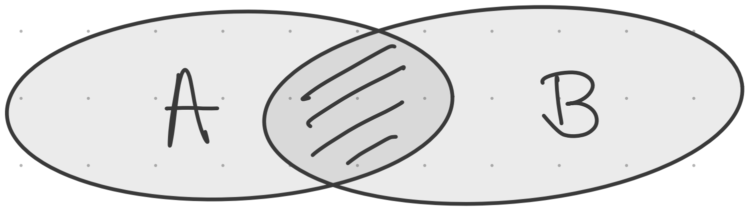
# Métodos Numéricos

——FI3104——

# Estimación Bayesiana de Parámetros



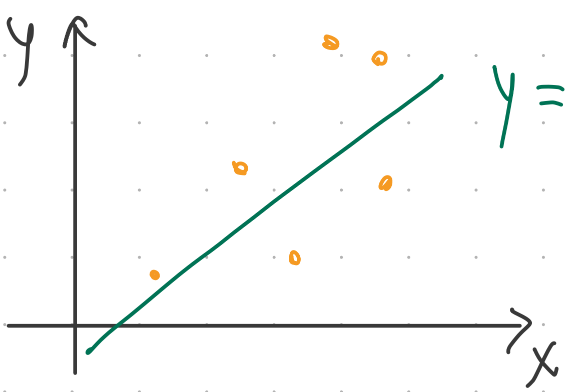

$$P(A \cap B) = P(B \cap A)$$

$$P(A|B)\,P(B) = P(B|A)\,P(A)$$

Teo. de Bayes.
$$\boxed{P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}}$$

$A \to \vec{\theta}$ : set de parámetros (ej $\theta = (m, n)$)

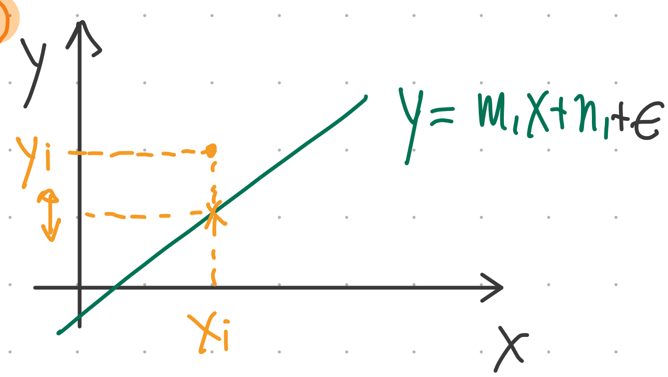
$B \to D$ : set de observaciones $\to (x_i, y_i)_{i=1\_N}$




$y = m_1 x + n_1$

¿$P((m_1, n_1) \mid (x_i, y_i)_{i=1\_N})$?

$$P(\vec{\theta}\mid D) = \frac{P(D\mid\vec{\theta})\, P(\vec{\theta})}{P(D)}$$



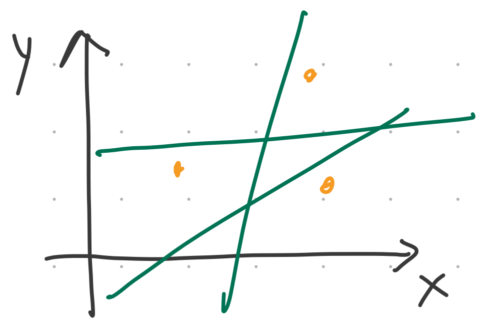

- $P(D\mid\vec{\theta})$ : verosimilitud (likelihood)

$$\epsilon_i \sim N(0, \sigma_i^2)$$

$$\mathcal{L}\left((x_i, y_i)\mid(m,n)\right) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(y_i-(mx+n))^2}{2\sigma_i^2}\right)$$

$$P(D\mid(m,n)) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(y_i-(mx+n))^2}{2\sigma_i^2}\right)$$

- $P(\vec{\theta})$ : prob. a priori (prior)




- $P(D)$ : evidencia

$$P(D) = \iint P(D\mid(m,n))\, dm\, dn$$

$$P(\theta\mid D) \propto P(D\mid\theta) \cdot P(\vec{\theta})$$

posterior　　verosimilitud　　Prior

# PRIORS FLEXIBLES USANDO GAUSSIANAS MULTIVARIADAS

En 1D: $G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

En 2D:

$$G(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(\frac{-1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right)$$

