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)

Recap: Contenido

(1) Introducción a los Modelos Lineales

- Definición de modelos lineales.
- Regresión lineal simple y múltiple.
- Métodos de ajuste de modelos lineales: mínimos cuadrados ordinarios (OLS), máxima verosimilitud, métodos Bayesianos
- Laboratorio: Ajuste de modelos lineales en R y Python

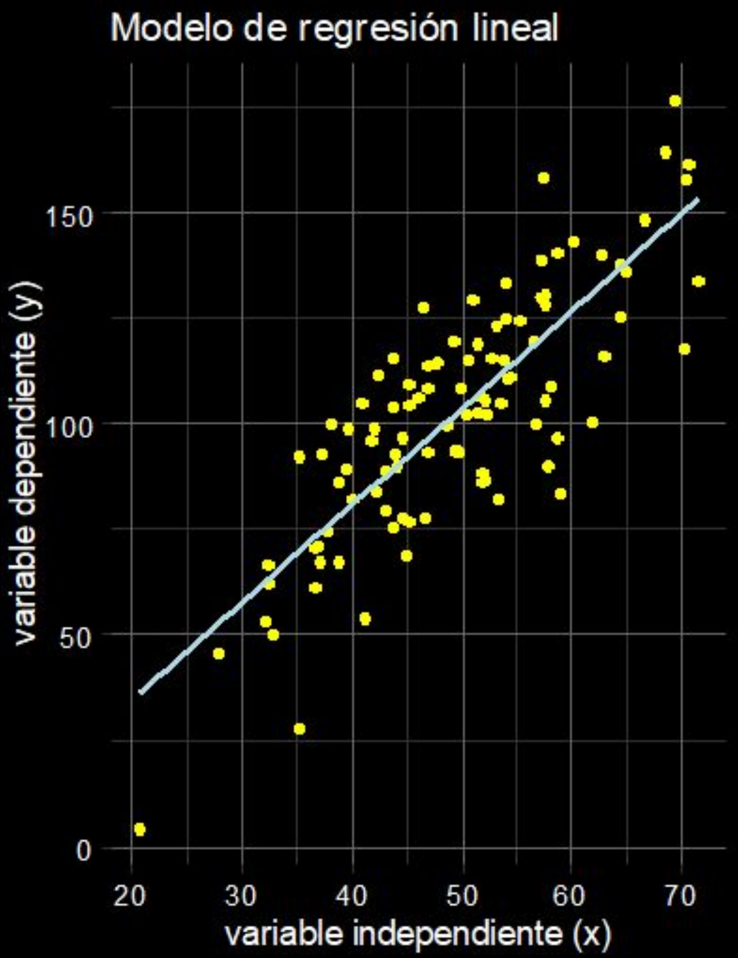
Objetivos de aprendizaje SMART de los conceptos y habilidades para las sesiones de modelos lineales

Al concluir las sesiones, se logrará:

- Conceptos:
  - Comprender qué son los modelos y qué son los modelos lineales
  - Entender los métodos de estimación de parámetros modelos lineales
  - Entender el concepto de estimador y sus propiedade
- Habilidades:
  - Estimar en la práctica modelos lineales
  - Analizar en la práctica los resultados de modelos lineales
  - Identificar en la práctica patologías en los modelos lineales

Definición de modelos lineales

- Modelos estadísticos que asumen una relación lineal entre variables.
- Los modelos lineales pueden ser simples, con una sola variable independiente, o múltiples, con varias variables independientes.
- Se utilizan en estadística y en diversas disciplinas científicas para describir y predecir relaciones entre variables.



Modelos

- Modelo Matemático: Abstracción para representar, comprender y predecir fenómenos utilizando matemáticas (ecuaciones, funciones, etc.)
    - Los modelos matemáticos son validados mediante la comparación de sus resultados/predicciones contra datos observacionales o experimentales, y pueden ser adaptados y refinados a medida que se dispone de más datos o se comprende mejor el fenómeno estudiado.
    - Epistemologicamente, los modelos son construcciones abstractas que permiten organizar y estructurar el conocimiento de manera sistemática y precisa.

Modelos

- Modelo Estadístico: Subtipo de modelo matemático que utiliza datos empíricos y probabilidades para entender, analizar relaciones y predecir fenómenos observables

Funciones Epistemológicas de un Modelo Estadístico:

- **Descriptiva:** Describe y resume las características principales de un conjunto de datos.
- **Inferencial:** Permite hacer inferencias sobre fenómenos de interés mediante estimaciones y pruebas de hipótesis.
- **Predictiva**
- **Explicativa:** Ayuda a entender las relaciones y mecanismos subyacentes entre variables.

Variables

La estadística analiza variables que fluctúan de una forma más o menos impredecible.

Variable aleatoria (definición). *Dado un espacio* $(S, \sigma_{\mathcal{B}}, \mathbb{P})$, *una variable aleatoria es una función del espacio muestral* S *a* $\mathbb{R}$,

$X : S \to \mathbb{R}$.

Ejemplo: Para $S = \{C, E\}$, es posible definir una función $X(m)$ tal que,

$$X(m) = \begin{cases} 1 \text{ si m = C} \\ 0 \text{ si m = E} \end{cases}$$

Variable

Cuando se realiza un experimento, la realización es un resultado en el espacio muestral.
Para cada evento A del espacio muestral S, puede asociarse un número entre cero y uno que se llamará probabilidad de A, $\mathbb{P}(A)$.
Para una definición más precisa, es necesario definir primero el concepto de sigma álgebras.
Sigma álgebra (definición). *Una colección de subconjuntos de S se llama sigma álgebra (σ-algebra o campo de Borel), denotado por $\mathcal{B}$, si satisface las siguientes propiedades:*

1. $\emptyset \in \mathcal{B}$
2. *Si* $A \in \mathcal{B}$, *entonces* $A^c \in \mathcal{B}$
3. *Si* $A_1, A_2, ... \in \mathcal{B}$, *entonces* $\cup_{i=1}^{\infty} A_i \in \mathcal{B}$

Variables

Función de probabilidad (definición). *Dado un espacio muestral S y una sigma álgebra asociada $\mathcal{B}$, una función de probabilidad será aquella que satisfaga,*

1. $\mathbb{P}(A) \geq 0$ *para todo* $A \in \mathcal{B}$
2. $\mathbb{P}(S) = 1$
3. *Si* $A_1, A_2, \ldots \in \mathcal{B}$ *son disjuntos, entonces* $\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$

Estas propiedades se denominan usualmente Axiomas de Probabilidad o Axiomas de Kolmogorov. La terna $(S, \sigma_{\mathcal{B}}, \mathbb{P})$ es un espacio de probabilidad en el que cada suceso $A \in \mathcal{B}$ recibe el nombre de probabilidad de A.

