Maestría en Ciencia y Análisis de Datos- Universidad Mayor de San Andrés

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.

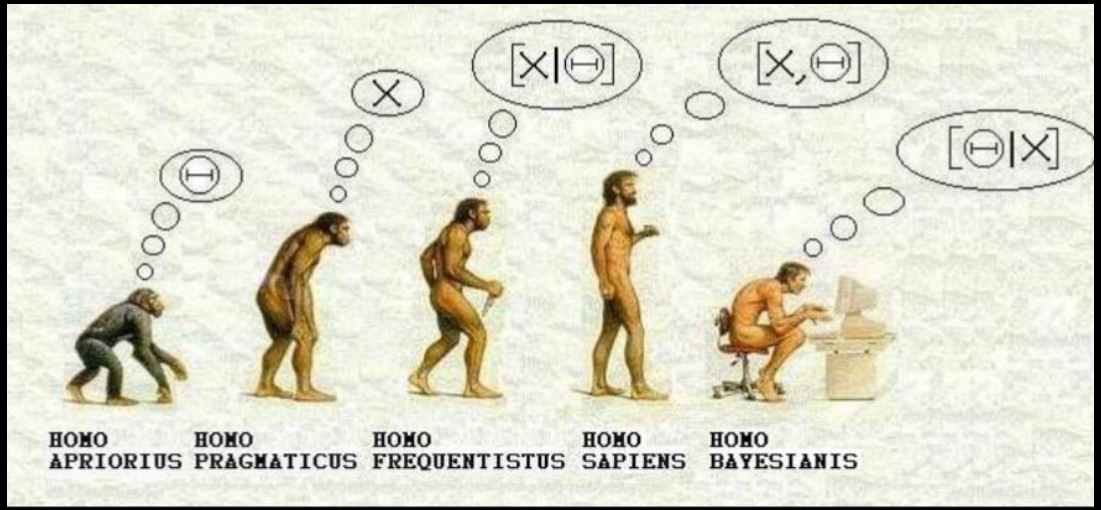
- Teorema de Bayes.

- Métodos de MCMC (Markov Chain Monte Carlo) para estimación Bayesiana.

- Laboratorio: Estimación Bayesiana de modelos lineales y modelos lineales generalizados

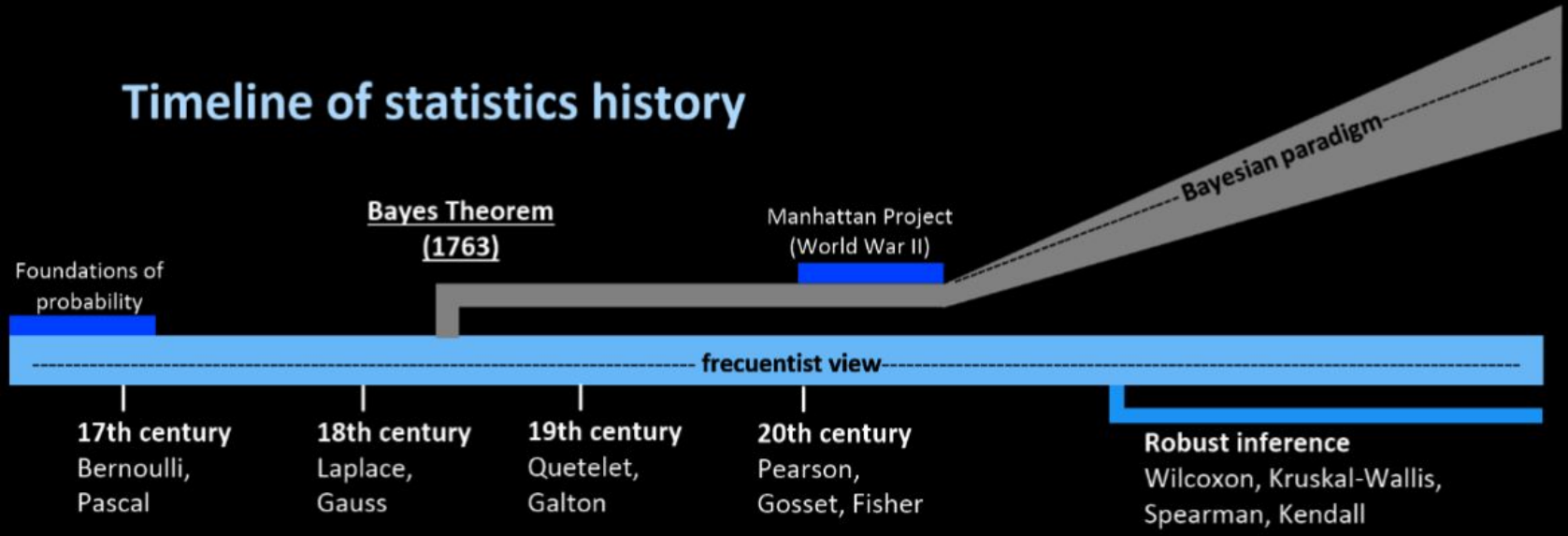
Estimación Bayesiana de modelos lineales

