

Modelos lineales y modelos lineales generalizados

Rolando Gonzales Martinez, PhD

Fellow postdoctoral Marie
Skłodowska-Curie

Universidad de Groningen
(Países Bajos)

Investigador (researcher)

Iniciativa de Pobreza y Desarrollo
Humano de la Universidad de
Oxford (UK)

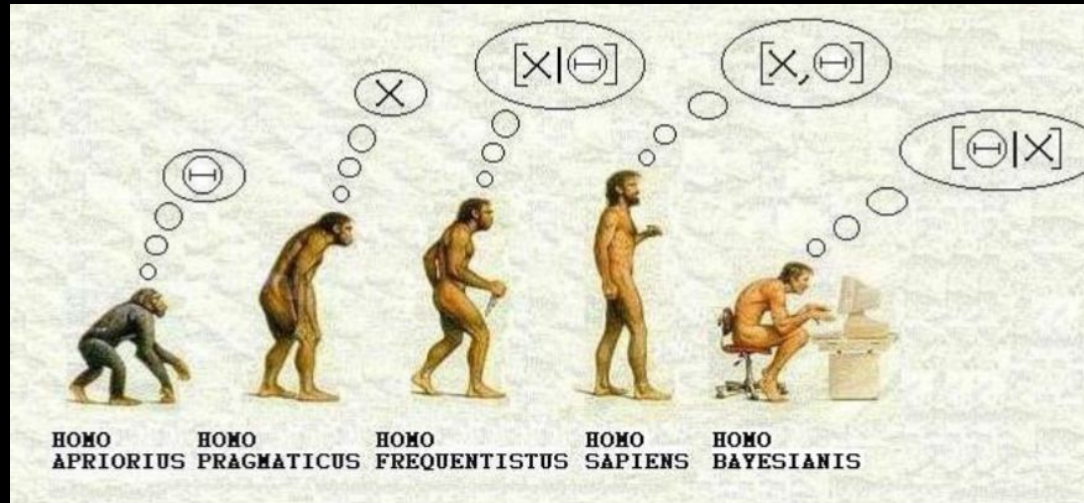
Contenido del curso

(4) Estimación Bayesiana

- Fundamentos de la inferencia Bayesiana.
- Teorema de Bayes.
- Métodos de MCMC (Markov Chain Monte Carlo) para estimación Bayesiana.
- Laboratorio: Estimación Bayesiana de modelos lineales y modelos lineales generalizados

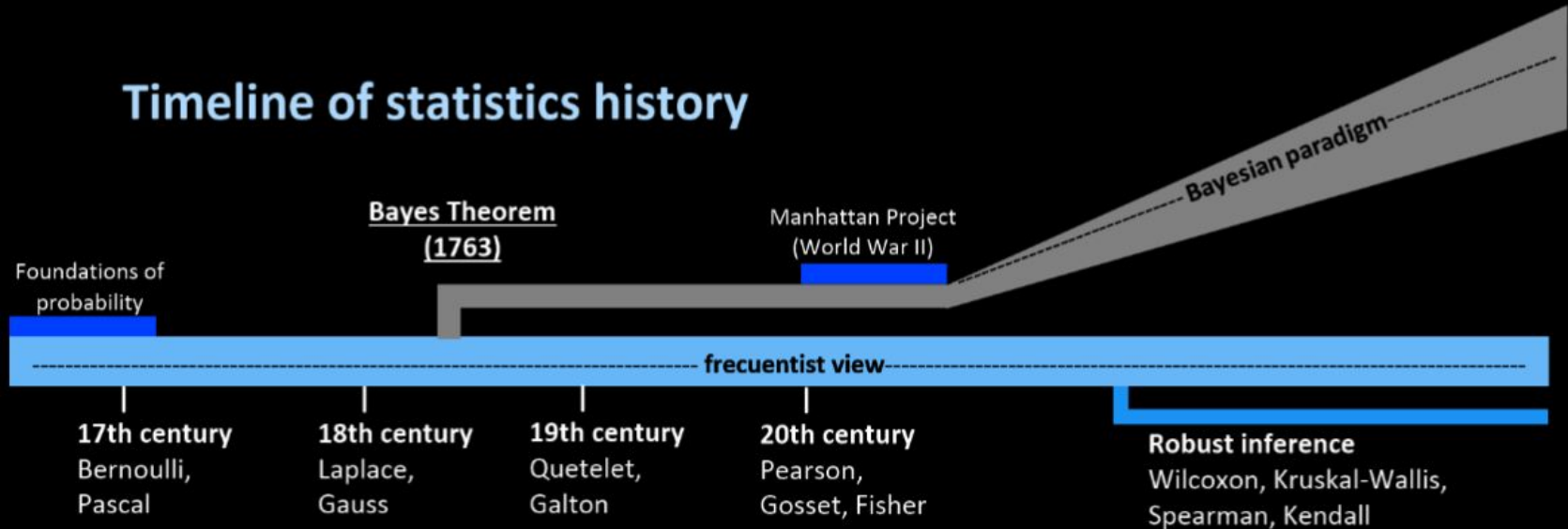
Estimación Bayesiana de modelos lineales