Variables

Las variables se dividen en varias categorías según su naturaleza y el tipo de valores que pueden tomar.

1.Variables Cualitativas (o Categóricas):

- Nominales: Son variables que representan categorías sin un orden inherente. Ejemplos incluyen el color de ojos (azul, verde, marrón) o el tipo de mascota (perro, gato, pájaro).
- Ordinales Son variables que representan categorías con un orden lógico, pero sin una distancia uniforme entre categorías. Ejemplos incluyen los niveles de satisfacción (insatisfecho, neutral, satisfecho) o las clasificaciones (primero, segundo, tercero).

2. Variables Cuantitativas (o Numéricas):

- Discretas: Son variables que toman valores contables, generalmente números enteros. Ejemplos incluyen el número de hijos en una familia o el número de coches en un garaje.
- Continuas: Son variables que pueden tomar cualquier valor dentro de un rango, incluyendo fracciones y decimales. Ejemplos incluyen la altura, el peso o la temperatura.

Variables

3. Variables Dependientes e Independientes:

- Independientes: Son variables que se manipulan o controlan para observar su efecto sobre otras variables. En un experimento, son las que el investigador cambia para ver cómo afectan a la variable dependiente.
- Dependientes: Son variables que se miden para ver el efecto de las variables independientes. En un experimento, son las que se observan y registran para ver cómo cambian en respuesta a las variables independientes.

4. Variables Dicotómicas: variables que sólo pueden tomar dos valores posibles.

5. Variables de Escala:

- Intervalo: Son variables cuantitativas donde la distancia entre dos valores tiene significado, pero no hay un verdadero cero. Ejemplo: temperatura en grados Celsius.
- Razón: Son variables cuantitativas donde hay un verdadero cero y se pueden realizar operaciones matemáticas significativas. Ejemplo: peso, altura, ingresos.

Función de distribución de una variable aleatoria

Asociada a una variable aleatoria X existe una función, denominada función de distribución acumulada de X.

Función de distribución acumulada (definición). *La función de distribución acumulada (cdf) de una variable aleatoria X, denotada por $F_X(X)$, se define como,*

$$F_X(X) = \mathbb{P}_X(X \leq x), \text{ para todo} x$$

Esta función satisface las siguientes propiedades,

1. $\lim_{x \to -\infty} F(x) = 0, \lim_{x \to \infty} F(x) = 1.$
2. $F(x)$ *es una función no decreciente de x.*
3. $\lim_{x \downarrow x_0} F(x) = F(x_0)$

Nótese que una variable aleatoria será continua si $F_X(x)$ es una función continua de x, y será discreta si $F_X(x)$ es una función en pasos de x.

Funciones de masa y densidad

Asociada a X y a su cdf F_X existe otra función, llamada función de densidad de probabilidad (pdf) o función de masa de probabilidad (pmf), refiriéndose al caso continuo y discreto, respectivamente. Ambas se refieren a probabilidades puntuales de variables aleatorias.

Función de masa de probabilidad (definición) . *La función de masa de probabilidad (pmf) $f_X(x)$ de una variable aleatoria discreta X e es*
$f_X(x) = \mathbb{P}(X = x)$ para todo x

Función de densidad de probabilidad (definición) . *La función de densidad de probabilidad (pdf) $f_X(x)$ de una variable aleatoria continua X es*
$F_X(x) = \int_{-\infty}^{x} f_X(t)dt$ para todo x *(nótese que $\frac{d}{dx}F_X(x) = f_X(x)$ y en el caso discreto, de manera similar, las probabilidades puntuales $f_X(x)$ se añaden para obtener $F_X(x)$)*

La expresión "X tiene una distribución $F_X(x)$", se escribe $X \sim F_X(x)$.

Distribuciones discretas y continuas

Las distribuciones de probabilidad pueden ser discretas o continuas. Una variable aleatoria X tiene una distribución discreta si el rango de X, el espacio muestral, es contable (en la mayoría de las situaciones, se encuentra en $\mathbb{Z}$). Por ejemplo, una variable aleatoria X tiene una *distribución uniforme discreta* si,

$$\mathbb{P}(X = x|N) = \frac{1}{N}, \quad x = 1, 2, ..., N$$

para $N \in \mathbb{Z}^+$. Esta distribución pone igual masa en todos los posibles resultados $1, 2, ..., N$.

En usual también trabajar con *variables continuas*, definidas en $\mathbb{R}$, o en $\mathbb{R}^{0,+}$ (como las variables de stock), por lo que se prestará atención a las distribuciones de variables continuas en esta sección, especialmente a las utilizadas usualmente en inferencia.

Distribuciones continuas

Distribución uniforme continua (definición). *La pdf de una distribución uniforme continúa es,*

$$f(x; a, b) = \begin{cases} \frac{1}{b-a} & \text{para } a \leq x \leq b, \\ 0 & \text{para } x < a \text{ o } x > b \end{cases}$$

por lo que si una variable aleatoria continua $X \sim \mathcal{U}(a, b)$, *su distribución queda definida por los parámetros* a *y* b. *Si* $a = 0$ *y* $b = 1$ *la distribución resultante* $\mathcal{U}(0, 1)$ *se denomina distribución uniforme estándar.*

Distribución Normal (Gaussiana)

Distribución gaussiana (definición). *Una variable aleatoria continua X se distribuye normalmente (sigue una distribución Gaussiana o Gauss-Laplace) con media μ y varianza σ^2, denotado por $X \sim \mathcal{N}(\mu, \sigma^2)$, si la pdf de X es,*

