Análisis Bayesiano

Naren Mantilla & Alexander Sepúlveda

Universidad Industrial de Santander

5 de diciembre de 2019

Fuente

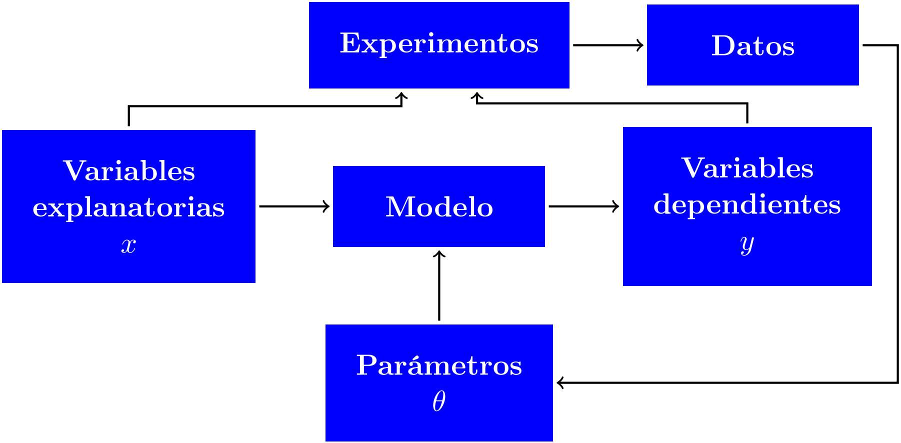
Curso: *Marco Bayesiano para el análisis de datos, calibración de parámetros y modelamiento inverso* **David A. Barajas-Solano**

- Egresado de la UIS – Ingeniería Civil
- Ph.D. en ciencias de la ingeniería, UC San Diego
- Científico en Pacific Northwest National Laboratory
- Perfil científico en `https://www.dbarajassolano.com`

Overview

Outline

- ***Modelos*** son objetos matemáticos usados para describir fenómenos en ciencia e ingeniería.
- Los modelos son ***calibrados*** utilizando datos experimentales.


Experimentos
Datos
Variables explanatorias x
Modelo
Variables dependientes y
Parámetros θ

Ejemplos de modelos

- *Ley de Ohm:*

$$V = R \times I$$

  - Parámetros: Resistencia R.

Ejemplos de modelos

- *Ley de Ohm:*

$$V = R \times I$$

  - Parámetros: Resistencia R.

- *Modelo logístico para variables categóricas:*

$$\log \frac{P(y=1)}{P(y=0)} = \alpha + \beta x$$

  - Variable categórica dependiente $y = \{0, 1\}$.
  - Variable explanatoria continua x.
  - Parámetros $\theta = \alpha, \beta$.

Outline

2. Teoría de Probabilidad

Ud. es el modelador

- Como toda teoría matemática, la teoría de probabilidades no representa realidades físicas. En cambio, es útil para representar conceptos.
- Nuestro objetivo es ***modelar*** información de manera consistente y fácil de interpretar.
- El modelador, ud., juega un rol fundamental.
- El modelador toma decisiones subjetivas.
- La validez de éstas suposiciones está en la utilidad de los resultados.

Experimento

Resultados Ω	**Variables aleatorias** $X : \Omega \to C$
Eventos $\mathcal{F}$	**Probabilidades** $P : \mathcal{F} \to [0, 1]$

Resultado: *Posible resultado de un experimento aleatorio*

Espacio de resultados Ω: "Conjunto" de resultados

Ejemplos

- Lanzar una moneda

$$\Omega = \{\text{cara}, \text{sello}\}$$

- Lanzar un dado

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

- Lanzar dos monedas

$$\Omega = \{(\text{cara}, \text{cara}), (\text{cara}, \text{sello}), (\text{sello}, \text{cara}), (\text{sello}, \text{sello})\}$$

Resultados

El espacio de resultados puede ser...