- La estimación Bayesiana no es simplemente un “método adicional o diferente” de estimación:
- El enfoque Bayesiano es un paradigma estadístico diferente.
- Epistemológicamente, un paradigma científico diferente.



Enfoque Bayesiano

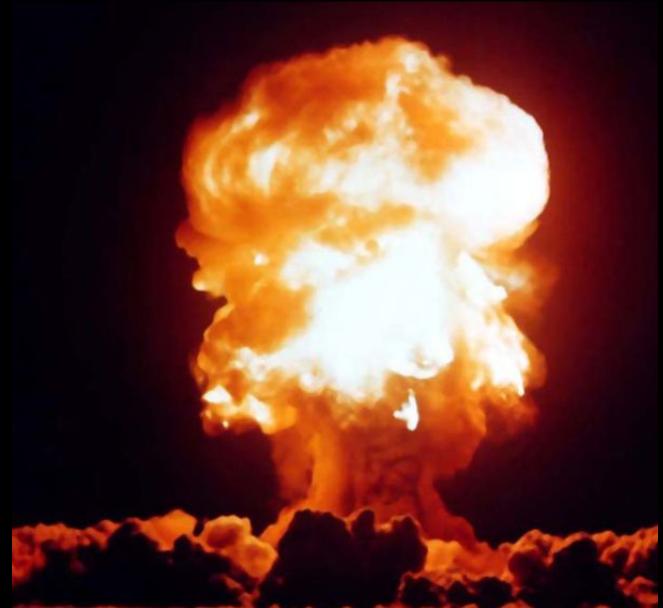
Antes de la segunda guerra mundial se utilizaban conjugados naturales



Enfoque Bayesiano

Antes de la segunda guerra mundial se utilizaban conjugados naturales

- Métodos de Monte Carlo
—desarrollados durante el El Proyecto Manhattan— permitieron aproximar la integrales multidimensionales del análisis Bayesiano.
- El crecimiento exponencial del software y hardware computacional han hecho el uso de los métodos de integración de Monte Carlo más accesible.



Enfoque Bayesiano

La regla de Bayes surge de los axiomas de probabilidad y no es una materia de controversia.

La división trata sobre la interpretación filosófica de probabilidad $P(A)$:

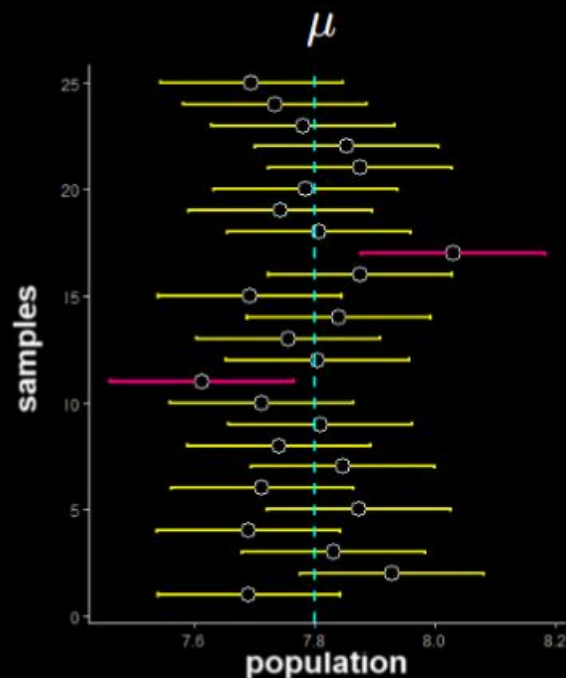
- Para un frecuentista, $P(A)$ es una frecuencia de largo plazo:

$$\mathbb{P}(A) := \lim_{n \rightarrow \infty} \frac{n_A}{n}$$

- Para un Bayesiano, $P(A)$ es cualquier conocimiento/información sobre el evento A , además del contenido en los datos, incluyendo la incertidumbre sobre A .

