Probabilidad y estadística

Clase 3

Estadística

¿Qué es la estadística?

Ahora lo que tenemos son observaciones (realizaciones) de las variables aleatorias.

El objetivo es poder hacer algún tipo de inferencia a partir de los valores observados.

Muestra aleatoria

Supongamos que tenemos un experimento aleatorio relacionado con una v.a. X. La v.a. X representa un observable del experimento aleatorio.

Los valores de X son la población de estudio, y el objetivo es saber como se comporta esa población.

Una muestra aleatoria de tamaño n, es una sucesión de n v.a **independientes** $\underline{X}_n = (X_1, \ldots, X_n)$, tal que $X_i \sim X$

Estimadores puntuales

Estimadores puntuales

$X \sim \overbrace{F(\theta)}$ } No lo conocem

aprender a partir de la m.a.

• m.a. $\to \{X_1, \dots, X_n\} = \underline{X}_n$; X_i son i.i.d.

$\overbrace{X_1 \sim F_X \quad \vdots \quad X_i \sim F_X}$

$\hat{\theta}(\underline{X}_n)$ }

Def: Dada una muestra aleatoria $\underline{X}_n$, un estimador $\hat{\theta}$ de un parámetro θ es una función de la muestra aleatoria que provee un valor aproximado del parámetro o característica desconocido.

Bondades de los estimadores

$\mathbb{E}_{X \sim F_\theta}\left[\hat{\theta}(\underline{X}_n)\right]$ como $X_1, \dots, X_n$ con $X_i \sim X$

poblacional

θ

Def: $\hat{\theta}$ es un estimador **insesgado** para θ si $B := \mathbb{E}[\hat{\theta} - \theta] = 0 \ \forall \, \theta$.

$\theta(\underline{X}_n)$

En caso contrario diremos que es **sesgado**. A B se lo conoce como **sesgo**

Def: $\hat{\theta}$ es un estimador **asintóticamente insesgado** para θ si

$$\lim_{n \to \infty} \mathbb{E}[\hat{\theta} - \theta] = 0 \ \forall \, \theta$$

Bondades de los estimadores

$\circledast\ 2E[\underbrace{\hat\theta - E\hat\theta}_{E\hat\theta - E\hat\theta = 0}](E\hat\theta - \theta) = 0$

Def: El error cuadrático medio (ECM) como $\mathbb{E}[(\hat\theta-\theta)^2] =$

$= E[((\hat\theta - E[\hat\theta]) + (E[\hat\theta] - \theta))^2]$

$= E[(\hat\theta - E\hat\theta)^2 + 2(\hat\theta - E\hat\theta)(E\hat\theta - \theta) + (E\hat\theta - \theta)^2]$

linealidad

$= var(\hat\theta) + B^2$

Obs: El ECM se puede descomponer como:

$$ECM = \underbrace{var(\hat\theta)}_{varianza} + \underbrace{B(\hat\theta)^2}_{sesgo}$$

$=0$

de compromiso sesgo - varianza

donde $var(\hat\theta) = \mathbb{E}[(\hat\theta - \mathbb{E}[\hat\theta])^2]$ y $B = \mathbb{E}[\hat\theta - \theta]$

Def: Un estimador $\theta^*(\underline{X}_n)$ es **óptimo** (en media cuadrática) si

$ECM(\theta^*) \leq ECM(\hat\theta)$ para todo $\hat\theta(\underline{X}_n)$

Bondades de los estimadores

$\hat{\theta}_1, \hat{\theta}_2 \dots \hat{\theta}_n$

Def: Dada una sucesión de estimadores $\hat{\theta}_n$ de θ, diremos que $T = \hat{\theta}$ es (débilmente) consistente si $\forall\, \varepsilon > 0, \mathbb{P}(|T - \theta| > \varepsilon) \to 0$ (es consistente en prob)

Teorema: Si $var(\hat{\theta}) \to 0$ y $\mathbb{E}[\hat{\theta}] \to \theta$, entonces $\hat{\theta}$ es consistente. ⇒ es consistente en

Def: Un estimador es consistente en media cuadrática si

$$\lim_{n\to\infty} ECM(\hat{\theta}_n) = 0, \forall\theta \quad \Rightarrow \text{consistencia en prob.}$$

$$E[(\hat{\theta}_n - \theta)^2] = \underbrace{Var(\hat{\theta})}_{\xrightarrow[n\to\infty]{} 0} + \underbrace{B^2}_{\xrightarrow[n\to\infty]{} 0} \xrightarrow[n\to\infty]{} 0$$