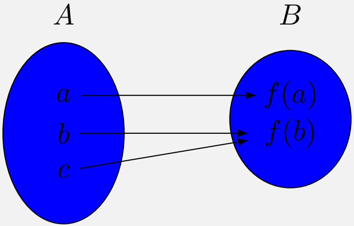
- **Finito contable**
  Ejemplo: Elegir un número de 1 a 10
  $$\Omega = \{1, 2, 3, \ldots, 10\}$$
- **Infinito contable**
  Ejemplo: Elegir un número de 0 a infinito
  $$\Omega = \mathbb{N}^0 = \{0, 1, 2, \ldots\}$$
- **Incontable**
  Ejemplo: Elegir un número real no negativo
  $$\Omega = \mathbb{R}^+ = [0, \infty)$$

Definición: Funciones

La notación

$$f\colon A \to B$$

define la función $f(\cdot)$ que asigna un elemento de B a un elemento A. E.g., $f(a) = b$ indica que la función asocia $b \in B$ a $a \in A$



Funciones de los resultados que representan cantidades asociadas al resultado del experimento

$$X : \Omega \to C$$

Ejemplos

- Lanzar dos dados

$$\Omega = \{(1,1),(1,2),(1,3),\ldots,(2,1),(2,2),\ldots(6,6)\}$$
$$\omega \in \Omega, \quad \omega = (\omega_1, \omega_2)$$

Variable aleatoria: Suma de los dos dados

$$X : \Omega \to \mathbb{N}, \quad X(\omega) = \omega_1 + \omega_2$$

Ejemplos

- Suma de dos dados: $X : \Omega \to \mathbb{N}$, $X(\omega) = \omega_1 + \omega_2$

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Pueden ser *discretas* o *contínuas* ***Ejemplo:*** Altura, ingreso y género de una población

$$\Omega = \{\text{personas en la población}\}$$

- **Discreta**: Género

$$G\colon \Omega \to \{\text{masc.}, \text{fem.}, \dots\}$$

- **Contínua**: Altura

$$A\colon \Omega \to \mathbb{R}$$

- **Ingreso**: Variable contínua o discreta?

Evento: *Subconjunto de resultados del experimento*

Campo de eventos $\mathcal{F}$: “Conjunto” de eventos

Definición: Conjunto-poder

Para Ω *contable*, una selección común de campo de eventos es el conjunto-poder 2^{Ω}, que contiene los siguientes eventos:

- Cada resultado (evento elemental)
- Todos los subconjuntos de Ω
- Todos los resultados (osea Ω)
- Ningún resultado (conjunto vacío, $\varnothing$)

Eventos

Ejemplo: Lanzar un dado

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Algunos eventos en el conjunto potencia (*power set*) $\mathcal{F} = 2^{\Omega}$:

- El número 3
- Un número par
- Un número menor a 4
- Un número entre 2 y 5
- Cualquier número (Ω)
- Ningún número ($\varnothing$)

Ejemplo: Lanzar dos dados

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Ambos dados menor o igual a 4.

Eventos

Ejemplo: Lanzar dos dados

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Dado # 1...

- E_1: Menor a 3.
- E_2: Mayor a 5.

Eventos E_1 y E_2 son *Mutuamente excluyentes*.

Eventos

También es posible definir eventos en términos de los valores de variables aleatorias.

Ejemplo: Lanzar dos dados, suma $X(\omega) = \omega_1 + \omega_2$

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Suma X mayor o igual a 7.

Eventos

También es posible definir eventos en términos de los valores de variables aleatorias.

Ejemplo: Lanzar dos dados

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Suma mayor o igual a 7, pero el dado # 2 es 3 o mayor.

Definición: Axiomas de probabilidades

Las probabilidades son...