Enfoque frecuentista:

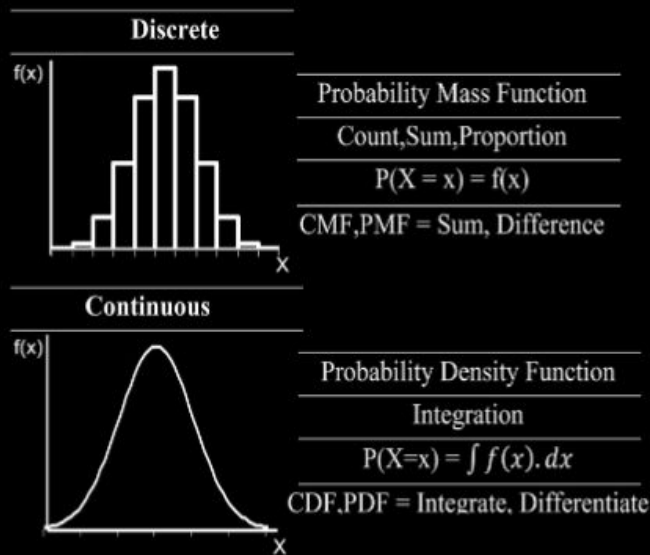
- Los datos son aleatorios
- Los parámetros son (puntos) fijos



Enfoque Bayesiano:

- Los datos son fijos
- Los parámetros son aleatorios

$$\mu \sim \mathcal{D}(\theta_1, \theta_2, \dots, \theta_d), \text{ e.g. } \mu \sim \mathcal{N}(\theta_\mu, \sigma_\mu^2), \theta_\mu \sim \mathcal{E}(\lambda_{\theta_\mu})$$



Función de verosimilitud y teorema de Bayes

$$L(\boldsymbol{\theta}|\mathbf{X}) = \prod_{i=1}^n p(\mathbf{X}_i|\boldsymbol{\theta})$$

$$\ell(\boldsymbol{\theta}|\mathbf{X}) = \log(L(\boldsymbol{\theta}|\mathbf{X}))$$

$$\dot{\ell}(\boldsymbol{\theta}|\mathbf{X}) = \frac{\partial}{\partial \boldsymbol{\theta}} \ell(\boldsymbol{\theta}|\mathbf{X})$$

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\mathcal{P}} \mathcal{N}(\mathbf{0}, \Sigma_{\boldsymbol{\theta}})$$

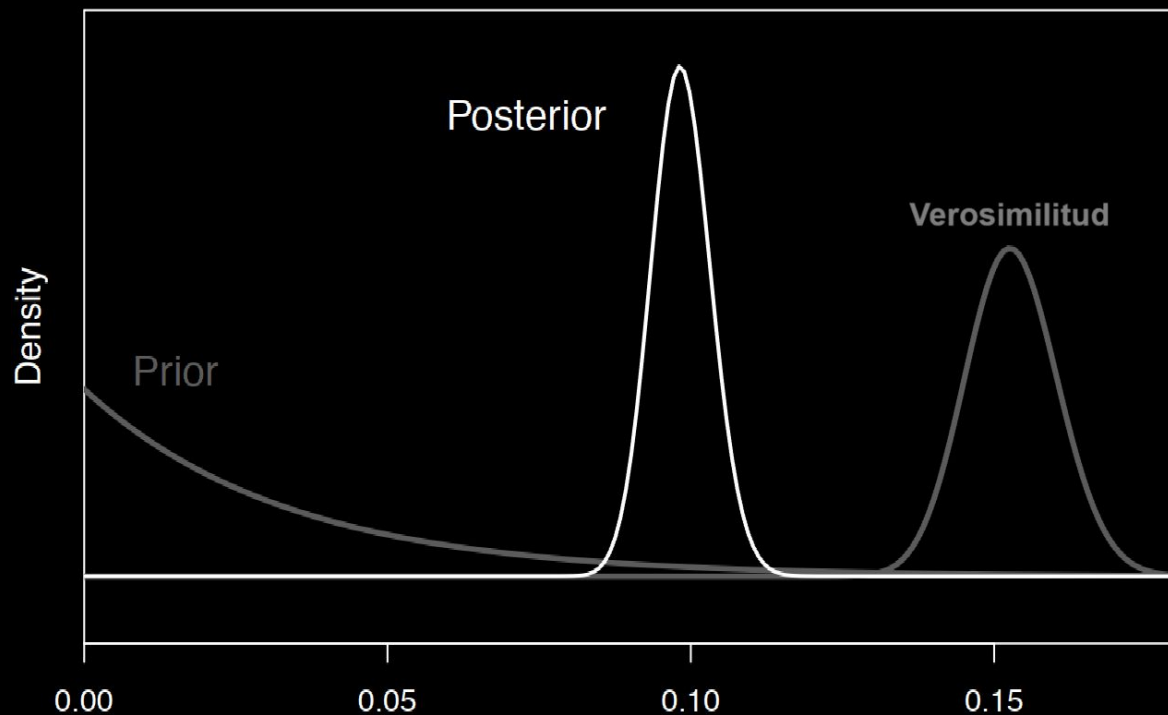
$$p(\boldsymbol{\theta}|\mathbf{X}) = p(\mathbf{X}|\boldsymbol{\theta}) \frac{p(\boldsymbol{\theta})}{p(\mathbf{X})}$$