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

Algunas propiedades útiles de $f(x)$ son,

1. $f(\mu + x) = f(\mu - x)$ (simetría alrededor de μ).
2. $\mathbb{P}(\mu - \sigma < X < \mu + \sigma) \cong 0.683$
3. $\mathbb{P}(\mu - 2\sigma < X < \mu + 2\sigma) \cong 0.954$
4. $\mathbb{P}(\mu - 3\sigma < X < \mu + 3\sigma) \cong 0.997$

Distribución Normal (Gaussiana) estandar

Distribución gaussiana estándar (definición). *Una variable aleatoria continua X sigue una distribución normal estándar si $\mu = 0$ y $\sigma^2 = 1$, por lo que $X \sim \mathcal{N}(0, 1)$.*

Distribución continua t de Student

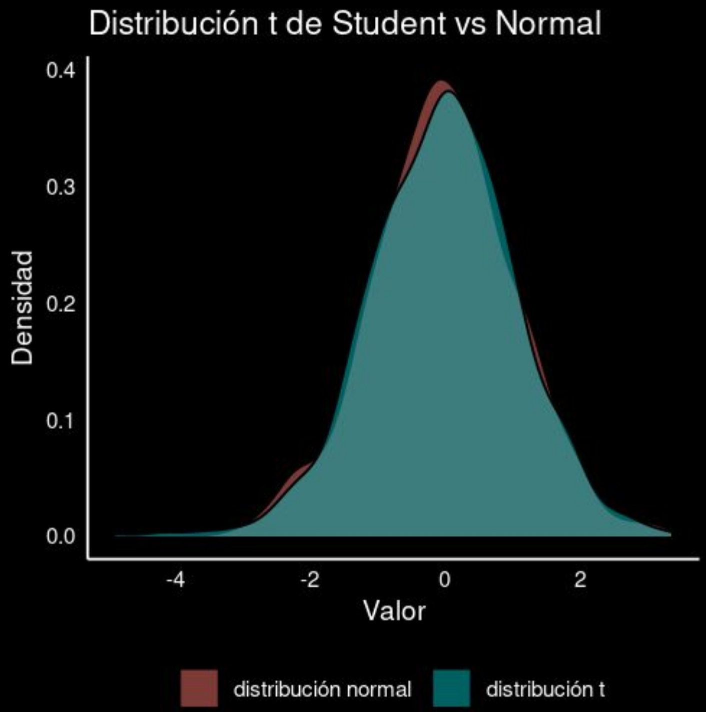
Distribución t de Student. *Sean dos variables aleatorias $Z \sim N(0,1)$ y $Y \sim \chi^2(k)$, la variable aleatoria continua T,*

$$T = \frac{Z}{\sqrt{\frac{Y}{k}}}$$

seguiría una distribución t-Student con una pdf,

$$f(T,k) = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{T^2}{k}\right)^{-\frac{1}{2}(k+1)}$$

Una propiedad importante e interesante de la distribución t de Student es que puede ser usada para muestras pequeñas que impliquen la distirbución Gaussiana.
Sea F_k la cdf de $t(k)$ y Φ la cdf de una distribución $\mathcal{N}(0,1)$. Entonces, $\lim_{k\to\infty} F_k(t) = \mathcal{N}(0,1)$.De hecho para $k > 40$ se lográ una buena aproximación.



Momentos

La esperanza de una variable aleatoria es una medida de su tendencia central, que resume el valor esperado de esta variable.

Esperanza de una variable aleatoria (definición). *La esperanza de una variable aleatoria $g(X)$, denotada por $\mathbb{E}(X)$,es*

$$\mathbb{E}(X)\begin{cases} \sum_{x\in X} g(X)f(X) = \sum_{x\in X} g(X)\mathbb{P}(X=x) \text{ si } X \text{ es discreta} \\ \int_{-\infty}^{\infty} g(X)f(X)dx \text{ si } X \text{ es continua,} \end{cases}$$

Sea una muestra $x_1, ..., x_n$. La esperanza $\mathbb{E}(X)$ está relacionada con el primer momento muestral (la media),

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

Segundo momento y momentos superiores

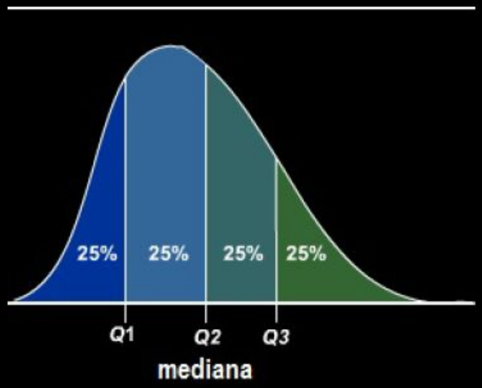
El segundo momento central es la varianza.
Varianza (definición). *La varianza de una variable aleatoria X es su segundo momento central, $Var(X) = \mathbb{E}(X - \mathbb{E}(X))^2$. La raíz cuadrada positiva de $Var(X)$ es la desviación estándar de X.*
La varianza proporciona una medida del grado de dispersión de una distribución alrededor de la media. La varianza muestral de una muestra $x_1, ..., x_n$ se calcula con,

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \overline{X}_n\right)^2$$

Momentos superiores son el sesgo y la curtosis de una distribución. El sesgo mide el grado de simetría de una distribución, y la curtosis el grado de apuntamiento de la distribución. Dependiendo del valor de la curtosis, una distribución puede ser leptocúrtica (curtosis $>$ 3), mesocúrtica (curtosis $=$ 3) o platicúrtica (curtosis $<$ 3).

Fractiles

Fractiles (definición). *Sea X una variable aleatoria continua y sea $\alpha \in (0,1)$. Si $q = q(X;\alpha)$ es aquel tal que $\mathbb{P}(X < q) = \alpha$ y $\mathbb{P}(X > q) = 1 - \alpha$, entonces q es llamado un fractil de X.*



Si se expresa las probabilidades en porcentaje, q será el 100α percentil de la distribución de X.

Estimación

La estimación es el proceso de extraer información acerca del valor de cierto parámetro poblacional a partir de la muestra $x_1, ..., x_n$, utilizando algún estadígrafo como estimador, calculado con los datos $x_1, ..., x_n$.