- 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

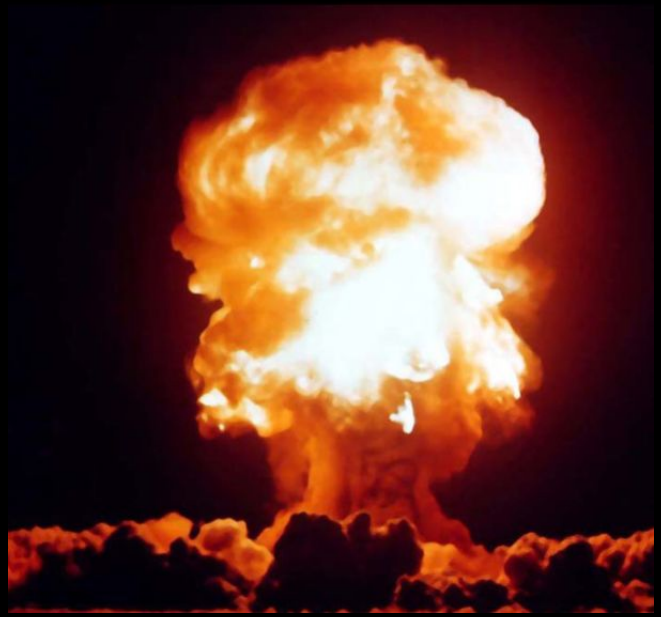
Antes de la segunda guerra mundial se utilizaban conjugados naturales



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):

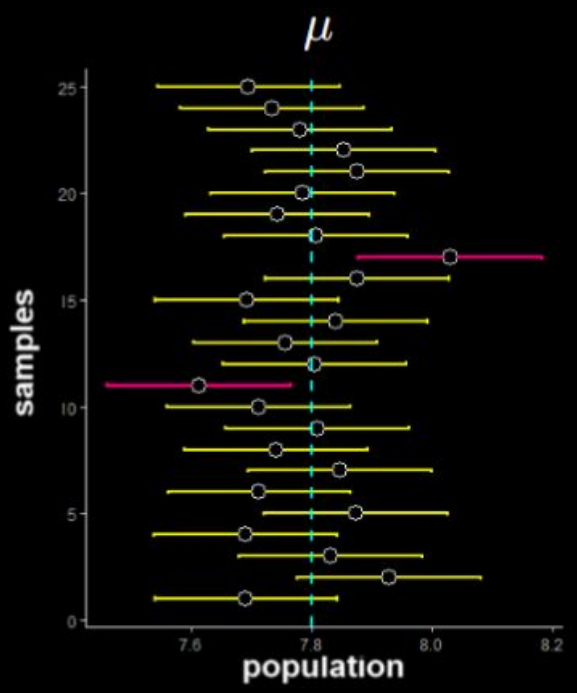
- Para un frecuentista, P(A) es una frecuencia de largo plazo:

$$\mathbb{P}(A) := \lim_{n \to \infty_+} \frac{n_A}{n}$$

- 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:

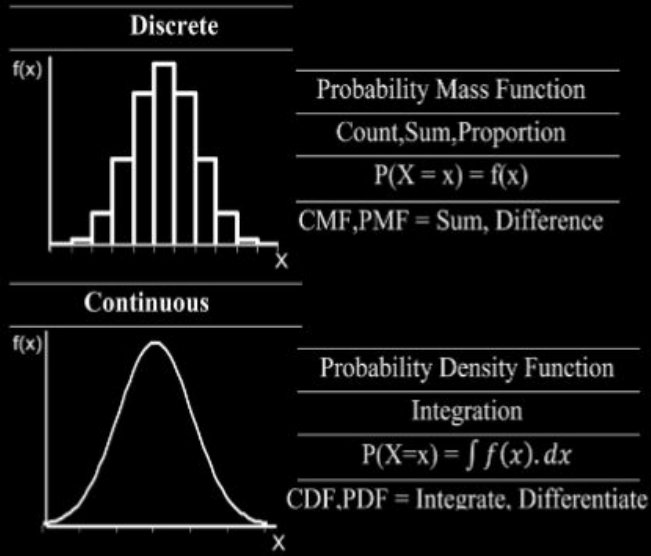
- Los datos son aleatorios
- Los parámetros son (puntos) fijos



Enfoque Bayesiano:

- Los datos son fijos
- Los parámetros son aleatorios

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



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)

Función de verosimilitud y teorema de Bayes

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

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

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

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

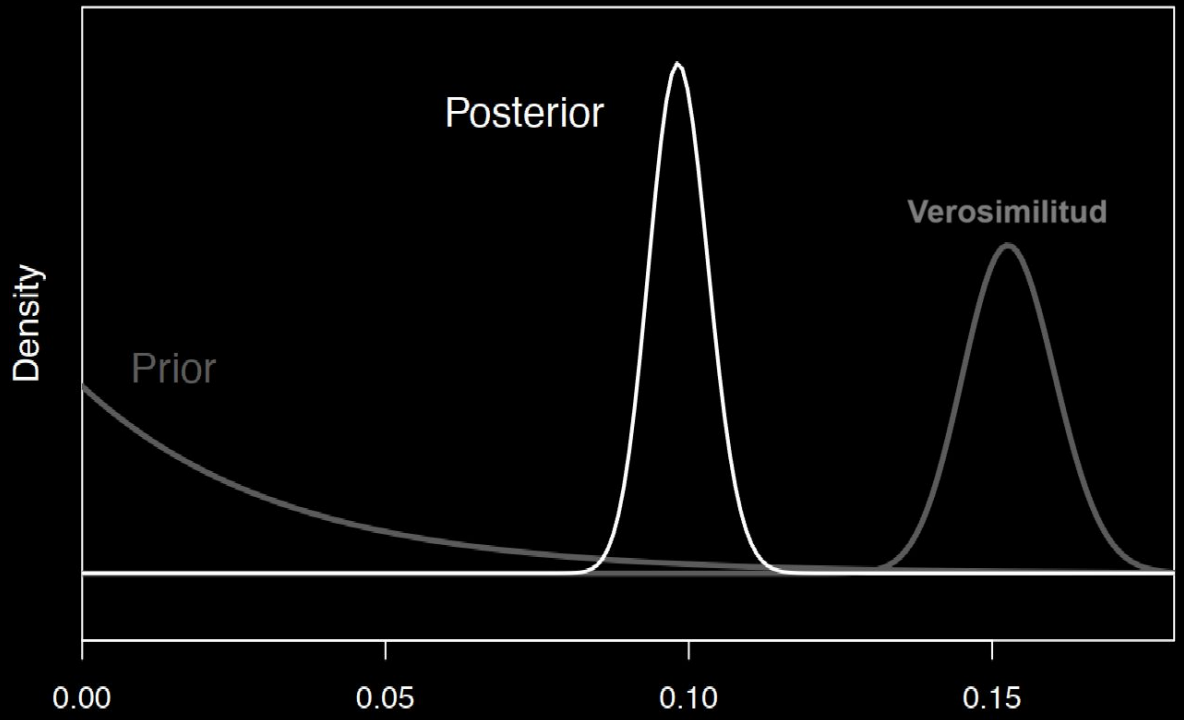
$$p(\boldsymbol{\theta}|\mathbf{X}) = p(\mathbf{X}|\boldsymbol{\theta}) \frac{p(\boldsymbol{\theta})}{p(\mathbf{X})}$$