1. ... cantidades reales no negativas asociadas a cada evento en $\mathcal{F}$...

$$P \colon \mathcal{F} \to \mathbb{R}^+$$

2. ... aditivas sobre eventos mutuamente excluyentes:

$$\text{Si } A \cap B = \varnothing, \text{ entonces } P(A \cup B) = P(A) + P(B)$$

3. ... que suman a 1 sobre el conjunto de resultados

$$P(\Omega) = 1$$

Definición: Axiomas de probabilidades

Las probabilidades son...

1. ... cantidades reales no negativas asociadas a cada evento en $\mathcal{F}$...

$$P\colon \mathcal{F} \to \mathbb{R}^{+}$$

2. ... aditivas sobre eventos mutuamente excluyentes:

$$\text{Si } A \cap B = \varnothing, \text{ entonces } P(A \cup B) = P(A) + P(B)$$

3. ... que suman a 1 sobre el conjunto de resultados

$$P(\Omega) = 1$$

Los axiomas implican que $0 \leq P(E) \leq 1$ para $E \in \mathcal{F}$

Probabilidades

Ejemplo: Lanzar dos dados, suma $X(\omega) = \omega_1 + \omega_2$

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

- E_i, $i \in [1, 36]$: Eventos elementales
- $P(E_i) = 1/36 \approx 28\,\%$
- $P(X \geq 7) = 21/36 \approx 58\,\%$

Probabilidades

- Hasta ahora probabilidades se han definido de manera enteramente *abstracta.*
- Probabilidad es una *medida*: "Área" o "masa" asociada a una región (evento) del espacio de resultados.
- ¿Qué significa que un evento tenga probabilidad de 58 %?

Probabilidades

- Interpretación ***frecuentista***: Si el experimento se ejecuta n veces, se observará un 58 % de las veces el evento en cuestión.

Probabilidades

- Interpretación ***frecuentista***: Si el experimento se ejecuta n veces, se observará un 58 % de las veces el evento en cuestión.
- Cuál es la interpretación adecuada para la probabilidad de eventos que sólo ocurren una vez?
- E.g. el evento "llover mañana": ¡Sólo hay –un– mañana!

- Interpretación ***frecuentista***: Si el experimento se ejecuta n veces, se observará un 58 % de las veces el evento en cuestión.
- Cuál es la interpretación adecuada para la probabilidad de eventos que sólo ocurren una vez?
- E.g. el evento "llover mañana": ¡Sólo hay –un– mañana!
- **Probabilidades como medida de (in)certidumbre.**

Eventos asociados a variables aleatorias continuas

Considérese la variable aleatoria continua

$$A\colon \Omega \to C$$

Eventos asociados a variables aleatorias continuas

Considérese la variable aleatoria continua

$$A \colon \Omega \to C$$

¿Qué eventos podemos definir en términos de los valores de A?

Para $C \equiv \mathbb{R}$ y cierto valor $a \in C$...

- $P(A \leq a)$, $P(A > a)$, $P(a \leq A \leq b)$, etc.
- ¿$P(A = a)$?
  En general, $A = a$ es un evento con probabilidad trivial $(= 0)$.

Resumen

- Elementos de la teoría de probabilidades:
  - *Resultados, variables aleatorias, eventos, y probabilidades.*
- El modelamiento probabilístico en la práctica se basa en variables aleatorias, eventos asociados a variables aleatorias, y sus probabilidades.
- Usualmente, la definición de espacio de resultados Ω se puede tratar de manera implícita, i.e., no es necesario definir Ω explícitamente.

Outline

3. Variables Aleatorias

Introducción

Ejemplo: Altura de una población

$$\Omega = \{\text{personas en la población}\}$$

- **Variable aleatoria**: Altura

$$A \colon \Omega \to \mathbb{R}$$

- **Eventos**:
  - Altura mayor o igual a cierto valor a: $P(A \geq a)$
  - Altura menor a cierto valor b: $P(A < b)$

Ejemplo: Altura de una población

$$\Omega = \{\text{personas en la población}\}$$

- **Variable aleatoria**: Altura

$$A \colon \Omega \to \mathbb{R}$$

- **Eventos**:
  - Altura mayor o igual a cierto valor a: $P(A \geq a)$
  - Altura menor a cierto valor b: $P(A < b)$

Si el interés es en las *propiedades* de la “altura”...