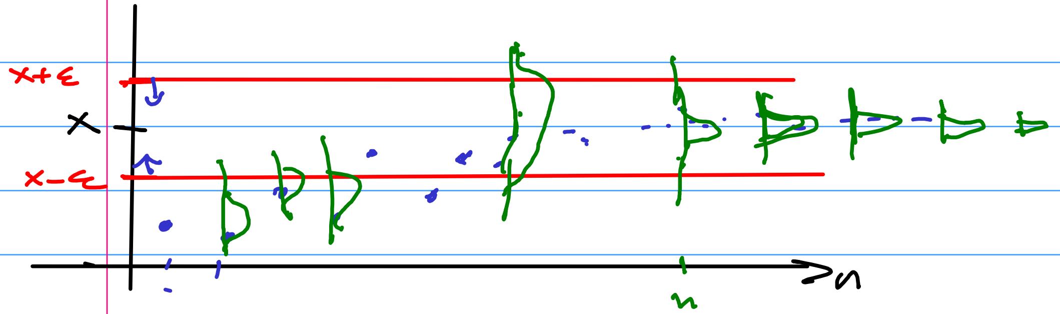
• Convergencia en media cuadrática

$\Downarrow$

• convergencia en $\mathbb{P}$:

$X_n \xrightarrow{P} X$ si dado $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$$



• convergencia en distribución

(TCL)

$$\lim_{n\to\infty} F_n = F$$

Ejercicio 1

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Se desea estimar la media de una variable con distribución N(μ,9) a partir del promedio de n realizaciones. Analizar las bondades de las que goza dicho estimador.

a) • $E[\overline{X}_n] \overset{\text{lin}}{=} \frac{1}{n}\sum_{i=1}^{n} E[X_i] \overset{\text{iid}}{=} \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{n\mu}{n} = \mu$

$\Rightarrow B = 0$, $\overline{X}_n$ es insesgado

b) • $Var(\overline{X}_n) \overset{\text{indep}}{=} \frac{1}{n^2}\sum_{i=1}^{n} \underbrace{Var(X_i)}_{9} = \frac{1}{n^2} n \cdot 9 = \frac{9 = \sigma^2}{n} \xrightarrow[n\to\infty]{} 0$

Por a) y b), $\overline{X}_n$ es consistente.

Algunos comportamientos límite

LGN + TCL

Ley (débil) de los grandes números

Sea $\underline{X}_n = (X_1, \ldots, X_n)$ una muestra aleatoria, con X_i i.i.d. con media μ y variancia σ^2. Para $\epsilon > 0$ se tiene que

$$\mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) \xrightarrow[n \to \infty]{} 0 \qquad (\bar{X}_n \text{ es consistente})$$

Es decir, a medida que crece la cantidad de muestras, el promedio $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ tiende a la media real de la distribución.

Teorema central del límite

$$\textcircled{*}\ \boxed{\lim_{n\to\infty} F_{\theta_n} = F_\theta}$$

$$\lim_{n\to\infty} F_{X_n} = F_X$$

Sea $\underline{X} = (X_1, \ldots, X_n)$ una muestra aleatoria, con X_i i.i.d. con media μ y varianza $\sigma^2 < \infty$. Sea $\boxed{\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i}$, luego

$X_n \sim N(\mu, \frac{\sigma^2}{n})$

$$\boxed{Z_n = \frac{\bar{X}_n - \mu}{\sqrt{\mathbb{V}(\bar{X}_n)}}} = \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \rightsquigarrow Z$$

estandarizada

con $Z \sim \mathcal{N}(0,1)$, o equivalentemente:

$$\textcircled{*}\ \lim_{n\to\infty} \overbrace{\mathbb{P}(Z_n \leq z)}^{F_{Z_n}} = \Phi(z) \doteq \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

Es decir, $\bar{X}_n$ tiene una distribución aproximadamente Normal, con media μ y variancia σ^2/n . Para n finito, grande podemos aproximar

$$\mathbb{P}\left(\frac{\overbrace{\sum_{i1}^{n} X_i}-n\mu}{\sqrt{n}\sigma} \leq z\right) \approx \Phi(z), \qquad \mathbb{P}\left(\sqrt{n}\frac{\bar{X}_i-\mu}{\sigma} \leq z\right) \approx \Phi(z),$$

Enfoque frecuentista

Estadístico suficiente

$\hat{\theta} \to \theta$

Def: Dada una muestra aleatoria $\underline{X}_n$, un estadístico es cualquier función $T_n = T(\underline{X}_n)$

Def: Sea una muestra aleatoria $\underline{X}_n$, cuya distribución es $F_\theta(\underline{x}), \theta \in \Theta$, se dice que $T = r(\underline{X}_n)$ es un estadístico suficiente para θ si $F_{\underline{X}|T=t}(\underline{x})$ no depende de θ.