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

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

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 varianza) o uniformes (no informativos)

$$\sqrt{n}(\tilde{\theta} - \hat{\theta}) \underset{n \to \infty}{\longrightarrow} 0$$

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

Sintetizando la distribución posterior

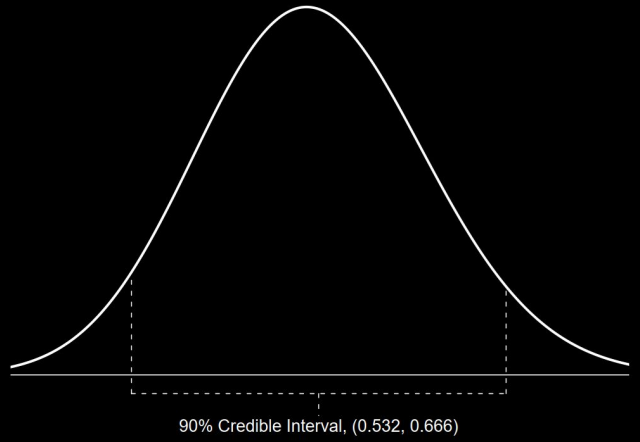
La media de la distribución posterior, intervalos de credibilidad y regiones de mayor densidad posterior (highest posterior density regions, HPD):

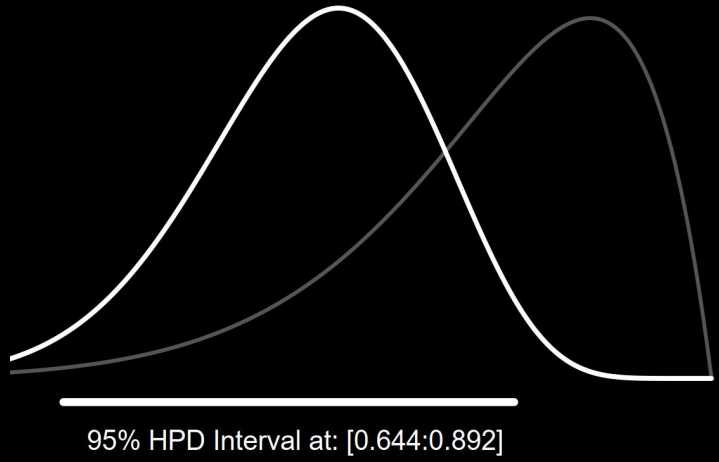
$$E[\theta|\mathbf{X}] = \int_{-\infty}^{\infty} \theta p(\theta|\mathbf{X}) d\theta$$

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

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

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





Beta(15,2) prior in grey

Ejemplo para una distribución exponencial

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

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

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

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

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

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

Ejemplo de estimación conjugada Beta-Binomial

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

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

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

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

Estimación Bayesiana del MLRM

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

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

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

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

Estimación Bayesiana del MLRM

Con priors informativos:

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

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

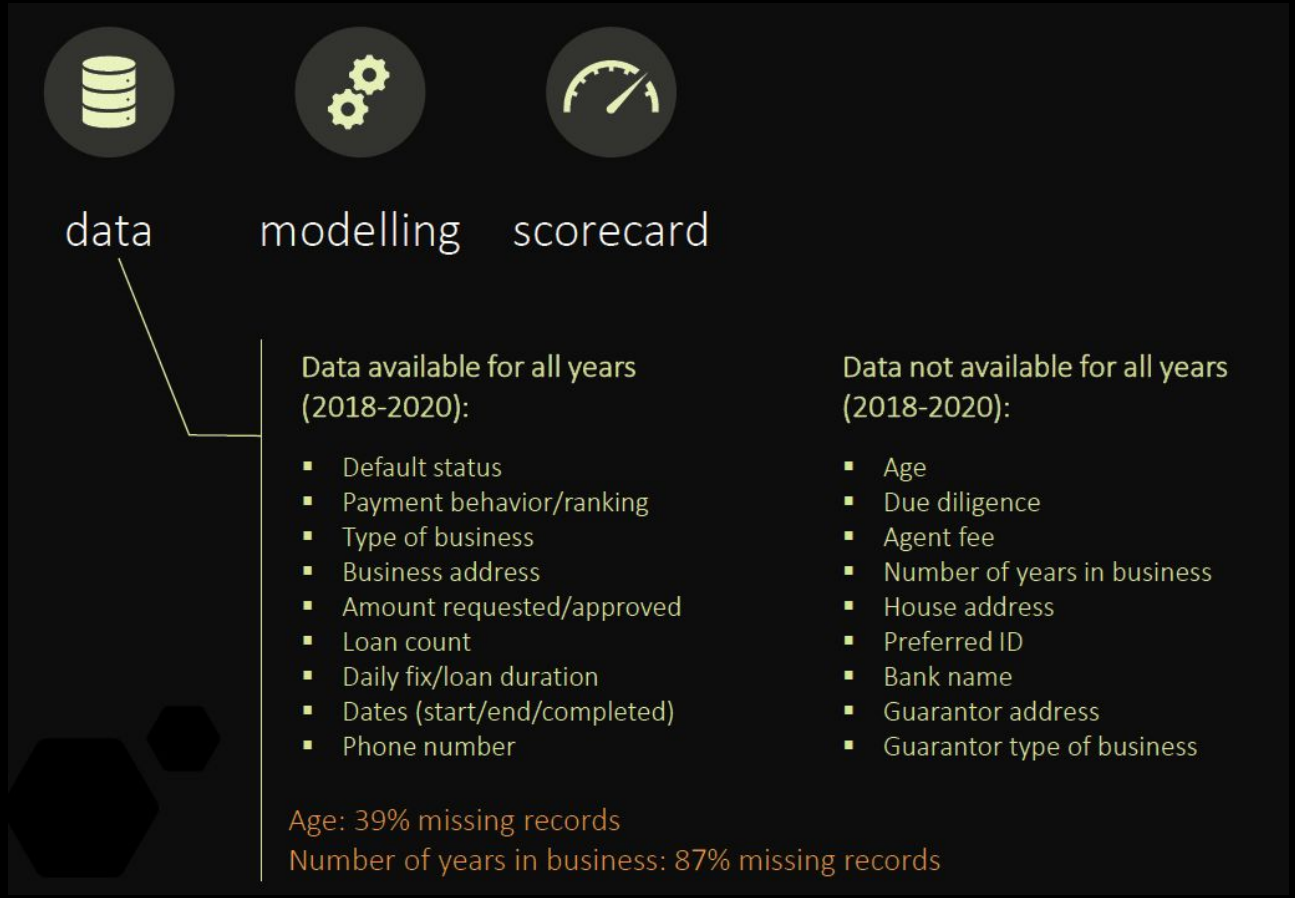
$$B_1 = [s^{-2}X'X + B_0^{-1}]^{-1},$$

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

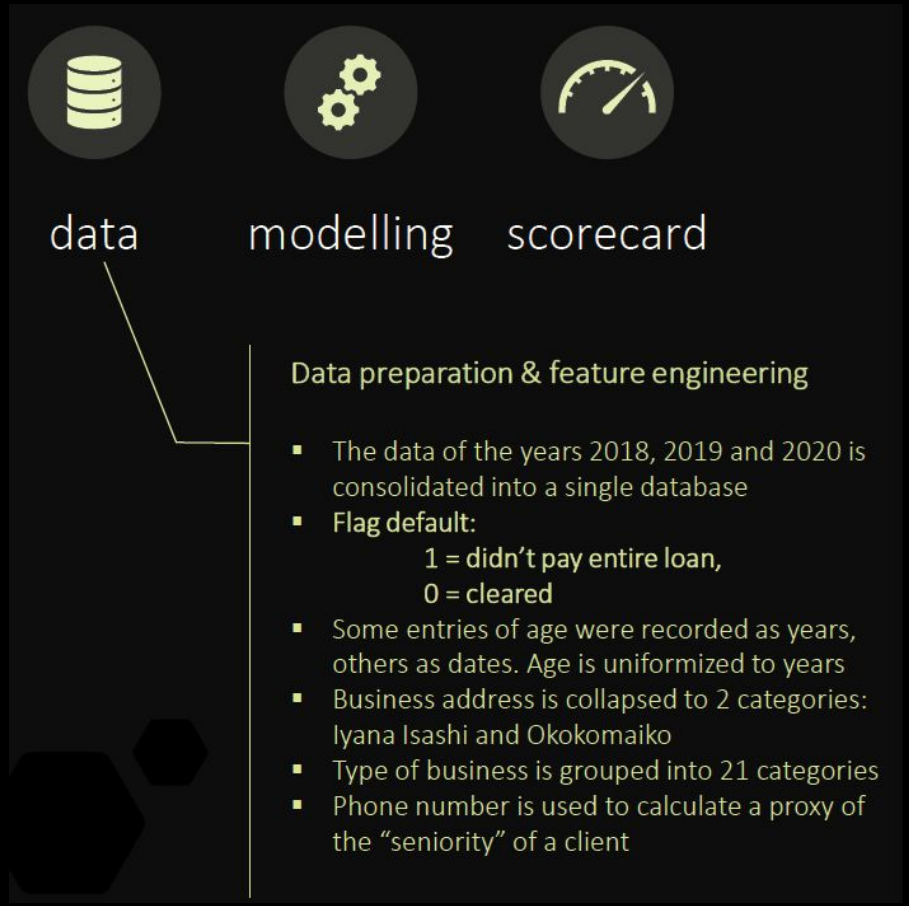
$$\alpha_1 = \alpha_0 + n,$$

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

Modelo logistico aplicado a credit scoring