Estadígrafo.

Considérese que se obtiene una muestra de tamaño n de una población. Un estadígrafo es una función que se obtiene utilizando como datos las observaciones de la muestra $x_1, ..., x_n$.

Estimación puntual e interválica

Dado un modelo de regresión:

$$y = \beta_0 + \beta_1 x + \epsilon$$

Los estimadores puntuales son:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Estimación puntual e interválica

Un intervalo de confianza (1 - α)% es:

$$\beta_1 \pm t_{n-2,\alpha/2} \times SE(\beta_1)$$

Donde:

$$SE(\beta_1) = \sqrt{\frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

$$\sigma^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-2} \qquad \begin{gathered} e_i = y_i - \hat{y}_i \\ \hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \end{gathered}$$

Ejercicios prácticos

R (posit cloud):

https://posit.cloud

R analytic flow:

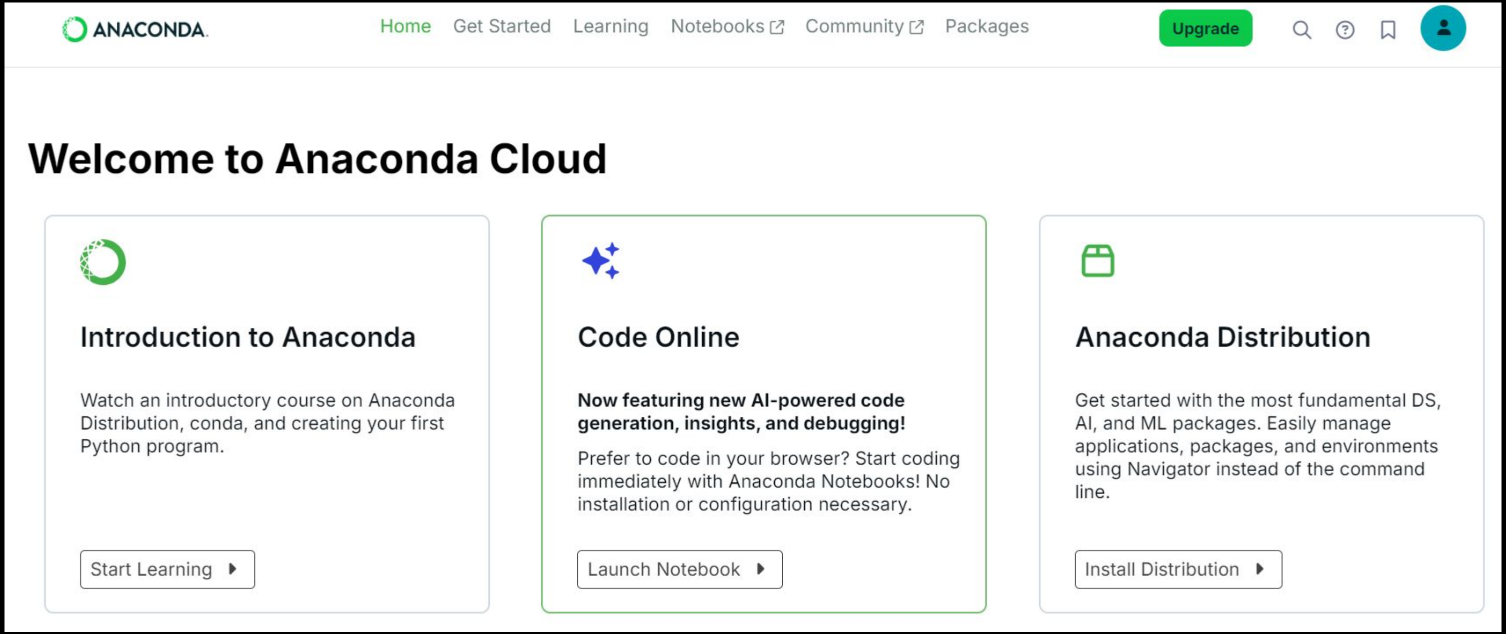
https://r.analyticflow.com/en/

Python (Anaconda cloud):

https://anaconda.cloud/sign-in

Anaconda cloud

Es necesario crear una cuenta o ingresar con un usuario de Google/GitHub y luego ingresar a Code Online:



Estimadores MCO (OLS)

Dado un modelo de regresión multivariante expresado en notación matricial:

$$\underset{(n\times 1)}{\mathbf{y}} = \underbrace{\underset{(n\times k)}{\mathbf{X}}\underset{(k\times 1)}{\beta}}_{(n\times 1)} + \underset{(n\times 1)}{\varepsilon}$$

$$\underset{(n\times 1)}{\mathbf{y}} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \underset{(n\times 1)}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \quad \underset{(n\times k)}{\mathbf{X}} = \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ x_{21} & \cdots & x_{2k} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nk} \end{bmatrix},$$

Estimadores MCO (OLS)

Se busca obtener estimadores que minimicen la suma de residuos al cuadrado

$$\underset{(n\times 1)}{\mathbf{y}} = \underbrace{\underset{(n\times k)}{\mathbf{X}}\underset{(k\times 1)}{\beta}}_{(n\times 1)} + \underset{(n\times 1)}{\varepsilon}$$

$$\underset{(k\times 1)}{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$$

$$\widehat{\beta} \equiv \arg\min_{\widehat{\beta}} SSR(\widehat{\beta})$$

$$SSR(\widehat{\beta}) = (\mathbf{y} - \mathbf{X}\widehat{\beta})'(\mathbf{y} - \mathbf{X}\widehat{\beta})$$

Estimadores MCO (OLS)

Se busca obtener estimadores que minimicen la suma de residuos al cuadrado

$$\underset{(n\times 1)}{\mathbf{y}} = \underbrace{\underset{(n\times k)}{\mathbf{X}}\underset{(k\times 1)}{\beta}}_{(n\times 1)} + \underset{(n\times 1)}{\varepsilon}$$

$$\underset{(k\times 1)}{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$$

$$\widehat{\beta} \equiv \arg\min_{\widehat{\beta}} SSR(\widehat{\beta})$$

$$\widehat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Estimación frecuentista con máxima verosimilitud

Qué es la función de verosimilitud?