Teorema de factorización: Diremos que $T = r(\underline{X}_n)$ es un est. suficiente para θ sii existen funciones h y g tales que:

$$\boxed{f_\theta(\underline{x})} = g(r(\underline{x}), \theta)h(\underline{x})$$

Ejercicio 2

$f_\theta(\underline{x}_n) \overset{\text{indep.}}{=} \prod_{i=1}^{n} f_\theta(x_i)$ — conjunta, indep., marginales

$f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$ si X e Y son indep

1) $f_\lambda(\underline{x}) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} \mathbb{1}\{x_i \geq 0\} = \underbrace{\lambda^n e^{-\lambda \sum x_i}}_{g(\lambda, T(\underline{x}))} \underbrace{\prod_{i=1}^{n} \mathbb{1}\{x_i \geq 0\}}_{h(\underline{x})}$

$\Rightarrow T(\underline{x}) = \sum_{i=1}^{n} x_i$.

1. Hallar un estadístico suficiente para el parámetro de una v.a. con distribución exponencial
2. Hallar un estadístico suficiente para el parámetro de una v.a. con distribución de Bernoulli
3. Hallar un estadístico suficiente para el parámetro de una v.a. con distribución Uniforme en el intervalo (0,θ)

2) $f_p(\underline{x}) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum x_i}(1-p)^{n-\sum x_i} \Rightarrow T(\underline{x}) = \sum_{i=1}^{n} x_i$

3) $f_\theta(\underline{x}) = \prod_{i=1}^{n} \frac{1}{\theta} \underbrace{\mathbb{1}\{0 \leq x_i \leq \theta\}}_{\mathbb{1}\{0 \leq x_i\} \mathbb{1}\{x_i \leq \theta\}} = \frac{1}{\theta^n} \overbrace{\mathbb{1}\{0 < \min_{i=1,\dots,n}(x_i)\}}^{h(\underline{x})} \mathbb{1}\{\max_{i=1,\dots,n}(x_i) < \theta\}$

$g(\theta, T(\underline{x}_n)) \Rightarrow T(\underline{x}_n) = \max_{i=1,\dots,n}(x_i)$

Método de máxima verosimilitud

Método de Máxima Verosimilitud

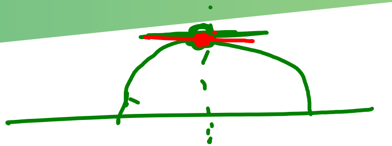
Def: Diremos que $\hat{\theta}(\underline{X})$ es un Estimador de Máxima Verosimilitud (MLE) si se cumple que:

$$f(\underline{X}, \hat{\theta}) = \max_{\theta} f_{\theta}(\underline{X})$$

La idea es que si observé una determinada muestra, entonces esta debería tener alta probabilidad de ocurrir, por lo tanto busco el θ que maximiza esa probabilidad de ocurrencia

Método de Máxima Verosimilitud

$1_{\{a<x<b\}}$



Def: Definimos la función de verosimilitud como

$\boxed{L(\theta) = f(\underline{x}, \theta)}$ (vista como función de θ) luego,

$\hat{\theta} = \arg\max L(\theta).$ obs: que $L(\Theta)$ es concava en Θ

$\theta \in \Theta$ → es el espacio de parámetros (ej. Bernoulli $p \in [0,1] = \Theta$)

Si el soporte de X no depende θ, Θ es un conjunto abierto y $f_\theta(x)$ es derivable respecto de θ, entonces para hallar el EMV puedo hallar θ tal que

$$\left[\ \frac{\partial\, ln(L(\theta))}{\partial\, \theta} = 0\ \right]$$

Ejercicio 3

continuar:
verificar que $\mathcal{L}(\hat{p}) > \mathcal{L}(0)$
y $\mathcal{L}(\hat{p}) > \mathcal{L}(1)$

La probabilidad de acertar a un blanco es p. Se realizan 10 tiros independientes, en los cuales se observaron 4 aciertos. A partir este valor observado estimar el valor de p por MV.