Modelo logistico aplicado a credit scoring

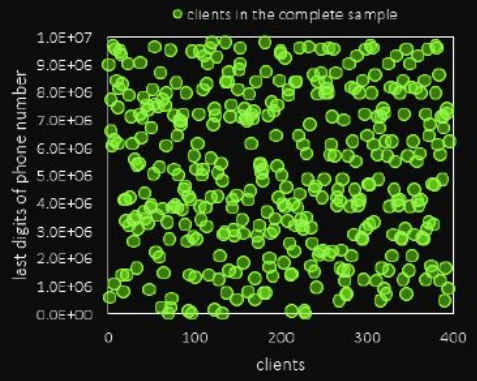


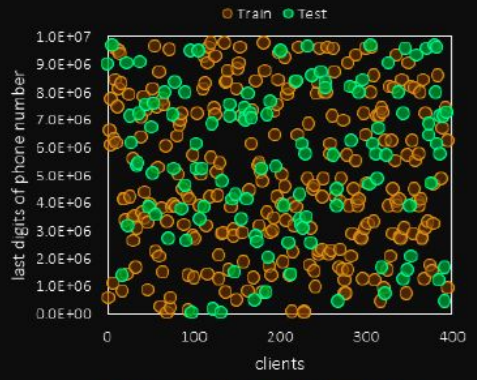
Modelo logistico aplicado a credit scoring

data modelling scorecard

Stratified sampling is used to split the data in train and test samples

- Tentative models are estimated in the train sample
- The best models of the train sample are evaluated in the test sample