$$\pi(\boldsymbol{\theta}|\mathbf{X}) = \frac{p(\boldsymbol{\theta})L(\boldsymbol{\theta}|\mathbf{X})}{p(\mathbf{X})}$$

$$\pi(\boldsymbol{\theta}|\mathbf{X}) \propto p(\boldsymbol{\theta})L(\boldsymbol{\theta}|\mathbf{X})$$

Probabilidad posterior \propto Probabilidad prior \times función de verosimilitud

Función de verosimilitud y teorema de Bayes



Probabilidad posterior \propto Probabilidad prior \times función de verosimilitud

Equivalencia asintótica entre los estimadores puntuales Bayesianos y los estimadores máximo verosímiles

Los estimadores bayesianos son asintóticamente equivalente a estimadores máximo verosímiles, si se emplean priors difusos (con varianza muy grande) o uniformes (no informativos)

$$\sqrt{n}(\tilde{\theta} - \hat{\theta}) \xrightarrow{n \rightarrow \infty} 0$$

$$M(\theta) = \operatorname{argmax}_{\theta} \pi(\theta | \mathbf{D})$$

Sintetizando la distribución posterior

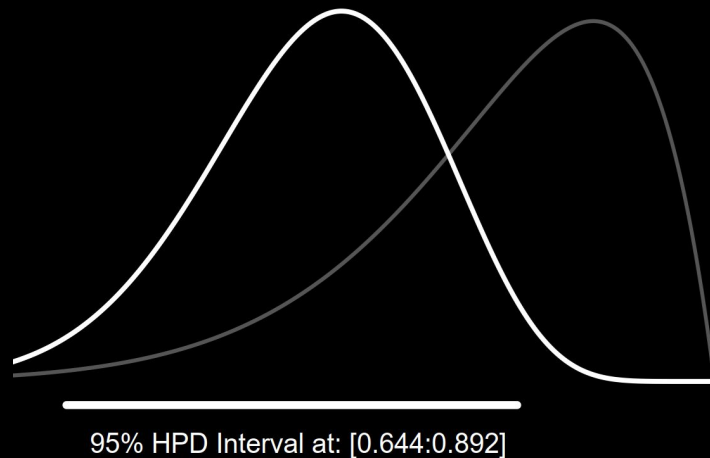
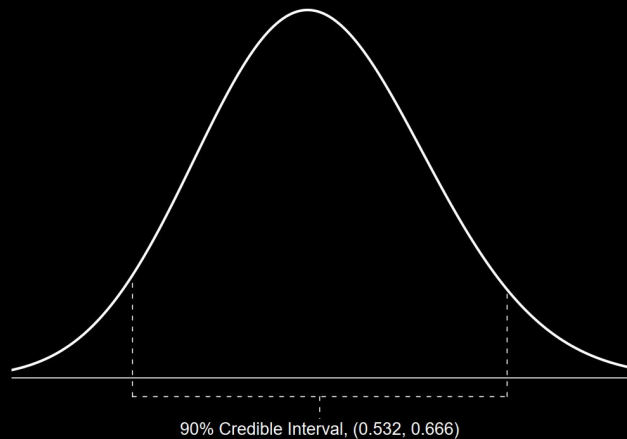
La media de la distribución posterior, intervalos de credibilidad y regiones de mayor densidad posterior (highest posterior density regions, HPD):

$$E[\theta|\mathbf{X}] = \int_{-\infty}^{\infty} \theta p(\theta|\mathbf{X}) d\theta$$

