

$$\Omega : E(\mathbf{Y}) = \mathbf{X}\theta \quad \Sigma_{\mathbf{Y}} = \sigma^2 \mathbf{I}$$

Teorema: La función paramétrica $\psi = \mathbf{c}'\boldsymbol{\theta}$ es estimable si y sólo si $\mathbf{c}$ es una combinación lineal de las filas de $\mathbf{X}$, o sea si existe $\mathbf{a} \in \Re^n$ tal que

$$\mathbf{c}' = \mathbf{a}'\mathbf{X}$$

Notemos que si el $rg(\mathbf{X}) = p$, toda función paramétrica $\psi = \mathbf{c}'\boldsymbol{\theta}$ es estimable .

Mejor Estimador Lineal Inssegado (BLUE)

Lema: Supongamos que vale el modelo Ω . Sean $\psi = \mathbf{c}'\boldsymbol{\theta}$ una función estimable y $\mathcal{V}_r$ el espacio generado por las columnas de $\mathbf{X}$ ($r = \text{rg}(\mathbf{X}) \leq p$).

Luego, existe un único estimador lineal inssegado de ψ , digamos $\mathbf{a}^{*\prime}\mathbf{Y}$ con $\mathbf{a}^* \in \mathcal{V}_r$. Más aún, si $\mathbf{a}'\mathbf{Y}$ es un estimador inssegado de ψ , $\mathbf{a}^*$ es la proyección ortogonal de $\mathbf{a}$ sobre $\mathcal{V}_r$.

Mejor Estimador Lineal Insesgado (BLUE)

Lema: Supongamos que vale el modelo Ω . Sean $\psi = \mathbf{c}'\boldsymbol{\theta}$ una función estimable y $\mathcal{V}_r$ el espacio generado por las columnas de $\mathbf{X}$ ($r = \text{rg}(\mathbf{X}) \leq p$).

Luego, existe un único estimador lineal insesgado de ψ , digamos $\mathbf{a}^{*\prime}\mathbf{Y}$ con $\mathbf{a}^* \in \mathcal{V}_r$. Más aún, si $\mathbf{a}'\mathbf{Y}$ es un estimador insesgado de ψ , $\mathbf{a}^*$ es la proyección ortogonal de $\mathbf{a}$ sobre $\mathcal{V}_r$.

Teorema de Gauss–Markov:

Supongamos que vale el modelo $\Omega : E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\theta} \quad \Sigma_{\mathbf{Y}} = \sigma^2\mathbf{I}$.

Toda función estimable $\psi = \mathbf{c}'\boldsymbol{\theta}$ tiene un único estimador $\hat{\psi}$ lineal insesgado de mínima varianza (BLUE). Este estimador $\hat{\psi}$ se puede obtener reemplazando a $\boldsymbol{\theta}$ en $\mathbf{c}'\boldsymbol{\theta}$ por $\hat{\boldsymbol{\theta}}$, el estimador de mínimos cuadrados.