$X \sim \text{Bernoulli}(p)$, $\Theta = [0,1]$ $\Rightarrow$

$\underline{X}_{10}$ m. a. con $X_1 \sim X$ se observó $T(\underline{x}) = \sum_{i=1}^{10} x_i = 4$

——o—— consideramos $\Theta = (0,1)$

$$\mathcal{L}(p) = \prod_{i=1}^{10} p^{x_i}(1-p)^{1-x_i} = p^{\sum x_i}(1-p)^{10-\sum x_i} = p^4(1-p)^6$$

$$\ell(p) = \log \mathcal{L}(p) = 4 \log p + 6 \log(1-p)$$

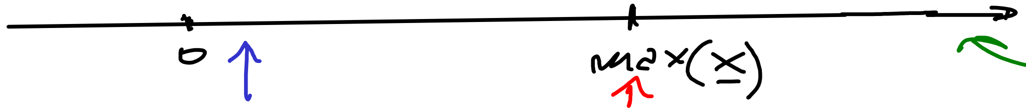
$$\frac{d}{dp}\ell(p) = \frac{4}{p} - \frac{6}{1-p} = 0 \Rightarrow 4(1-p) = 6p \Rightarrow 4 = 10p \Rightarrow \boxed{\hat{p} = 0{,}4}$$

Ejercicio 4

1. Sea X~U(0, θ). Hallar el EMV para θ basado en una muestra de tamaño n.
2. Si en una muestra de tamaño 3 se observaron los valores (0.756, 1.1, 3,11), hallar el valor estimado de θ.

$X \sim U(0,\theta)$ y $\underline{X}_n$ es una m.a con $X_i \sim X$

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} \mathbb{1}_{\{0<x_i<\theta\}} = \frac{1}{\theta^n} \mathbb{1}_{\{\min(x_i) \geq 0\}} \mathbb{1}_{\{\max(x_i) \leq \theta\}}$$



$\mathcal{L}(\theta) = 0$

$$\boxed{\mathcal{L}(\max \underline{x}) = \frac{1}{\max(\underline{x})^n}} > \mathcal{L}(\theta)$$

$\therefore \hat{\theta} = \max(\underline{x})$

Principio de invariancia

Supongamos que ahora queremos estimar por máxima verosimilitud a $\lambda = q(\theta)$.

EMV

EMV

x
cmzto

Teorema: Si $\hat{\theta}$ es MLE de θ, entonces $\hat{\lambda} = q(\hat{\theta})$.

¿Por qué es útil? Por ejemplo, podría querer estimar una probabilidad de la v.a. X, que en general no puedo porque desconozco el parámetro de la distribución.

$$P(X \in A) = \int_A f_\theta(x)\,dx = q(\theta) \Rightarrow \hat{P}(X \in A) = q(\hat{\theta}) = \int_A f_{\hat{\theta}}(x)\,dx$$

Ejercicio 5

$$P(Y=y) = p\,(1-p)^{y-1}$$

Siguiendo el ejercicio 3, estimar la probabilidad de que se necesiten al menos 2 tiros para observar el primer acierto.

$(\hat{p} = 0{,}4$ en el MLE$)$

$Y \sim Geom(p)$

$$P(Y \geq 2) = 1 - P(Y < 2) = 1 - \underbrace{P(Y=1)}_{p} = 1 - p$$

ppio de invarianza

$$\hat{P}(Y \geq 2) \stackrel{\downarrow}{=} 1 - \hat{p} = 0{,}6$$

Estimadores de cuadrados mínimos

Estimador de cuadrados mínimos

Es común querer estimar el valor de una v.a. X a partir de una medición Y. Ejemplo: Y es una versión ruidosa de X.

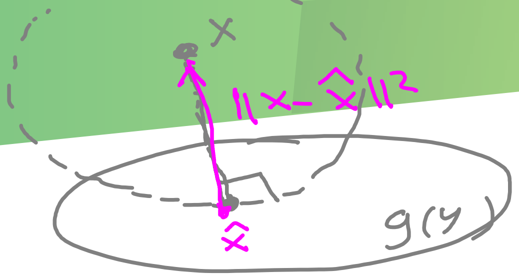
Buscamos un estimador $\hat{X}(y)$ de X tal que tenga mínimo error cuadrático medio

$$\mathrm{ECM} = \mathbb{E}[(X - \hat{X})^2 \mid Y] = \|x - \hat{x}\|^2$$

Observar que se corresponde con la distnacia asociada al p.i. canónico para v.a. en $\langle x, y \rangle = E[xy]$

$$\|x\|^2 = E[x^2]$$

Estimador de mínimos cuadrados



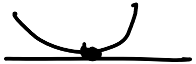
En otras palabras, queremos $\hat{X} = g^*(Y)$ tal que

$$\mathbb{E}[(X-\hat{X})^2 \mid Y] \leq \mathbb{E}[(X-g(Y))^2 \mid Y] \qquad \forall g(Y) \text{ (medible)}$$

