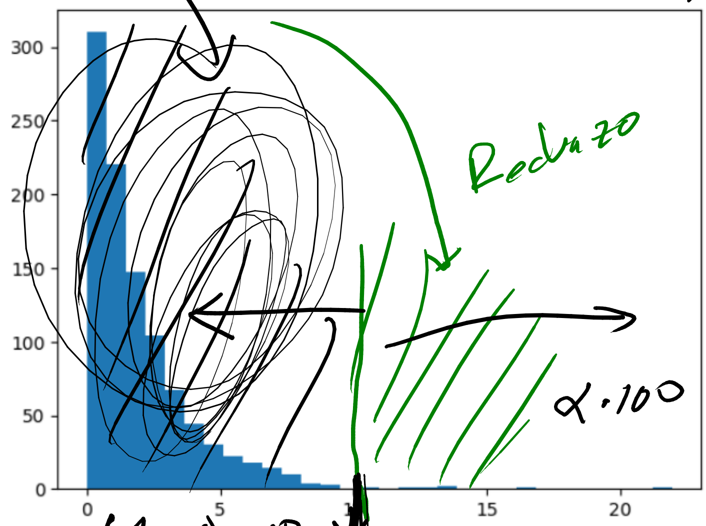
Aceptación

Al 95% de confianza x es un outliers.