- La función $\mathcal{L}(\theta|\mathbf{x})$ –denominada función de verosimilitud– es una cantidad matemática que expresa la verosomilitud, plausibilidad, de que el parámetro θ tome un valor concreto en base a la información contenida en la muestra $\mathbf{x} = (x_1, ..., x_n)$
- En la función de verosimilitud $\mathcal{L}(\theta|\mathbf{x})$, la muestra permanece constante, y θ varía en el espacio paramétrico Θ. Situación contraria a $\mathbb{P}(\mathbf{x}|\theta)$, ya que en este caso, θ permanece constante y lo que se compara son los posibles sucesos $\mathbf{x}_1, \mathbf{x}_{2,}...$ para el mismo valor de θ.

Estimación frecuentista con máxima verosimilitud

Sea una variable aleatoria ξ con una función de densidad $f(x;\theta)$. Si se toma una muestra de tamaño n, $x_1, ..., x_n$, la función de densidad conjunta de la muestra será,

$$f(x_1, ..., x_n; \theta) = f(x_1; \theta) \cdots f(x_n; \theta)$$

esta función de densidad conjunta será la función de verosimilitud muestral,

$$\mathcal{L}(\theta|\mathbf{x}) := f(x_1; \theta) \cdots f(x_n; \theta)$$

La elección del estimador $\widehat{\theta}$ entre los posibles valores que puede tomar el parámetro θ se sgue el criterio de elegir aquel $\widehat{\theta}$ tal que,

$$\mathcal{L}(\widehat{\theta}|\mathbf{x}) = \max_{\theta \in \Theta} \mathcal{L}(\theta|\mathbf{x})$$

Estimación frecuentista con máxima verosimilitud

Ya que en general la función de verosimilitud es complicada, para mayor sencillez se calcula el valor de θ en su logaritmo,

$$\ln \mathcal{L}(\widehat{\theta}|\mathbf{x}) = \max_{\theta \in \Theta} \ln \mathcal{L}(\theta|\mathbf{x})$$

$$\ln \mathcal{L}(\theta|\mathbf{x}) = \ln f(x_1;\theta) + \cdots + \ln f(x_n;\theta) = \sum\nolimits_{i=1}^{n} \ln f(x_i;\theta)$$

Una vez hallado el logaritmo se deriva respecto a θ,

$$\frac{\partial \ln \mathcal{L}(\theta|\mathbf{x})}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial \ln f(x_i;\theta)}{\partial \theta} = 0$$

Ejemplo: Estimación frecuentista con máxima verosimilitud del primer momento

Sean una muestra i.i.d. $x := \{x_t\}_{t=1}^{m} = x_1, ..., x_m$, si $x \sim \mathcal{N}(\mu, \sigma^2)$, la función de verosimilitud $\mathcal{L}\left(\mu, \sigma^2\right)$ será,

$$
\begin{aligned}
\mathcal{L}\left(\mu, \sigma^2\right) &= f(x_1, ..., x_m | \mu, \sigma^2) \\
&= \prod_{i=1}^{m} f(x_i | \mu, \sigma^2) \\
&= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(-\frac{\sum_{i=1}^{m} (x_i - \mu)^2}{2\sigma^2}\right)
\end{aligned}
$$

Ejemplo: Estimación frecuentista con máxima verosimilitud del primer momento estadístico

$\mathcal{L}\left(\mu,\sigma^2\right)$ puede escribirse en función de $\overline{x}$,

$$\mathcal{L}\left(\mu,\sigma^2\right) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(-\frac{\sum_{i=1}^{m}\left(x_i-\overline{x}\right)^2 + m\left(\overline{x}-\mu\right)^2}{2\sigma^2}\right).$$

Si se aplica logaritmos a $\mathcal{L}\left(\mu,\sigma^2\right)$ y se maximiza derivando respecto al parámetro de interés μ e igualando a cero:

$$\frac{\partial}{\partial\mu}\log\left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(-\frac{\sum_{i=1}^{m}\left(x_i-\overline{x}\right)^2 + m\left(\overline{x}-\mu\right)^2}{2\sigma^2}\right) = 0$$

se tendrá el estimador máximo verosimil,

$$\widehat{\mu} = \overline{x} = \frac{1}{m}\sum_{i=1}^{m} x_i$$

Estimación frecuentista máximo verosímil del MLRM

Dado un modelo lineal de regresión múltiple (MLRM) en el que:

$$\underset{(n\times 1)}{\mathbf{y}} = \underbrace{\underset{(n\times k)}{\mathbf{X}}\underset{(k\times 1)}{\beta}}_{(n\times 1)} + \underset{(n\times 1)}{\varepsilon}$$

$$\varepsilon|\mathbf{X} \sim \mathcal{N}(0, \sigma^2\mathbf{I}_n)$$

$$\mathbf{y}|\mathbf{X} \sim \mathcal{N}(\mathbf{X}\beta, \sigma^2\mathbf{I}_n)$$

La función de verosimilitud es:

$$\mathcal{L}\left(\beta, \sigma^2 \middle| \mathbf{X}, \mathbf{y}\right) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \left[-\frac{1}{2\sigma^2}\left(\mathbf{y} - \mathbf{X}\beta\right)'\left(\mathbf{y} - \mathbf{X}\beta\right)\right]$$

Y los estimadores que maximizan la función son:

$$\widehat{\beta}_{MV} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Docimasia de hipótesis en el MLRM

En el MLRM la significancia de los parámetros del modelo se evalúa con pruebas de hipótesis basadas el estadístico t:

$$t_i = \frac{\hat{\beta}_i}{\mathrm{SE}(\hat{\beta}_i)}$$

$$H_0 : \beta_i = 0$$

$$H_A : \beta_i \neq 0$$

El p-value se obtiene comparando el estadístico t de cada parámetro contra el fractal de una distribución t de Student con n−k grados de libertad, para n el número de observaciones y k el número de parámetros estimados (incluyendo la intersección)