¿Quién era $\hat{X}$? $\boxed{\hat{X} = \mathbb{E}[X|Y]}$

Idea de demostración: [Ejercicio]

1. Probar que el mejor estimador constante es $\mathbb{E}[X]$
2. Probar que el mejor estimador condicional es $\mathbb{E}[X|Y=y]$.
3. Dejar que Y tome todos los valores posibles (i.e. reemplazo y por Y), recupero la esperanza condicional.



1) $ECM = E\left[(x-c)^2\right] \Rightarrow \frac{d}{dc} ECM = 0$

$$E\left[\partial (x-c)^2\right] = E\left[2(x-c)\right] = 0$$

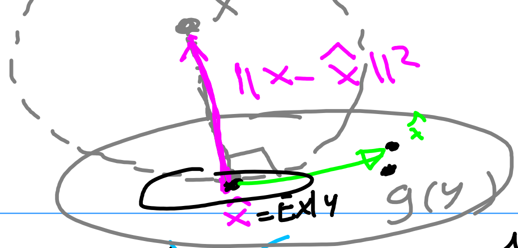
$$\boxed{E\,x = c}$$

2) $ECM = E\left[\left(x - \underbrace{\hat{x}(y)}\right)^2 \,\middle|\, Y = y\right] \Rightarrow \frac{d}{d\hat{x}(y)} ECM = 0$

$$\Rightarrow E\left[2\left(x - \hat{x}(y)\right) \middle| Y\right] = 0$$

$$\boxed{E\left[X \mid Y = y\right] = \hat{x}(y)}$$

3) $\hat{X} = E\left[X \mid Y\right]$



$$ECM = E[(x-\hat{x})^2 \mid y] = E[((x - E\,x|y) + (E\,x|y - \hat{x}))^2 \mid y]$$

$$= E[(x-E\,x|y)^2] + 2E[\underbrace{(x-E\,x|y)(E\,x|y-\hat{x})}_{\langle x - E\,x|y,\; E\,x|y - \hat{x} \rangle = 0}] + E[(E\,x|y-\hat{x})^2]$$

$$\min_{\hat{x}\in g(y)} E[(x-E\,x|y)^2] + E[(E\,x|y-\hat{x})^2] = \min_{\hat{x}} E[(E\,x|y-\hat{x})^2]$$

$$= E\,x|y$$

Mínimos cuadrados: caso lineal

$x_1, x_2 \sim N(\underline{\mu}, \Sigma)$, $x_1 | x_2 \sim N(E x_1|x_2, \Sigma_{12})$

$\hat{X} = \frac{cov(XY)}{var(Y)}(Y - EY) + E[X]$

A veces obtener $\mathbb{E}[X|Y]$ puede ser muy complicado, entonces nos restringimos a los estimadores lineales.

Buscamos a, b tq $\overset{ECM}{=} \mathbb{E}[(X - \underbrace{(aY + b)}_{\hat{X}})^2 \mid Y]$ sea mínima.

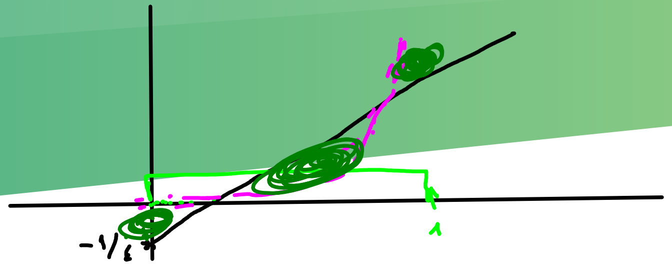
Resulta que $a = \frac{cov(X,Y)}{var(Y)}$ y $b = -\frac{cov(X,Y)}{var(Y)}\mathbb{E}[Y] + \mathbb{E}[X]$

Obs: Si se asume que X e Y son conjuntamente gaussianas, el estimador de mínimos cuadrados (la esperanza condicional) coincide con el estimador de mínimos cuadrados asumiendo un modelo lineal y puede obtenerse a partir del modelo condicional de $\mathbb{E}[X \mid Y]$