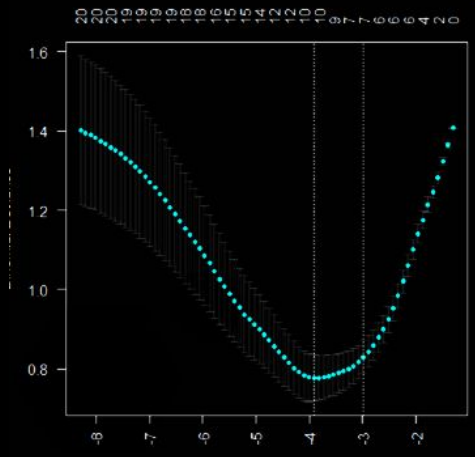


Train sample: 70% of records

Test sample: 30% of records

Modelo logistico aplicado a credit scoring

data modelling scorecard



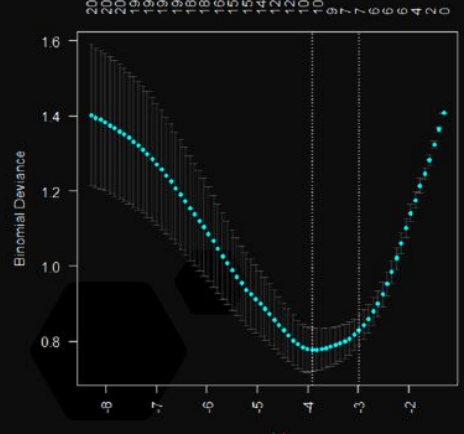
Feature selection in the train sample is performed with AUC and machine-learning

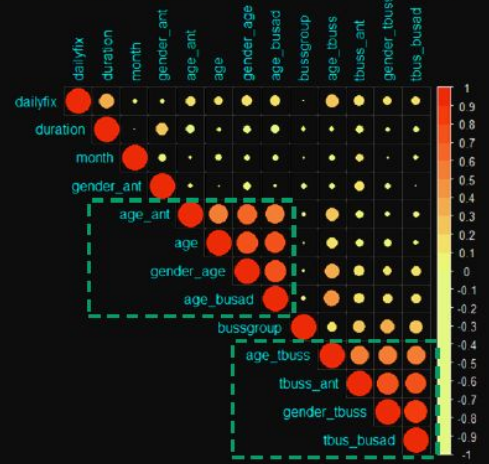
Variable	Description	AUC rank	Tree-based (Random Forests) Boruta	XtraTrees	XGBoost	Elastic nets Ridge	Elastic	Lasso	Average
tbuss_ant	Business group x seniority	1	2	2	2	20	10	10	3.86
tbus_busad	Business group x business address	5	1	1	1	10	1	1	3.93
age_tbuss	Age x business group	2	3	5	7	19	9	9	4.86
age	Age	3	4	17	4	14	6	4	5.21
bussgroup	Business group	4	9	4	8	16	7	7	5.93
gender_age	Gender x age	6	7	9	15	9	2	2	6.57
duration	Loan duration	7	11	6	10	17	8	8	8.29
age_busad	Age x business address	8	5	12	9	12	3	11	8.29
dailyfix	Daily fix	9	10	8	5	15	5	5	8.57
gender_tbuss	Gender x business group	10	6	7	3	4	11	11	8.71
age_ant	Age x seniority	11	8	10	14	3	11	11	10.36
gender_ant	Gender x seniority	12	18	15	6	8	11	11	11.79
month	Month	13	16	18	11	13	4	3	12.07
amountapp	Loan amount	14	13	14	13	2	11	11	12.57
antproxph	Seniority	15	14	13	16	7	11	11	13.71
gender	Gender	17	20	11	18	1	11	6	14.50
gender_busad	Gender x business address	16	19	19	18	11	11	11	15.50
loancount	Loan count	20	17	3	12	5	11	11	15.64
busad_ant	Business address x seniority	19	15	16	17	6	11	11	16.29
bussaddress	Business address	18	12	20	18	18	11	11	16.71

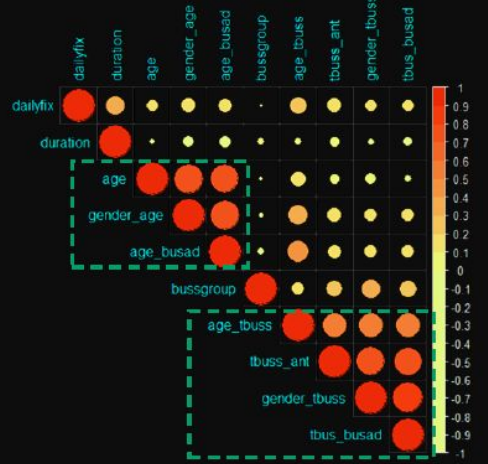
Modelo logistico aplicado a credit scoring

data modelling scorecard

Candidate models in the train sample are estimated with the (quasi-)orthogonal sub-set of best variables