- ...no es necesario pensar en Ω
- Sólo hace falta pensar en A y los valores que puede tomar

Asocian probabilidades a los valores de variables aleatorias

- **Variables discretas**:
  Función de *masa* de probabilidad (pmf)
- **Variables contínuas**:
  Función de *densidad* acumulada (cdf)
  Función de *densidad* de probabilidad (pdf)

Función de masa de probabilidad (pmf)

- **Variable aleatoria**

$$X : \Omega \to C$$

- **pmf** $p_X : C \to [0, 1]$ definida como

$$p_X(c) = P(X = c) = P\left(\{\omega \in \Omega \colon X(\omega) = c\}\right)$$

Función de masa de probabilidad (pmf)

Ejemplo: Lanzar dos dados. Suma $X : \Omega \to \mathbb{N}$

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

pmf:

- $p_X(1) = 0$
- $p_X(2) = 1/36$
- $p_X(3) = 2/36$
- $p_X(4) = 3/36$

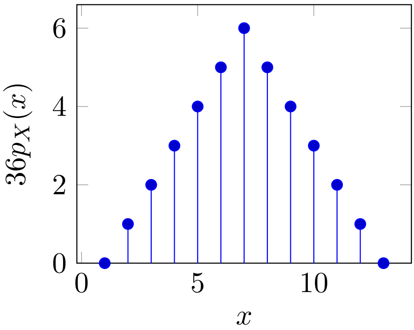
$\ldots$

Función de masa de probabilidad (pmf)

Ejemplo: Lanzar dos dados. Suma $X : \Omega \to \mathbb{N}$

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

pmf:



Función de masa de probabilidad (pmf)

- Para la variable aleatoria discreta $X\colon \Omega \to C$, se puede utilizar la pmf para calcular la probabilidad del evento del valor de X estar en cualquier subconjunto $B \subseteq C$:

$$P(X \in B) = \sum_{b \in B} p_X(b)$$

Función de masa de probabilidad (pmf)

- Para la variable aleatoria discreta $X\colon \Omega \to C$, se puede utilizar la pmf para calcular la probabilidad del evento del valor de X estar en cualquier subconjunto $B \subseteq C$:

$$P(X \in B) = \sum_{b \in B} p_X(b)$$

- Como consecuencia,

$$\sum_{b \in C} p_X(b) = 1$$

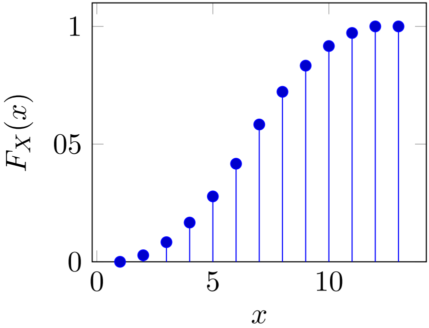
- **Ejemplo**: Lanzar dos dados

$$p_X(1) + p_X(2) + \cdots + p_X(11) + p_X(12) = 1$$

- **Variable aleatoria**

$$X : \Omega \to C$$

- **cdf** $F_X(c) := P(X \leq c) \in [0, 1]$
- **Ejemplo**: Lanzar dos dados



Función cuantil

Para la variable aleatoria $X \colon \Omega \to C$,

$$Q(p) \coloneqq \text{ínf}\{x \in C : p \leq F_X(x)\}$$

- $Q(p)$ determina el menor valor x de X para el cual $P(X \leq x) \geq p$
- **Ejemplo**: Lanzar dos dados