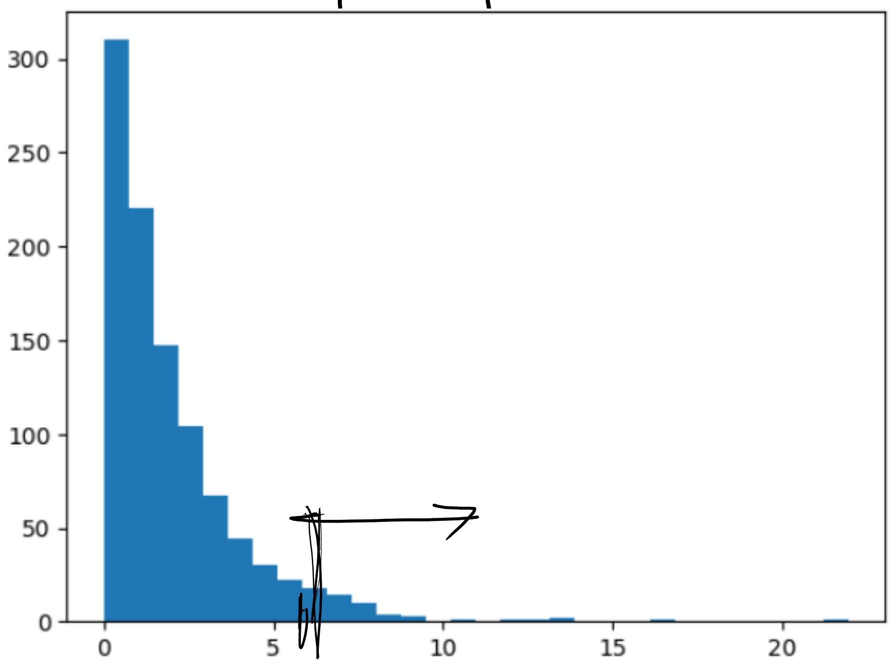
x

Rechazo



$x_{1-\alpha}$

```
alfa = 0.05
```




```
alfa = 0.05
```

Si $x \sim N(\mu, \sigma)$

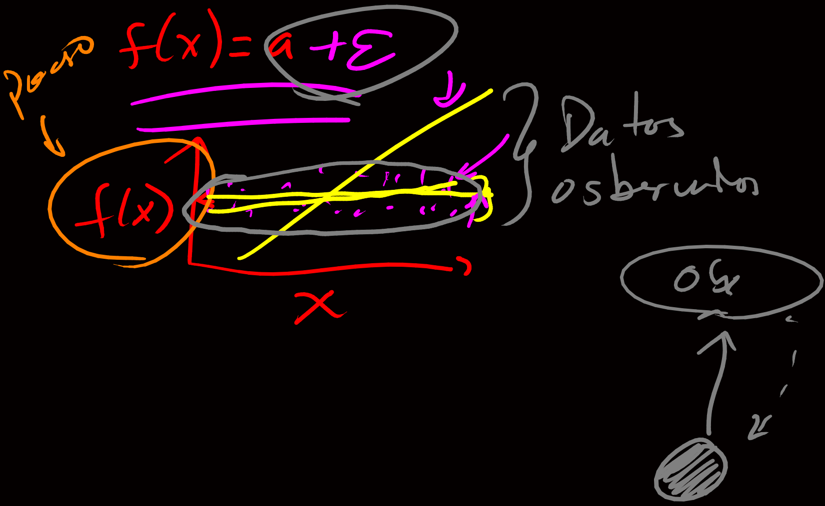
$$\Rightarrow x^2 \sim \chi^2(df)$$

$$(\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu}) \qquad (\vec{x}_0)^2 \text{ — chi}^2$$

$\vec{x}_1$ $\vec{x}_0$

3. Teoría de Estimación

3.1 Función de Verosimilitud



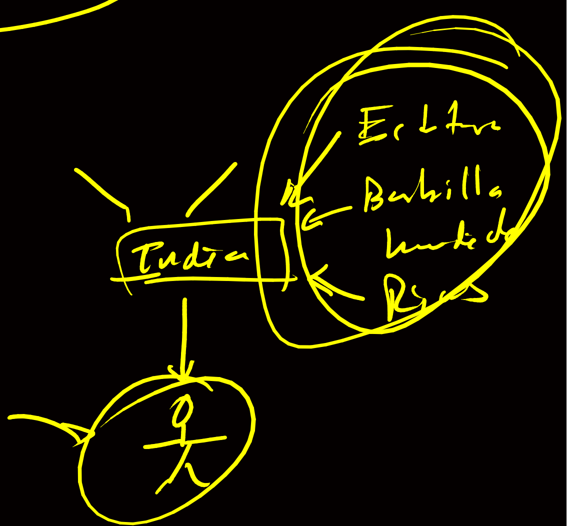
- Cuando tenemos datos, la cantidad de procesos generadores (modelos) candidatos es muy grande

- **Modelos de probabilidad:** ¿Cuál la probabilidad de observar x datos dados los parámetros que conocemos? $X \sim chi2(k)$ $p(x;k)$

- **Modelos estadísticos:** ¿Cuáles son los valores más plausibles de los parámetros dados los datos que observamos?

$\Rightarrow$

$$X \sim chi2(?)$$
$$X \sim Exp(?)$$
$$\vdots$$



Ejemplo: ¿Cuán probable es que obtengamos 9 caras si lanzamos 10 veces una moneda "justa" ($\theta = 0.5$)?

$n = 10$

$K = 9$

$\theta = 0.5$

0.5 _ _ _ _ _ _ _ _ _

0.5^9

Obtuve 9 éxitos al lanzar 10 veces una moneda ¿crees que la moneda es justa?

= ¿$\theta = 0.5$? y si no, ¿qué otro valor de θ es más creíble?

likely ← likelihood function

$X = \{0, 1\}$ ⇐ ensayos de Bernoulli

Distribución binomial

$K = \sum_{i=1}^{n} x_i$

Constante de normalización

$$P(K, n, \theta) = \binom{n}{k} \underbrace{\theta^K (1-\theta)^{n-K}}_{} \Big\} \text{Dist. Bino.}$$

Coeficiente Binomial

$n! = n(n-1)(n-2)\cdots 1$

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

n: tamaño de muestra (total de ensayos)

K: cantidad de éxitos

θ: define la probabilidad de un éxito

El valor más probable de K

$$\{\bar{x} = n \cdot \theta\} \Leftarrow 4 = 20 \cdot \theta$$

$$\theta = \frac{4}{20} = 0.2$$

Estimador de máxima verosimilitud

Maximum likelihood Estimation

MLE

$$\text{Datos} \xrightarrow{?} \text{parámetros}$$

- ¿Hay alguna forma de estimar qué θ(s) originaron un resultado?
- Si tenemos la hipótesis

$$f = 0.5$$

¿hay forma de saber si esa hipótesis podría ser falsa?

Función de verosimilitud

likelihood

$$L(\theta | X)$$

La fn. de verosimilitud deja fijo $\underline{X}$ (ya fue observado) y trata como variable a θ

θ no es una variable aleatoria, es tratado como desconocida.