Errores Tipo I y Tipo II

- Error Tipo I: rechazar una hipótesis nula verdadera.
- Error Tipo II: no rechazar una hipótesis nula falsa (beta).
- La probabilidad de cometer un error de Tipo I (α) es el nivel de significancia
- La potencia de una prueba de hipótesis es igual a (1 - beta)

	H_0 **es cierta**	H_0 **no es cierta**
Se escogió H_0	No hay error (1 − α o verdadero negativo)	Error de tipo II (β o falso negativo)
Se escogió H_1	Error de tipo I (α o falso positivo)	No hay error (1 − β o verdadero positivo)

Importancia de las variables en un MLRM

- La importancia relativa de cada variable en el modelo puede compararse con coeficientes estandarizados.
- Los coeficientes estandarizados permiten comparar la importancia de las variables en el modelo al eliminar las unidades de medida.
- Los coeficientes estandarizados se obtienen al convertir todas las variables a una escala común (media 0 y desviación estándar 1) antes de ajustar el modelo de regresión:

$$Z = \frac{X - \mu}{\sigma}$$

- Los coeficientes estandarizados indican cuántas desviaciones estándar cambia la variable dependiente por un cambio de una desviación estándar en la variable independiente.

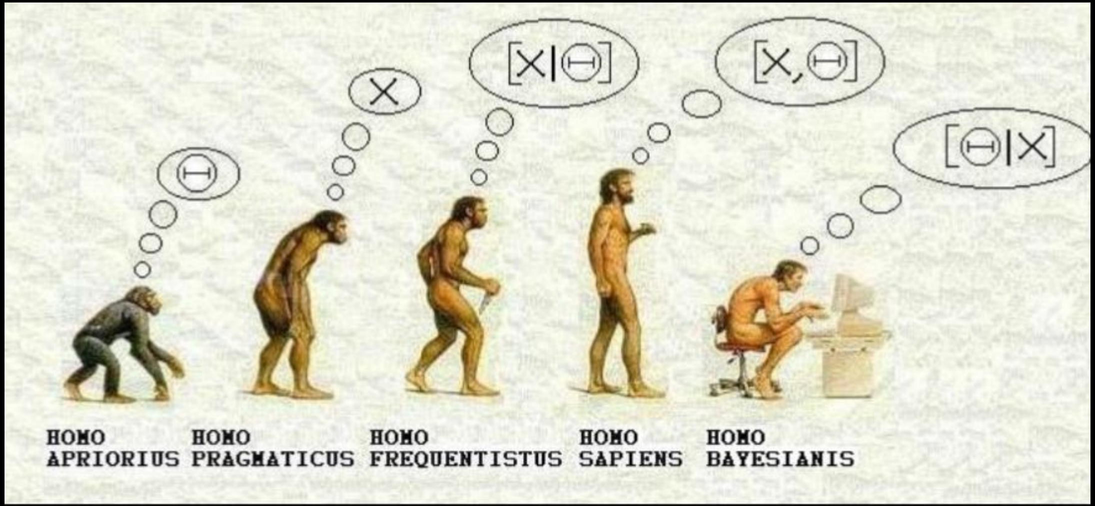
Enfoque Bayesiano

A philosophical theory, like a physics or a mathematical theory, is accessible only to the initiated.

Simone the Beauvoir

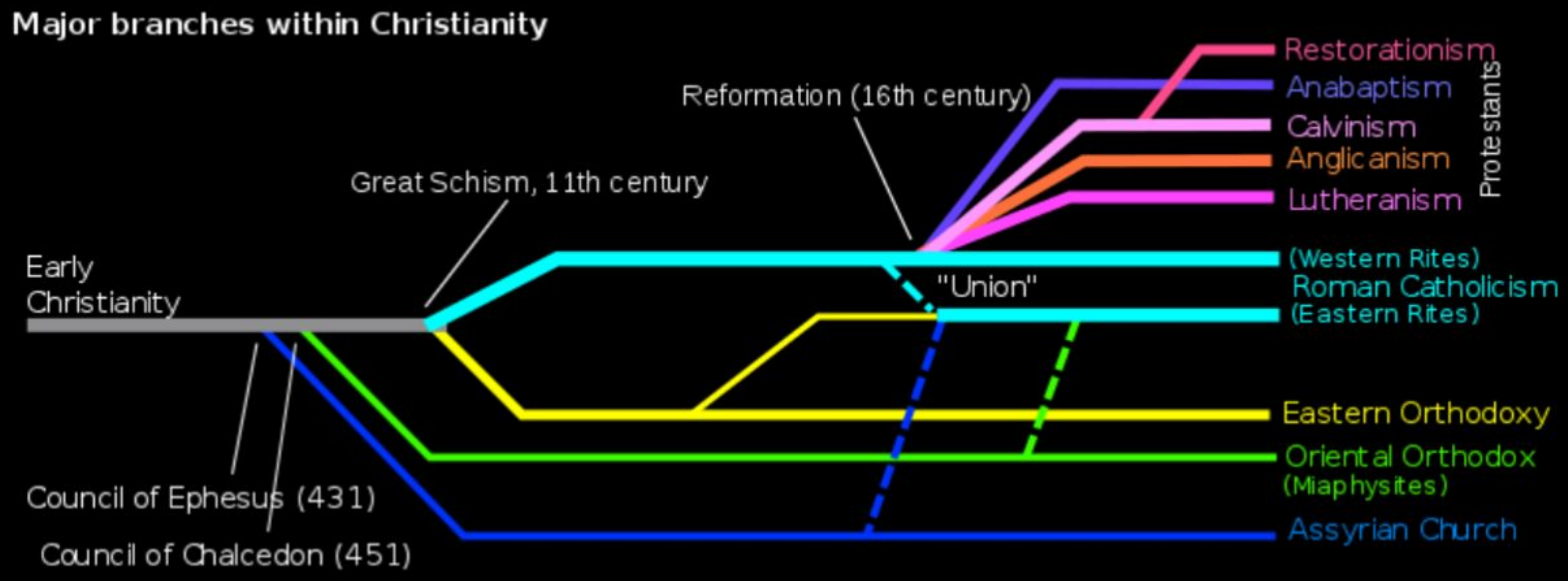
Estimación Bayesiana de modelos lineales

- 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.