EN ND:

$$G(\vec{x}) \; ; \; \vec{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_N \end{pmatrix} \; ; \; \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12}^2 & \cdots \\ & \sigma_2^2 & \cdots \\ & & \ddots \\ & & & \sigma_N^2 \end{bmatrix}$$

$$G(\vec{x}) = \frac{1}{\sqrt{(2\pi)^N \det(\Sigma)}} \exp\left(\frac{-1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu})\right)$$

scipy.stats.multivariate_normal

Generalmente usaremos $\Sigma = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_N^2 \end{bmatrix}$

# COMPARACION BAYESIANA DE MODELOS

$$P(M|D) = \frac{P(D|M)\,P(M)}{P(D)}$$

$P(D)$ ← bajo cualquier modelo posible!

Razón de prob. entre 2 modelos:

$$O_{12} = \frac{P(M_1|D)}{P(M_2|D)} \quad ; \quad O_{12} > 1 \Rightarrow M_1 \text{ tiene mayor prob.}$$

$$O_{12} = \frac{\dfrac{P(D|M_1)P(M_1)}{\cancel{P(D)}}}{\dfrac{P(D|M_2)P(M_2)}{\cancel{P(D)}}} = \overbrace{\left(\frac{P(D|M_1)}{P(D|M_2)}\right)}^{\text{Factor de Bayes}} \overbrace{\left(\frac{P(M_1)}{P(M_2)}\right)}^{\text{razón a priori generalmente} = 1}$$

$$P(\theta|D,M) = \frac{P(D|\theta,M)\,P(\theta|M)}{P(D|M)}$$

evidencia → $P(D|M)$

Factor de Bayes es el cociente de las evidencias

$$P(\theta|D,M)\,P(D|M) = P(D|\theta,M)\,P(\theta|M) \Big/ \int_{\Omega} d\theta$$

$\Omega$ ← todo el espacio de param. posible

$$\int P(\theta|D,M)\,\underbrace{P(D|M)}_{\text{indep } \theta}\,d\theta = \int P(D|\theta,M)\,P(\theta|M)\,d\theta$$

$$P(D|M) \times 1 = \int \underbrace{P(D|\theta,M)}_{\text{verosimilitud}}\,\underbrace{P(\theta|M)}_{\text{prior}}\,d\theta$$

- $O_{12}$ toma en cuenta la complejidad de los modelos

Ejemplo: - $M_1$ no tiene parámetros
- $M_2$, 1 parámetro, $\lambda$ ← obviamente mejores verosimilitudes

¿Debemos preferirlo siempre?

$$O_{12} = \frac{P(D|M_1)}{P(D|M_2)} \frac{P(M_1)}{P(M_2)}$$

$$P(D|M_2) = \int_{\Omega} P(D|\lambda, M_2) \underbrace{P(\lambda|M_2)}_{\text{uniforme}[\lambda_{min}, \lambda_{max}]} d\lambda$$

$$\frac{1}{\lambda_{max} - \lambda_{min}} = \frac{1}{\Delta\lambda}$$

gaussiana

max verosimilitud

$$P(D|\lambda, M_2) = P(D|\lambda_0, M_2) \exp\left(-\frac{(\lambda - \lambda_0)^2}{2\sigma_\lambda^2}\right)$$

incertidumbres observacionales

$$P(D|M_2) = \int_{\lambda_{min}}^{\lambda_{max}} P(D|\lambda_0, M_2) \underbrace{\exp\left(-\frac{(\lambda - \lambda_0)^2}{2\sigma_\lambda^2}\right)} \cdot \frac{1}{\Delta\lambda} d\lambda$$

$$P(D|M_2) = \frac{P(D|\lambda_0, M_2)}{\Delta\lambda} \cdot \sqrt{2\pi}\,\sigma_\lambda$$

$$O_{12} = \underbrace{\left(\frac{P(M_1)}{P(M_2)}\right)}_{\text{típicamente } = 1} \underbrace{\left(\frac{P(D|M_1)}{P(D|\lambda_0, M_2)}\right)}_{\text{compara los mejores ajustes}} \left(\frac{\Delta\lambda}{\sqrt{2\pi}\,\sigma_\lambda}\right)$$

Penaliza $M_2$ si el espacio de prámetros a explorar es muy grande (comparado con las incertidumbres experimentales).