$x_1, x_2, \dots, x_n$

$P(x_1;\theta)$ $\quad p(x_2;\theta)$ ....

$p(A_1 \cap A_2) = p(A_1)\,p(A_2)$

La función de verosimilitud se define como la función de densidad conjunta

$$L(\theta \mid X) = f(x_1, x_2, \dots, x_n)$$

$$L(\theta \mid X) = \underbrace{f(x_1 \mid \theta) \cdot f(x_2 \mid \theta) \cdot \ldots \cdot f(x_n \mid \theta)}_{\text{Producto}}$$

$$L(\theta) = \prod_{i=1}^{n} f(x_i \mid \theta)$$

$$\log(x \cdot y) = \log(x) + \log(y)$$

Regla log

$$\log\left(\frac{x}{y}\right) = \log(x) - \log(y)$$

$$\log(x \cdot y \cdot z \cdot w) = \log(x) + \log(y) + \log(z) + \log(w)$$

$$\log(x_1 \cdot x_2 \cdot x_3 \cdots x_n) = \sum_{i=1}^{n} \log(x_i)$$

$$\underbrace{\log[L(\theta)]} = \log\left[\prod_{i=1}^{n} f(x_i|\theta)\right]$$

$$l(\theta) = \sum_{i=1}^{n} \log(f(x_i|\theta))$$

log-likelihood

Matemáticamente (comp.) es más sencillo tratar con sumas.

El $\log(x)$ es monotónico con respecto a x.

Si en $f(x)$ existe un máximo/mínimo, en $f(\log(x))$ también existe.

$$p(x|y) = \frac{\boxed{p(x,y)}}{p(y)}$$ Verosim.

$$\mathcal{L}(\theta)\uparrow$$

$\mathcal{L}(\theta=0.5)$, $\mathcal{L}(\theta=0.9)$, $K=9$, $n=10$

Proceso de Bernoulli

$X_1 = 1$, $X_2 = 1$, $X_3 = 0$, $X_4 = 1$, $X_5 = 1$, $X_6 = 1$, $X_7 = 1$, $X_8 = 1$,
$X_9 = 1$, $X_{10} = 1$

$X = \{0, 1\}$

distribución de Bernoulli

Binarios

Bernoulli ⟷ Binomial

$$f(x) = \theta^x (1-\theta)^{1-x}$$

$x = \{0, 1\}$

Binomial = Bernoulli
$n = 1$, $k = \{1, 0\}$

$K = \sum_{i=1}^{n} x_i$, si $x = 1$

La función de verosimilitud

$$\mathcal{L}(\theta) = f(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} \theta^{x_i} (1-\theta)^{1-x_i}$$

$$\sum_{i=1}^{n} 1 = n \cdot 1$$

Dado que x es iid

inde. e identicamente distribuido

$$a_1 \cdot a_2 \cdot a_3 \cdots a_n = a^n$$

$$a^x \cdot a^x \cdots = a^{n \cdot x}$$

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} \theta^{x_i} (1-\theta)^{1-x_i}$$

$= a^{x+x+x\dots}$

$\theta^{x_1} \theta^{x_2} \dots \theta^{x_n} = \theta^{x_1+x_2\dots}$

$= \theta^{\boxed{\sum_{i=1}^{n} x_i}}$

$=$

Producto de $\theta^{x_1} \cdot \theta^{x_2} \dots \theta^{x_n}$

Producto de $\underbrace{(1-\theta)^{(1-x_1)} \cdot (1-\theta)^{(1-x_2)} \dots (1-\theta)^{(1-x_n)}}$

$(1-\theta)^{\sum_{i=1}^{n}(1-x_i)}$

$$\sum_{i=1}^{n}(1-x_i) = \boxed{\sum_{i=1}^{n} 1} - \sum_{i=1}^{n} x_i$$

$$\mathcal{L}(\theta) = \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n-\sum_{i=1}^{n} x_i} \quad ; \quad \sum_{i=1}^{n} x_i = 9, \quad n = 10$$

$$\mathcal{L}(\theta = 0.5) = 0.5^{9} (1-\theta)^{10-9} =$$

3.2 Máxima verosimilitud

Maximum Likelihood Estimation (MLE)

El problema de MLE consiste en lo siguiente.