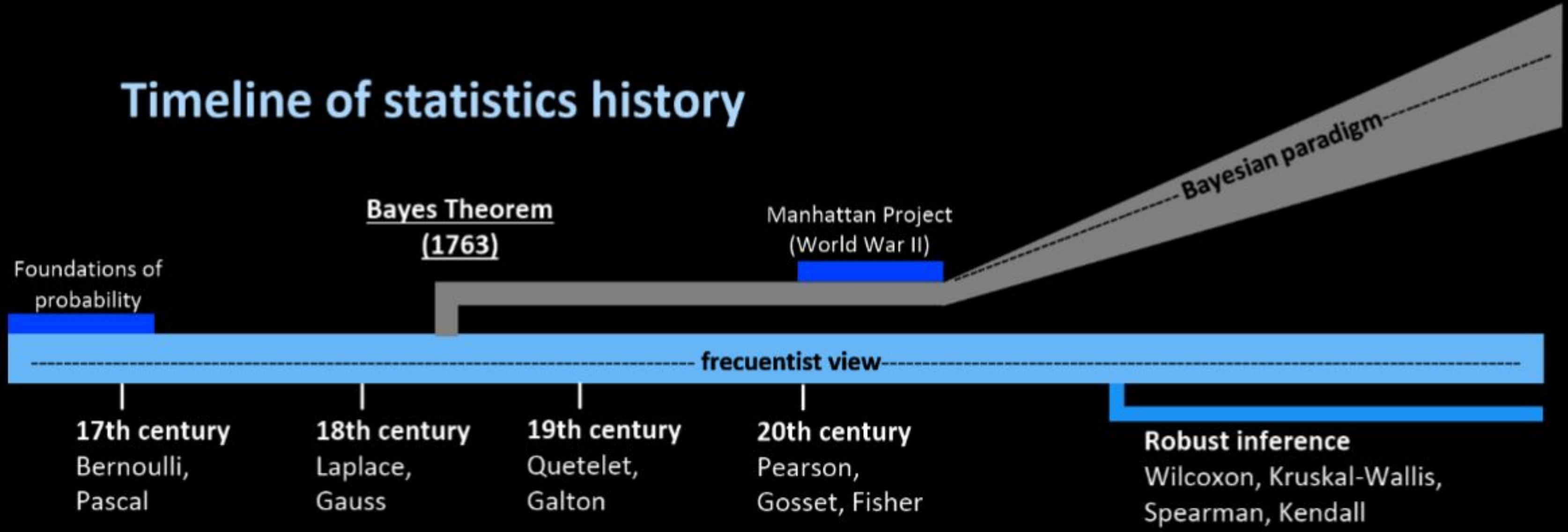
Enfoque Bayesiano

Analogía con el cristianismo



Enfoque Bayesiano

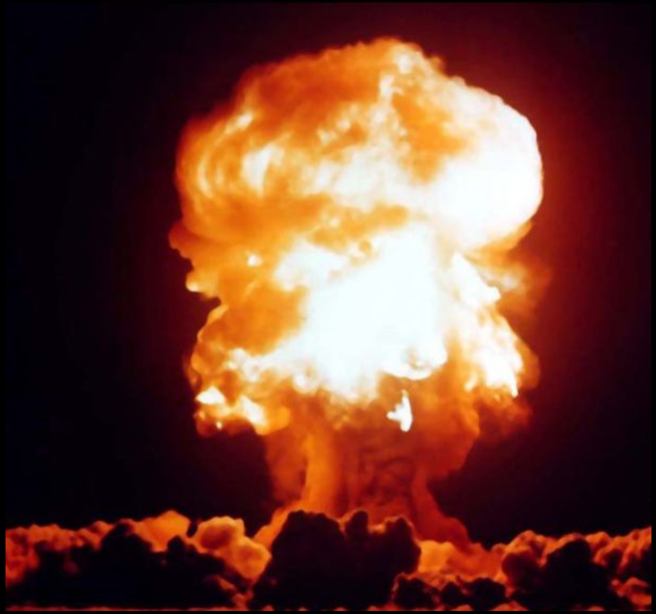
Antes de la segunda guerra mundial se utilizaban conjugados naturales



Enfoque Bayesiano

Antes de la segunda guerra mundial se utilizaban conjugados naturales

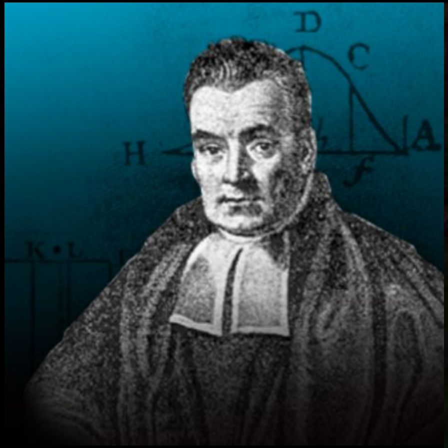
- Métodos de Monte Carlo —desarrollados durante el El Proyecto Manhattan— permitieron aproximar la integrales multidimensionales del análisis Bayesiano.
- El crecimiento exponencial del software y hardware computacional han hecho el uso de los métodos de integración de Monte Carlo más accesible.



Enfoque Bayesiano

- El término bayesiano se refiere a Thomas Bayes (1702 - 1761), quien propuso una solución al problema de probabilidad inversa utilizando lo que hoy se conoce como Teorema de Bayes.
- La solución de Bayes fue publicada póstumamente por Richard Price en Transacciones filosóficas vol. 53 (1763), págs. 370-418
- El mecanismo de Bayes fue descubierto de forma independiente, refinado y publicado en 1774 por Pierre Simon de Laplace

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A)\mathbb{P}(B|A)}{\mathbb{P}(B)}$$



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):