p	0.1	0.2	0.5	0.7	0.9
$Q(p)$	4	5	7	8	10

Estadísticos

- **Media/Esperanza/Valor esperado**:
$$\mu \equiv \mathbb{E}[X] := \sum_{x \in C} x p_X(x)$$
El promedio de n muestras de la variable aleatoria en el límite $n \to \infty$, i.e.,
$$\frac{x_1 + x_2 + \cdots + x_n}{n} \to \mathbb{E}[X]$$
- **Varianza**:
$$\sigma^2 \equiv \mathbb{E}\left[(X-\mu)^2\right] = \sum_{x \in C} (x-\mu)^2 p_X(x)$$
Medida de desviación de la distribución alrededor de la media
- **Desviación estándar**: $\sigma = \sqrt{\sigma^2}$
- **Mediana**: $Q(05)$
- **Moda**: Valor mas probable, i.e, $\arg\max p_X(x)$

Ejemplo: Variables Aleatorias Discretas

Revisemos el *Notebook* de llamado "*rvs.ipynb*".

- Ingrese a *Google Colaboratory*.
- Cargue el archivo `rvs.ipynb`.
- Siga paso a paso la parte de la guía correspondiente a *Variables aleatorias discretas*.

- **Variable aleatoria**

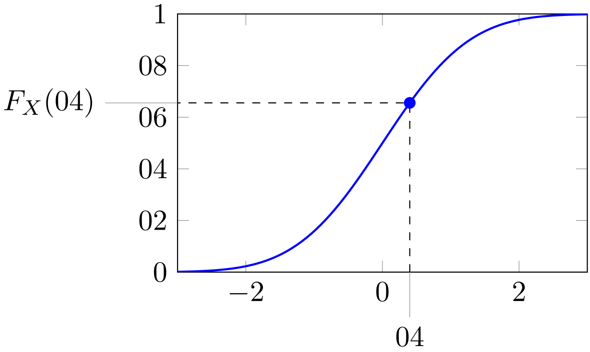
$$X : \Omega \to C$$

- **cdf** $F_X(c) = P(X \leq c) \in [0, 1]$

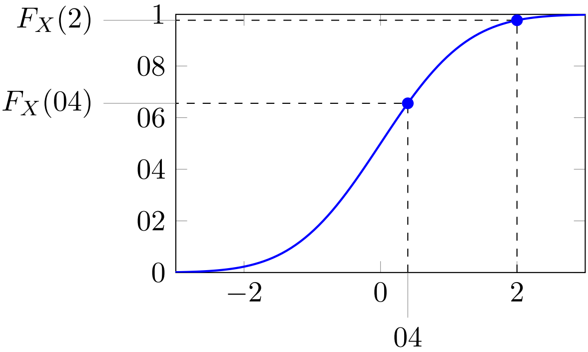
- **Variable aleatoria**

$$X : \Omega \to C$$

- **cdf** $F_X(c) = P(X \leq c) \in [0, 1]$
- **Ejemplo**: Variable aleatoria normal estándar



Función de densidad cumulativa



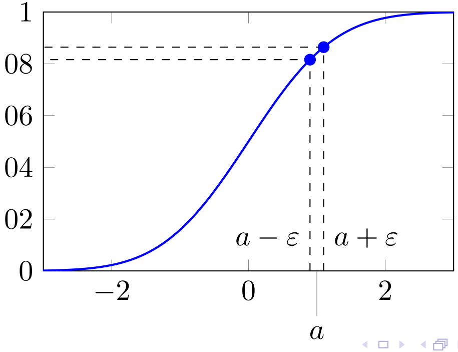
La cdf puede utilizarse para evaluar la probabilidad de varios escenarios:

- $P(X > 2) = 1 - P(X \leq 2) = 1 - F_X(2)$
- $P(04 \leq X \leq 2) = P(X \leq 2) - P(X \leq 04) = F_X(2) - F_X(04)$

Función de densidad de probabilidad (pdf)

- Para generalizar la pmf a variables aleatorias contínuas, hay que introducir el concepto de *densidad*
- La pdf (o *densidad* o *distribución*) $p_X(a)$ cuantifica la cantidad de probabilidad en un intervalo alrededor de a relativa a la longitud del intervalo