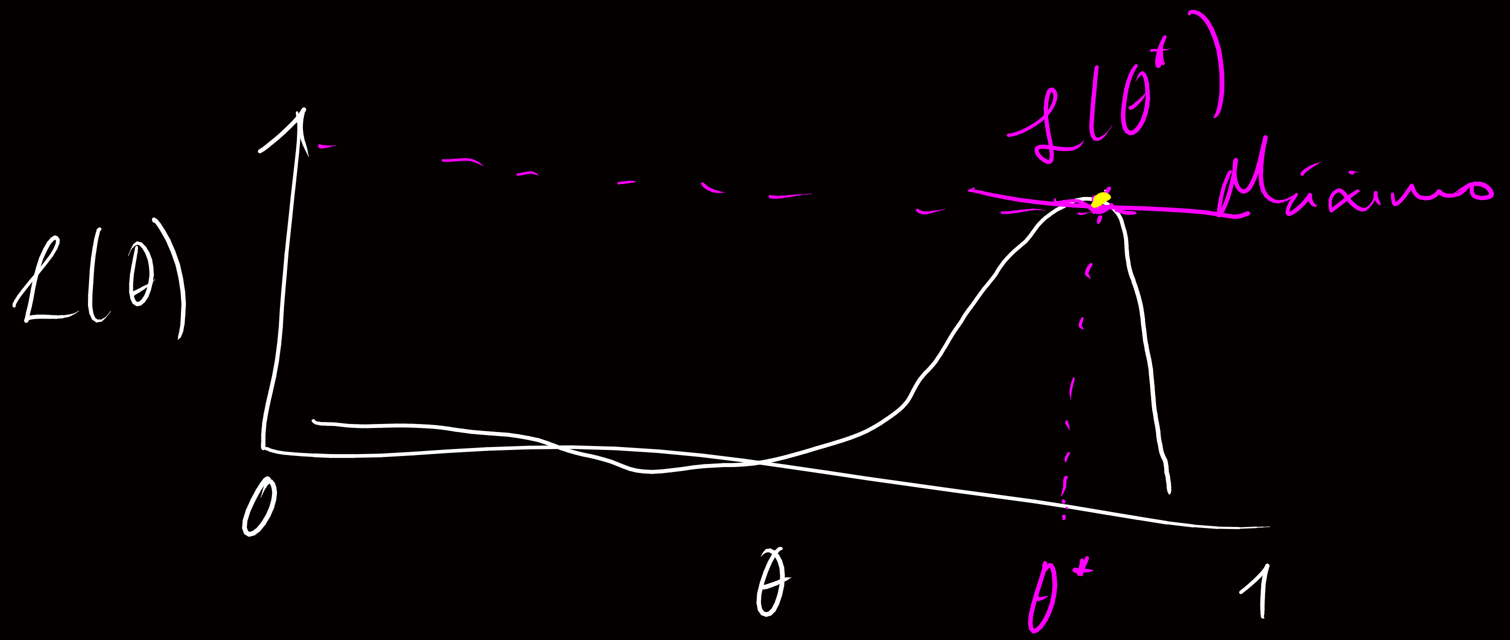
Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria de variables iid con distribución conjunta $f(x_1, x_2, \dots, x_n)$ con función de verosimilitud $\mathcal{L}(\theta) = \prod_{i=1}^{n} f(x_i|\theta)$

El estimador de máxima verosimilitud de θ es la solución al problema de optimización

θ^*

$$\theta^* = \arg\max \mathcal{L}(\theta)$$

θ^* es el argumento de la función $\mathcal{L}(\theta)$ que la maximiza



MLE para binomial

$$k = \sum_{i=1}^{n} x_i$$

Media aritmética

$$\theta^* = k/n = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\boxed{\sum_{i=1}^{n} p(x_i)\, x_i}$$

Esperanza de X

$$p(k; n, \theta) = \boxed{\binom{n}{k}} \theta^k (1-\theta)^{n-k} \quad \text{(Binomial)}$$