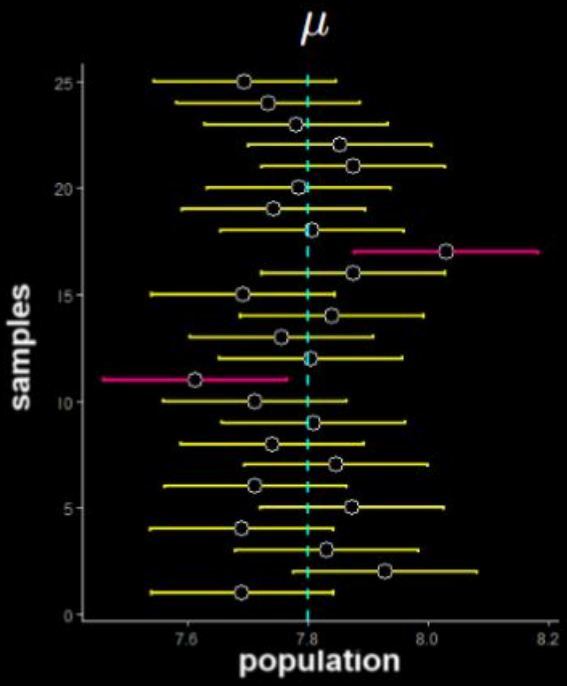
- Para un frecuentista, P(A) es una frecuencia de largo plazo:

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

- 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:

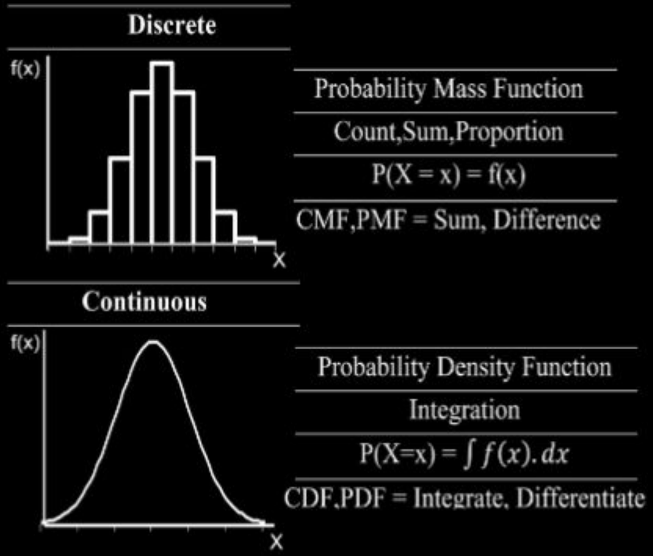
- Los datos son aleatorios
- Los parámetros son (puntos) fijos



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 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:

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

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

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

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

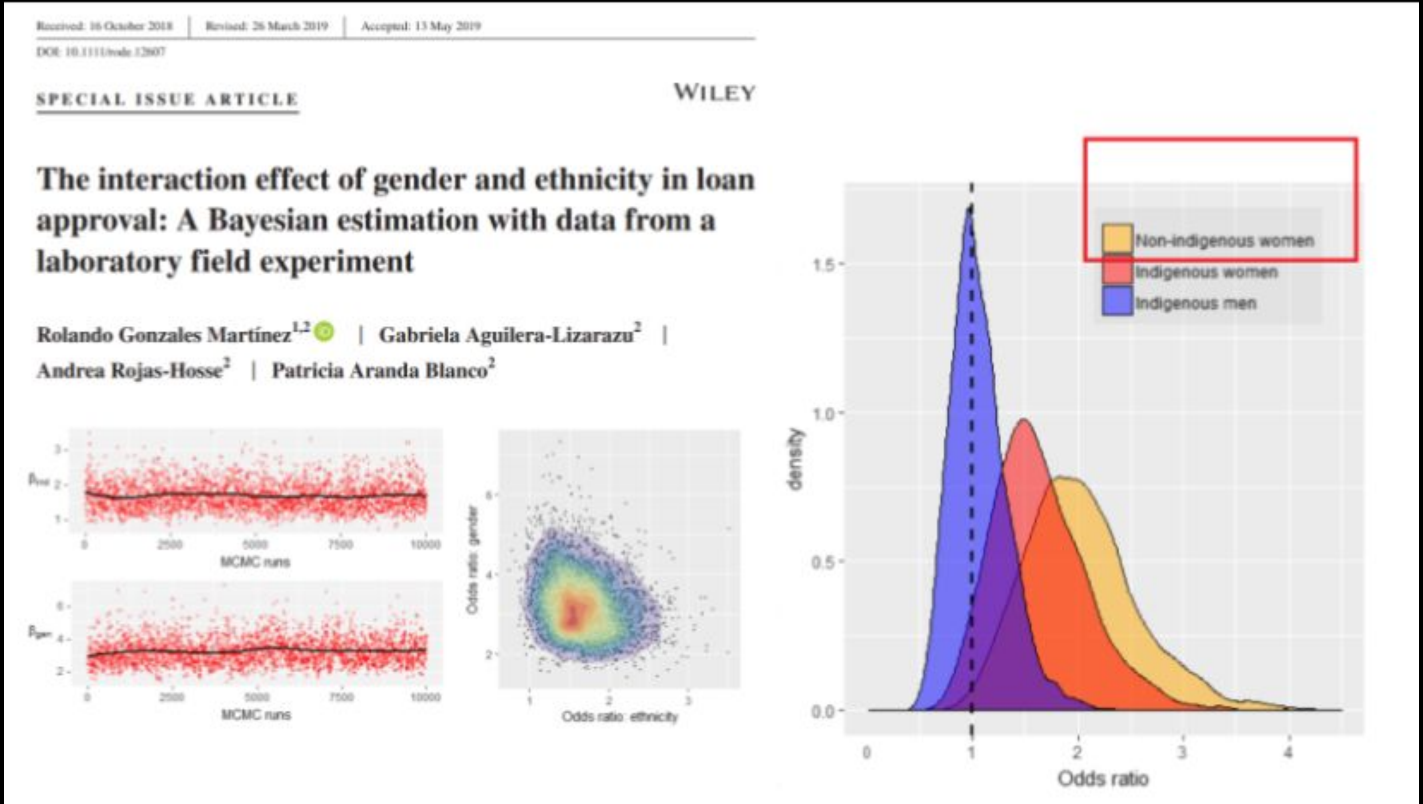
$$\alpha_1 = \alpha_0 + n,$$

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

Estimación Bayesiana de modelos lineales

- 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/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]