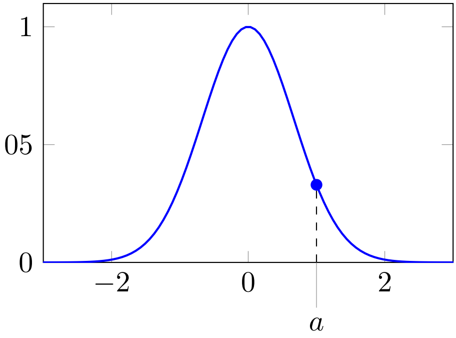
Ejemplo: Variable aleatoria normal estándar



Función de densidad de probabilidad (pdf)

- Para generalizar la pmf a variables aleatorias contínuas, hay que introducir el concepto de *densidad*
- La pdf (o *densidad* o *distribución*) $p_X(a)$ cuantifica la cantidad de probabilidad en un intervalo alrededor de a relativa a la longitud del intervalo

Ejemplo: Variable aleatoria normal estándar



Función de densidad de probabilidad (pdf)

- Si entendemos la probabilidad como la "masa" de una región del espacio de resultados (evento), la distribución $p_X(\cdot)$ indica la densidad en cada punto de esa región
- Para variables aleatorias *discretas*,

$$P(X \in B) = \sum_{b \in B} p_X(b)$$

- Para variables aleatorias *contínuas*,

$$P(X \in B) = \int_B p_X(x)\, dx$$

- La integral $\int (\cdot) dx$ generaliza la suma $\sum$ a variables aleatorias continuas.

- Para la variable aleatoria $X : \Omega \to C$, la pdf y la cdf están conectadas a través de la relación

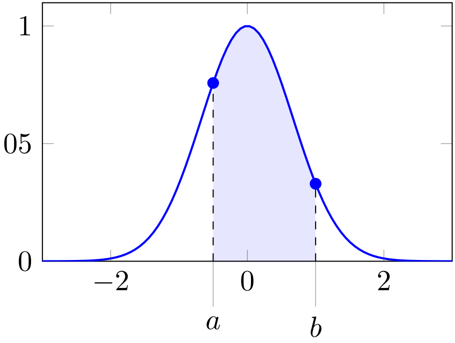
$$F_X(a) = P(X \leq a) = \int_{-\infty}^{a} p_X(x)\, dx$$

- Como consecuencia de la definición,

$$P(X \in C) = \int_C p_X(x)\, dx = 1$$

Área bajo la curva

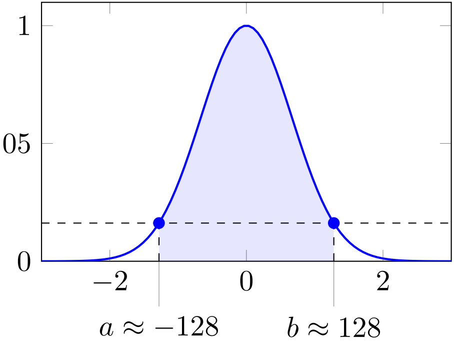
- La integral $\int_B p_X(x)\,dx$ puede interpretarse como el área bajo la curva $y = p(x)$ en el intervalo B.
- E.g., para el intervalo $B = (a, b)$,



- Dado que $\int_C p_X(x)\,\mathrm{d}x = 1$, eso indica que el área bajo la pdf es 1.
- La cdf $F_X(a)$ corresponde al área bajo la curva entre $-\infty$ y a.