← no está en función de θ

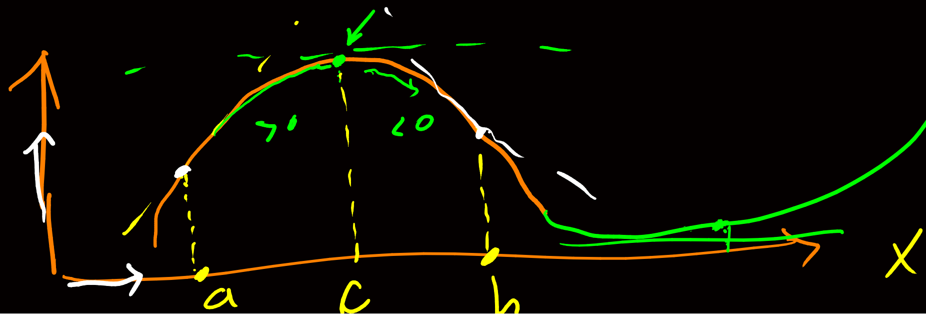
$$X_1, X_2, \ldots, X_n$$

$$L(\theta) = f(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} f(x_i \mid \theta)$$

cóncava   convexa

$$L(\theta) = \theta^k (1-\theta)^{n-k}$$

si sabemos que $L(\theta)$ es cóncava, podemos usar el criterio de 1a derivada para encontrar el máximo de $L(\theta)$



$f'(a) > 0, \quad f'(b) < 0$

$\boxed{f'(c) = 0}$ ← El criterio de 1° derivada

1ª Derivada: es la pendiente de recta tangente que pasa por un punto.

$$\boxed{\left.\frac{\partial L(\theta)}{\partial \theta}\right|_{k,n} = 0}$$

Para MLE

1. Encontrar $L(\theta)$
2. Convertirla en $\ell(\theta)$ — conf. más sencillo
3. Aplicar el CPD
4. Resolver para θ^*

Paso 2

$$\ell(\theta) = \log(L(\theta)) = \log\left(\underbrace{\theta^k}\underbrace{(1-\theta)^{n-k}}\right)$$

$$= \log(\theta^k) + \log\left[(1-\theta)^{n-k}\right]$$

$$\ell(\theta) = k\log(\theta) + (n-k)\log(1-\theta)$$

$\log(x^a) = a\log(x)$

$\log\left(\frac{x}{y}\right) = \log(x) - \log(y)$

$\log(x \cdot y) = \log(x) + \log(y)$

$\log(x^2) = \log(x \cdot x)$

$= \log(x) + \log(x)$

$= 2\log(x)$

Paso 3 $\left.\dfrac{\partial l(\theta)}{\partial \theta}\right|_{n,k} = 0$

$$\left.\frac{\partial l(\theta)}{\partial \theta}\right|_{n,k} = \frac{\partial}{\partial \theta}\left[\underbrace{k \log(\theta)} + (n-k)\log(1-\theta)\right]$$

$$= k\frac{\partial}{\partial \theta}\log(\theta) + \boxed{(n-k)}\frac{\partial}{\partial \theta}\log(1-\theta)$$

$$= \frac{k}{\theta} + \frac{(n-k)}{1-\theta}(-1)$$

$$\boxed{= \frac{k}{\theta} - \boxed{\frac{(n-k)}{1-\theta}} = 0}$$

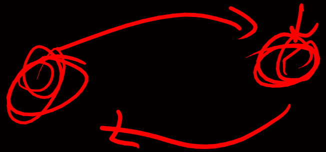
$\dfrac{\partial}{\partial x}$

$\dfrac{1}{u}\dfrac{du}{\;} =$

$\dfrac{d(1-\theta)}{d\theta} = -1$

Paso 4 Resolver para θ

$$\frac{K}{\theta} = \frac{(n-k)}{(1-\theta)} \rightarrow (1-\theta)\frac{K}{\theta} = (n-k)$$



$$K - \cancel{K\theta} = n\theta - \cancel{K\theta}$$

$$K = n\theta \Longrightarrow \boxed{\theta^* = \frac{K}{n}} \quad \text{q.e.d.}$$