$$1 - \alpha = \int_C p(\theta|\mathbf{X}) d\theta$$

$$1 - \alpha = \int_{\theta: \pi(\theta|\mathbf{x}) > k} \pi(\theta|\mathbf{x}) d\theta$$

$$C = \{\theta : \pi(\theta|\mathbf{x}) \geq k\}$$



Beta(15,2) prior in grey

Ejemplo para una distribución exponencial

$$p(X|\theta) = \theta e^{-\theta X}$$

$$p(\theta) = 1/\theta$$

$$\pi(\theta|\mathbf{X}) \propto p(\theta)L(\theta|\mathbf{X}) = \left(\frac{1}{\theta}\right) \theta^n \exp \left[-\theta \sum_{i=1}^n x_i \right]$$

$$= \theta^{n-1} \exp \left[-\theta \sum_{i=1}^n x_i \right]$$

$$\pi(\theta|\mathbf{X}) = \frac{(\sum x_i)^n}{\Gamma(n)} \theta^{n-1} \exp \left[-\theta \sum x_i \right]$$

$$\frac{\alpha}{2} = \int_0^L \pi(\theta|\mathbf{X}) d\theta$$

$$\frac{\alpha}{2} = \int_H^\infty \pi(\theta|\mathbf{X}) d\theta$$

Ejemplo de estimación conjugada Beta-Binomial

$$\theta \sim \text{Beta}(\alpha, \beta) \quad x \mid \theta \sim \text{Binomial}(n, \theta)$$

$$p(\theta \mid \sum_{i=1}^n x_i) \propto p(\sum_{i=1}^n x_i \mid \theta) \cdot p(\theta)$$

$$p(\theta \mid \sum_{i=1}^n x_i) \propto \theta^{\sum_{i=1}^n x_i + \alpha - 1} (1 - \theta)^{n - \sum_{i=1}^n x_i + \beta - 1}$$

$$\theta \mid \sum_{i=1}^n x_i \sim \text{Beta}(\alpha + \sum_{i=1}^n x_i, \beta + n - \sum_{i=1}^n x_i)$$

Estimación Bayesiana del MLRM

Con priors no informativos los estimadores Bayesianos coinciden con los estimadores máximo verosímiles:

$$\mathbb{P}(\beta) \propto c \text{ y } \mathbb{P}(\sigma^2) \propto \sigma^{-1}$$

$$\mathbb{P}(\beta, \sigma^2) \propto \mathcal{L}(\beta, \sigma^2 | \mathbf{X}, \mathbf{y}) \mathbb{P}(\beta) \mathbb{P}(\sigma^2)$$

$$\propto \sigma^{-n-1} \exp \left[-\frac{1}{2\sigma^2} (\hat{\sigma}^2(n-k) + (\beta - \hat{\beta})' \mathbf{X}' \mathbf{X} (\beta - \hat{\beta})) \right]$$

Estimación Bayesiana del MLRM

Con priors informativos:

$$\beta \sim \mathcal{N}_k(\beta_0, B_0), \quad \sigma^2 \sim \mathcal{IG}(\alpha_0/2, \delta_0/2),$$

$$\beta|\sigma^2, y \sim \mathcal{N}(\bar{\beta}, B_1), \quad \sigma^2|\beta, y \sim \mathcal{IG}(\alpha_1/2, \delta_1/2)$$

$$B_1 = [s^{-2}X'X + B_0^{-1}]^{-1},$$

$$\bar{\beta} = B_1[\sigma^{-2}X'y + B_0^{-1}\beta_0],$$

$$\alpha_1 = \alpha_0 + n,$$

$$\delta_1 = \delta_0 + (y - X\beta)'(y - X\beta)$$

Estimación Bayesiana de modelos lineales y modelos lineales no generalizados

- Basada en el teorema de Bayes
- Estimación con conjugados naturales
- Estimación con MCMC (Markov Chain Monte Carlo: Monte Carlo con Cadenas de Markov)
- Estimación MCMCMC (MC3)

Ejemplo

Received: 16 October 2018 | Revised: 26 March 2019 | Accepted: 13 May 2019
DOI: 10.1111/ode.12807

SPECIAL ISSUE ARTICLE

WILEY

The interaction effect of gender and ethnicity in loan approval: A Bayesian estimation with data from a laboratory field experiment

Rolando Gonzales Martinez^{1,2} | Gabriela Aguilera-Lizarazu² |
Andrea Rojas-Hosse² | Patricia Aranda Blanco²