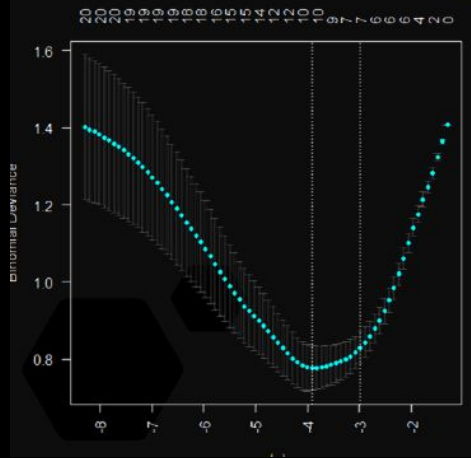




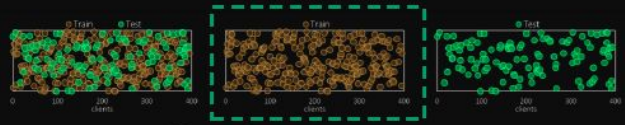


Modelo logistico aplicado a credit scoring

data modelling scorecard



Candidate models in the train sample



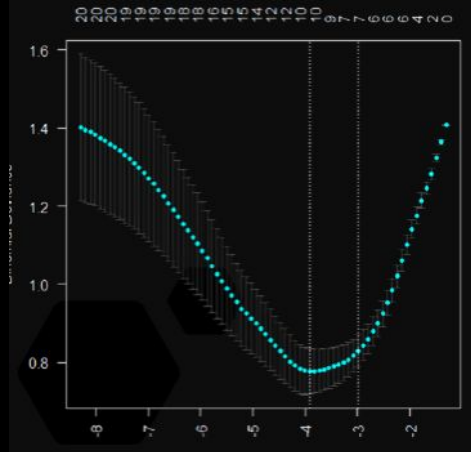
Group	Description	Variable weight M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	Average importance
Main covariates	Age	44%	28%	35%	33%	25%	27%	17%	37%	23%	25%	12%
	Business group	41%	30%	22%	27%	26%	45%	40%	29%	23%	21%	12%
	Seniority	15%			12%	15%	18%	21%	12%	14%	10%	6%
	Loan duration		21%									8%
	Daily fix		21%									8%
Interaction terms	Business group x seniority			43%							22%	13%
	Business group x business address				28%					13%		8%
	Age x business group					33%				26%	23%	11%
	Gender x age						10%					4%
	Age x business address							22%				9%
	Gender x business group								23%			9%
	Discrimination power	59.7	73.5	75.7	67.8	76.9	61.0	61.8	63.6	79.1	80.8	
	Covariates	3	4	3	4	4	4	4	4	5	5	

All variables are statistically significant with a significance level of less than 1%

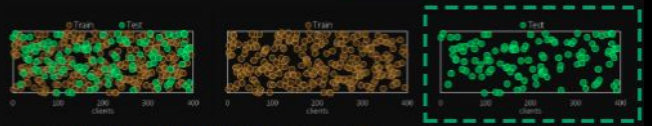
Gender was not found to be statistically significant at conventional significance levels

Modelo logistico aplicado a credit scoring

data modelling scorecard



Candidate models in the test sample



The model chosen wit the team of Ajo card were evaluated with the test sample:

- Models M5 and M9 lost their discrimination power
- The interaction term age x type of business is not statistically significant

The results suggest to choose model M10 without the interaction term of age x type of business:

Group	Description	Variable weight M5	M9	M10	Average importance
Main covariates	Age	21%	21%	14%	13%
	Business group	25%	18%	13%	13%
	Seniority	43%	42%	35%	28%
Interaction terms	Business group x seniority			38%	28%
	Business group x business address		17%		12%
	Age x business group	12%	2%		5%
	Discrimination power	52.0	53.9	64.2	
	Covariates	4	5	4	

Ejemplo

Received: 16 October 2018 | Revised: 26 March 2019 | Accepted: 13 May 2019

DOI: 10.1111/rode.12607

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 Martínez[1,2] | Gabriela Aguilera-Lizarazu[2] | Andrea Rojas-Hosse[2] | Patricia Aranda Blanco[2]