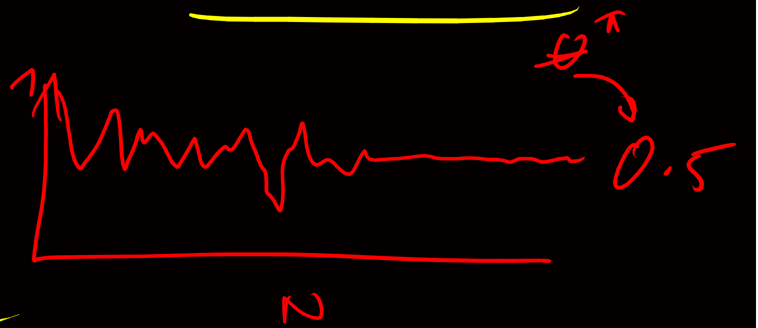
$$\boxed{\theta^* = \frac{1}{n}\sum_{i=1}^{n} x_i}$$

MLE

La media de éxitos

el MLE θ^* es una media muestral

proporción muestral



$\theta \in [0, 1]$

opt. restringida

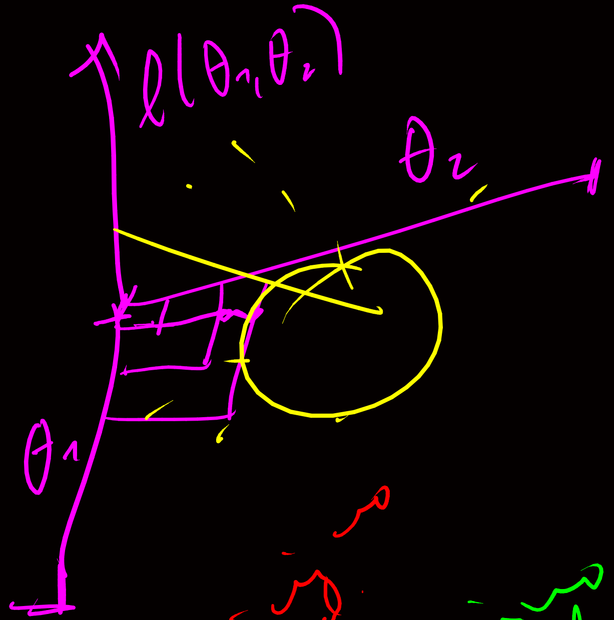
$$\theta^* = \operatorname{argmax} \mathcal{L}(\theta)$$
$$\text{s.a.} \quad \theta \in [0,1] \Big\} \text{ Restricciones}$$