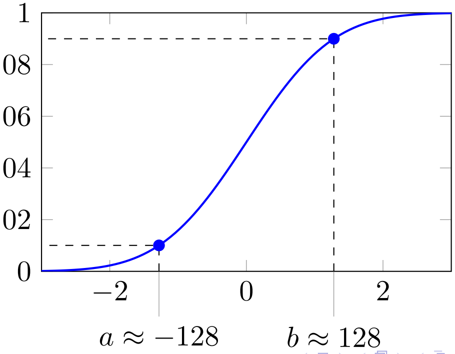
Ejemplo

- Variable aleatoria normal estándar $X \sim \mathcal{N}(0, 1)$.
- **Intervalo de mayor densidad (HDI)**: Para un valor $\alpha \in [0, 1]$, hay un intervalo $B = (a, b)$ para el cual:
  - $p_X(a) = p_X(b)$.
  - $P(X \in B) = 1 - \alpha$.
- Para $\alpha = 02$:



Ejemplo

- Variable aleatoria normal estándar $X \sim \mathcal{N}(0, 1)$.
- **Intervalo de mayor densidad (HDI)**: Para un valor $\alpha \in [0, 1]$, hay un intervalo $B = (a, b)$ para el cual:
  - $p_X(a) = p_X(b)$.
  - $P(X \in B) = 1 - \alpha$.
- Para $\alpha = 02$:



Ejemplo

- Para la variable aleatoria $X \sim \mathcal{N}(\mu, \sigma)$ y $\alpha = 005$,

$$\text{HDI} = (\mu - 196\sigma, \mu + 196\sigma)$$

- Éste intervalo se conoce comúnmente como el *intervalo de credibilidad de 95 %.*

Ejemplo: Variables Aleatorias Continuas

Revisemos el *Notebook* de llamado “*rvs.ipynb*".

- Ingrese a *Google Colaboratory*.
- Cargue el archivo `rvs.ipynb`.
- Siga paso a paso la parte de la guía correspondiente a *Variables aleatorias continuas*.

Outline

Resultados de un referendo en dos regiones

		Región 1	
		Gana G_1	Pierde P_1
Región 2	Gana G_2	A 02	B 01
	Pierde P_2	C 04	D 03

Podemos pensar en términos de eventos

- $P(G_1) = P(A) + P(C) = 06.$
- $P(P_1) = P(B) + P(D) = 04.$
- Etc.

Resultados de un referendo en dos regiones

		Región 1	
		Gana $Y_1 = 1$	Pierde $Y_1 = 0$
Región 2	Gana $Y_2 = 1$	$(1,1)$ 02	$(0,1)$ 01
	Pierde $Y_2 = 0$	$(1,0)$ 04	$(0,0)$ 03

O podemos pensar en variables aleatorias

- Y_1: Resultado en la región 1
- Y_2: Resultado en la región 2
- Variable vectorial

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$$

equivalente a $Y = [Y_1, Y_2]^\top$

		Región 1	
		Gana $Y_1 = 1$	Pierde $Y_1 = 0$
Región 2	Gana $Y_2 = 1$	$(1,1)$ 02	$(0,1)$ 01
	Pierde $Y_2 = 0$	$(1,0)$ 04	$(0,0)$ 03

Introducimos la *pmf conjunta*

$$\begin{aligned} p_Y(y) &= p_{Y_1Y_2}(y_1, y_2) \\ &= P\left(\{Y_1 = y_1\} \& \{Y_2 = y_2\}\right) \end{aligned}$$

- $p_{Y_1Y_2}(1,0) = 04$
- $p_{Y_1Y_2}(0,0) = 03$
- Etc.

Considérese las dos variables aleatorias

$$X : \Omega \to C_X, \quad Y : \Omega \to C_Y$$

- **pmf conjunta**

$$p_{XY}(x, y) = P\left(\{X = x\} \& \{Y = y\}\right)$$

Probabilidad conjunta y marginal

Considérese las dos variables aleatorias

$$X : \Omega \to C_X, \quad Y : \Omega \to C_Y$$

- **pmf conjunta**

$$p_{XY}(x, y) = P\left(\{X = x\}\&\{Y = y\}\right)$$

La pmf conjunta completamente define la distribución de probabilidad de ambas variables aleatorias