$X, Y \sim N(\underline{\mu}, \Sigma)$
Normal multiv.

es lineal

Ejercicio 6



Sea $X \sim U(0,1)$ e $Y=X^2$. Hallar la mejor aproximación lineal de Y basada en X. Comparar con la mejor estimación de Y basada en X.

$$\hat{Y} = \frac{\mathrm{cov}(X,Y)}{\mathrm{Var}\,X}(X - EX) + EY$$

$$\mathrm{cov}(X,Y) = E[(X-EX)(Y-EY)] = E[XY] - E[X]E[Y] = \frac{1}{4} - \frac{1}{2}\,\frac{1}{3} = \frac{1}{12}$$

$$E[XY] = E_X\big[E[XY|X]\big] = E\big[X\,E[\underbrace{Y}_{X^2}|X]\big] = E[X^3] = \int_0^1 x^3\,dx = \frac{x^4}{4}\Big|_0^1 = \frac{1}{4}$$

$$E[Y] = E[X^2] = \int_0^1 x^2\,dx = \frac{x^3}{3}\Big|_0^1 = \frac{1}{3}$$

$$EX = 1/2$$

$$\mathrm{Var}\,X = \frac{1}{12} \qquad \hat{Y} = \left(X - \frac{1}{2}\right) + \frac{1}{3} = X - \frac{1}{6}$$

Regresión lineal (OLS)

$\langle a, b \rangle = a^T b$

$\|a\|^2 = \sum a_i^2$ (Norma 2)

$\boxed{\varepsilon_i \sim \mathcal{N}(0, \sigma^2)}$ iid

Contamos con n observaciones conjuntas $(y_1, x_1), (y_2, x_2), \cdots, (y_n, x_n)$
el modelo de regresión lineal es

$\boxed{p+1 \times r}$

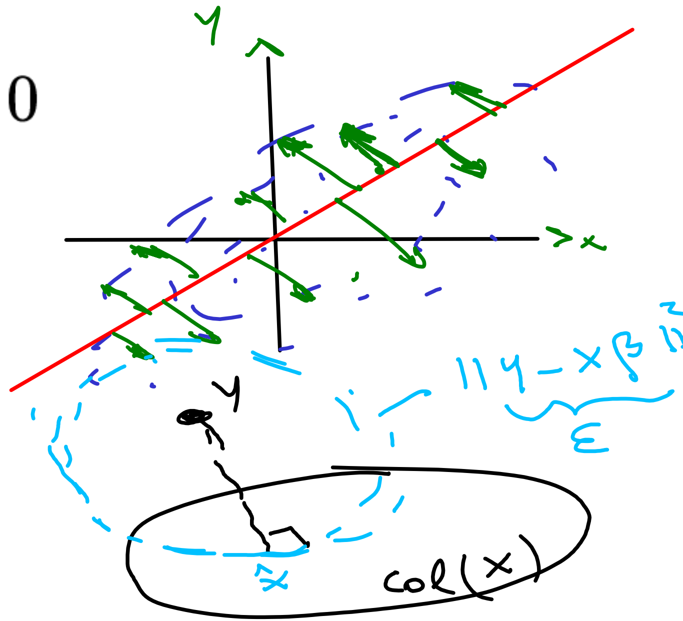
$$Y = X\beta + \varepsilon \quad \text{con} \quad \mathbb{E}(\varepsilon_i \mid X_i) = 0$$

donde $n \times r$ $\quad n \times (p+1)$

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

El estimador de cuadrados mínimos de $\hat{\beta}$ está dado por

$$\hat{\beta} = (X^\top X)^{-1} X^T Y$$



- **Obs**: Si se asume que $\boxed{\varepsilon_i \sim \mathcal{N}(0, \sigma^2)}$ y son i.i.d., este estimador coincide con el estimador de máxima verosimilitud. $\rightarrow Y|X \sim \mathcal{N}(X\beta, \cancel{\sigma^2 I})$ $\rightarrow \sum \{(x_1, y_1), \ldots, (x_n, y_n)\}$

Bibliografía

Bibliografía

- Notas de estadística. Graciela Boente y Víctor Yohai
- "All of Statistics : A Concise Course in Statistical Inference", Larry Wasserman, Springer-Verlag New York Inc.
- "Mathematical statistics with applications", Mendenhall, W., Scheaffer, R. L., & Wackerly, D. D. Boston: Duxbury Press.