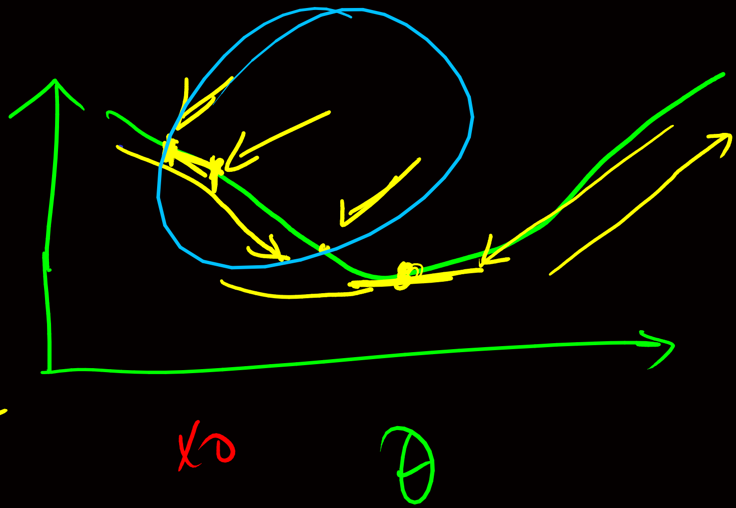
{random search}

$100 \cdot 100 \cdot 1000$ | θ_1 | θ_2 | θ_3

gradient free

no usa derivates





MLE de Normal

$x \sim N(\mu, \sigma)$

Dos parámetros

$\vec{\theta} = \{\mu, \sigma\}$

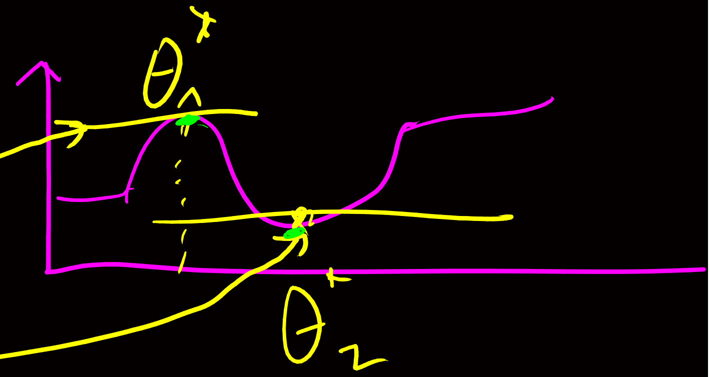
- Obtener MLE μ si σ es conocida

Si $L(\theta)$ es diferenciable en θ_k

tenemos $\boxed{\text{candidatos posibles}}$ para θ_k

$$\exists\, (\theta_1^*, \theta_2^*, \theta_3^*, \dots, \theta_p^*)$$

CPO
$$\frac{\partial L(\theta)}{\partial \theta} = 0$$



En general no se considera suficiente para mostrar el MLE

Es suficiente si sabemos en qué rango de θ, $\mathcal{L}(\theta)$ es convexa (concava)

σ es conocida

Sea $X_1, X_2, \dots, X_n$ una variable i.i.d. $\sim N(\mu, \sigma)$

normalmente distribuida con media μ y varianza de 1, $X_i \sim N(\mu, 1)$

y sea $L(\mu)$ la función de verosimilitud

Hallar $\mu^* = \arg\max L(\mu)$

$$L(\mu) = \boxed{\prod_{i=1}^{n} f\,(x_1, x_2, \dots, x_n)}$$

$$f(x) = \frac{1}{\sqrt{2\pi}\,(\sigma)} \cdot \exp\left(-\frac{1}{2\sigma^2}(x_i-\mu)^2\right)$$

$$f(x) = \boxed{\frac{1}{\sqrt{2\pi}} \cdot \exp\left[-\frac{1}{2}(x-\mu)\right]}$$

Si $f(x_1, x_2, \dots, x_n)$ es conjunta y x_i es $\boxed{i.}$ i.d.

$$f(x_1, x_2 \dots x_n) = \underbrace{f(x_1) \cdot f(x_2) \cdots f(x_n)}$$

$$L(\mu) = \prod_{i=1}^{n} \frac{1}{(2\pi)^{1/2}}\, \underbrace{\exp\left\{-\frac{1}{2}(x_i-\mu)^2\right\}}$$

El término exponencial:

$e^{x_1} e^{x_2} \cdots e^{x_n} = e^{\boxed{\sum_{i=1}^{n} x_i}}$

$$\mathcal{L}(\mu) = \frac{1}{(2\pi)^{n/2}} \prod_{i=1}^{n} \underbrace{e^{-\frac{1}{2}(x_i-\mu)^2}}$$

$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots x_n$

$e^{x_1 + x_2 \cdots + x_n}$

$$\mathcal{L}(\mu) = \underline{\frac{1}{(2\pi)^{1/2}} e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2}}$$

$\rightarrow \log, \ln$

$\log\left(\frac{x}{y}\right) =$

$\log(x) - \log(y)$

transformar logarítmicamente:

$$\ell(\mu) = \log_{e}^{\text{base}} \left\{ \underbrace{\frac{1}{(2\pi)^{1/2}}}_{x} \underbrace{e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2}}_{y} \right\}$$

$\log(x \cdot y) = \log(x) + \log(y)$

$$\ell(\mu) = \log\left(\underbrace{\frac{1}{(2\pi)^{1/2}}}_{\log\left(\frac{x}{y}\right)}\right) + \log\left[e^{\underline{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2}}\right]$$

$$\log(e^a) = a$$

$$= \log(\cancel{1}^{\,0}) - \log\left([2\pi]^{1/2}\right) - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\ell(\mu) = -\log\left[(2\pi)^{1/2}\right] - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2$$

CPO. $\left.\dfrac{\partial \ell(\mu)}{\partial \mu}\right|_{x^*} = 0$ y resolver para μ

$$\frac{\partial \ell(\mu)}{\partial \mu}\bigg|_{x_i} = \frac{\partial}{\partial \mu}\left\{\cancel{-\log\left[(2\pi)^{n/2}\right]} - \underbrace{\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2}\right\}$$

$$= -\frac{1}{2}\frac{\partial}{\partial \mu}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$= -\frac{1}{\cancel{2}}\left[\cancel{2}\sum_{i=1}^{n}(x_i-\mu)(-1)\right]$$

$$= \sum_{i=1}^{n}(x_i-\mu) = 0$$

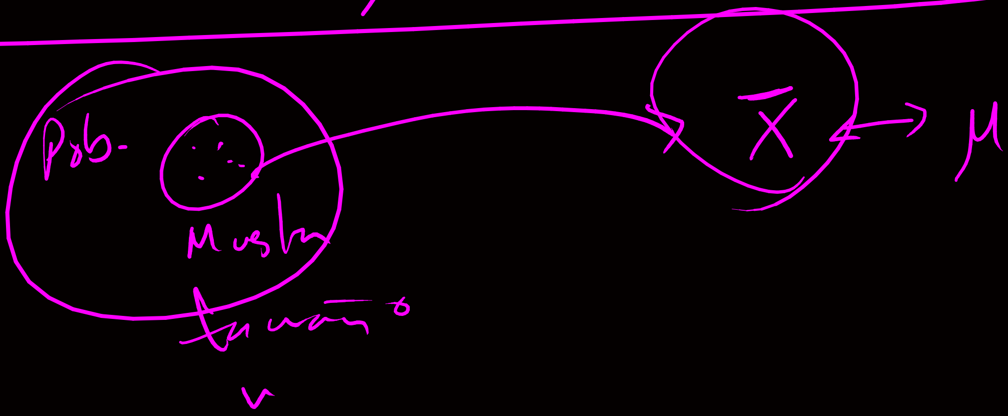
$$\sum_{i=1}^{n} k = nk$$

$$\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \mu = 0 \Rightarrow \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \mu$$

$$\sum_{i=1}^{n} x_i = n\mu \Rightarrow \mu^* = \frac{1}{n}\sum_{i=1}^{n} x_i \to \sum_{i=1}^{n} \frac{x_i}{n}$$

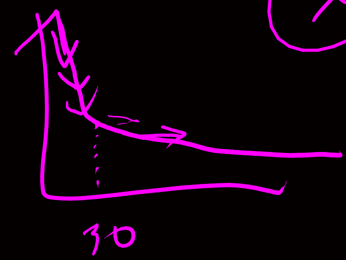
Media

* La media muestral es el MLE de μ.



$n \geq 30$ $n = 200$

$\bar{X}$



Conteo rápido

$$\bar{X} = \sum_{i=1}^{n} \frac{\text{Votos } X}{m}$$

switch

$n = 1300$

5

Encontrar MLE para μ y σ

$X_i \sim N(\mu, \sigma)$

$$f(x_i; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}$$

$$\mathcal{L}(\mu, \sigma) = f(\underbrace{x_1, x_2, \dots, x_n}_{\text{Conjunta}})$$

$X_1, X_2, \dots, X_n$ son i.i.d.

$$\mathcal{L}(\mu, \sigma) = f(x_1) \cdot f(x_2) \cdots f(n) = \prod_{i=1}^{n} \boxed{f(x_i)}$$

$$\mathcal{L}(\mu,\sigma) = \prod_{i=1}^{n} \underbrace{\frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}}_{\text{Función de densidad normal para } x_i}$$

$x_i \sim N(\mu,\sigma)$ ¿por qué no μ_i, σ_i?

$X_1, X_2, \dots, X_n$ son i.i.d

con μ y σ iguales

X_i proviene de $\underbrace{\text{una muestra}}$ de n

La función L parte de que

x fue observado, pero desconocemos

de qué parámetro proviene

$$L(\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2\sigma^2}(x_i - \mu)^2}$$

$$\longrightarrow \frac{1}{\sigma(2\pi)^{1/2}} \Rightarrow \frac{1}{\sigma(2\pi)^{n/2}}$$

1) Resolver para μ^* tomando σ como constante

2) Resolver para σ^* sustituyendo $\mu \leftarrow \mu^*$, siguiendo la misma lógica

¿Por qué tenemos que resolver primero para μ?

Resp. ¿Cuál es la una ...)

$\sigma(\mu, x)$

$$\left.\frac{\partial \mathcal{L}(\mu,\sigma)}{\partial \mu}\right|_{\sigma}$$

Simplificar $\mathcal{L}(\mu,\sigma)$, log-vero, C.P.O