Considérese las dos variables aleatorias

$$X : \Omega \to C_X, \quad Y : \Omega \to C_Y$$

- **pmf conjunta**

$$p_{XY}(x, y) = P\left(\{X = x\} \& \{Y = y\}\right)$$

La pmf conjunta completamente define la distribución de probabilidad de ambas variables aleatorias

- **pmfs marginales**

$$p_X(x) = \sum_{y \in C_Y} p_{XY}(x, y), \quad p_Y(y) = \sum_{x \in C_X} p_{XY}(x, y)$$

Resultados de un referendo en dos regiones

<table>
<tr><td></td><td></td><td colspan="2">Región 1</td></tr>
<tr><td></td><td></td><td>Gana
$Y_1 = 1$</td><td>Pierde
$Y_1 = 0$</td></tr>
<tr><td rowspan="2">Región 2</td><td>Gana
$Y_2 = 1$</td><td>$(1,1)$
02</td><td>$(0,1)$
01</td></tr>
<tr><td>Pierde
$Y_2 = 0$</td><td>$(1,0)$
04</td><td>$(0,0)$
03</td></tr>
</table>

pmf marginal de Y_1

$$
\begin{aligned}
p_{Y_1}(1) &= \sum_{y_2} p_{Y_1Y_2}(1, y_2) \\
&= p_{Y_1Y_2}(1,1) + p_{Y_1Y_2}(1,0) \\
&= 02 + 04 = 06 \\
p_{Y_1}(0) &= \sum_{y_2} p_{Y_1Y_2}(0, y_2) \\
&= p_{Y_1Y_2}(0,1) + p_{Y_1Y_2}(0,0) \\
&= 01 + 03 = 04
\end{aligned}
$$

Independencia

Dos variables aleatorias son independientes si y sólo si

$$p(x,y) = p(x)p(y) \ \ \text{para todo} \ \ x \in C_X, \ y \in C_Y$$

i.e., si la pmf conjunta se factoriza en términos de las pmfs marginales

- Independencia implica que observar una variable aleatoria no provee información acerca del valor de la otra variable aleatoria

		Región 1	
		Gana $Y_1 = 1$	Pierde $Y_1 = 0$
Región 2	Gana $Y_2 = 1$	$(1,1)$ 02	$(0,1)$ 005
	Pierde $Y_2 = 0$	$(1,0)$ 01	$(0,0)$ 065

- $p_{Y_1}(1) = 03$
- $p_{Y_2}(1) = 025$
- $p_{Y_1 Y_2}(1,1) = 02 \neq 03 \times 025$

... por lo tanto, Y_1 y Y_2 **no** son independientes

Probabilidad condicional

Introducimos la *pmf condicional* $p_{X|Y}(x \mid y)$

- Indica la probabilidad de $X = x$ dado que se ha observado que $Y = y$
- Relacionada con la pmf conjunta y las pmfs marginales a través de la relación

$$p_{XY}(x, y) = p_{X|Y}(x \mid y)p_Y(y)$$

- Naturalmente,

$$p_{XY}(x, y) = p_{Y|X}(y \mid x)p_X(x)$$

Resultados de un referendo en dos regiones

		Región 1	
		Gana $Y_1 = 1$	Pierde $Y_1 = 0$
Región 2	Gana $Y_2 = 1$	$(1,1)$ 02	$(0,1)$ 005
	Pierde $Y_2 = 0$	$(1,0)$ 01	$(0,0)$ 065

Se observa que el referendo *gana* en la región 2. Cuál es la probabilidad de ganar en la región 1?

$$
\begin{aligned}
p_{Y_1|Y_2}(1 \mid 1) &= \frac{p_{Y_1 Y_2}(1,1)}{p_{Y_2}(1)} \\
&= \frac{02}{025} \approx 08
\end{aligned}
$$

Resultados de un referendo en dos regiones

		Región 1	
		Gana $Y_1 = 1$	Pierde $Y_1 = 0$
Región 2	Gana $Y_2 = 1$	$(1, 1)$ 02	$(0, 1)$ 005
	Pierde $Y_2 = 0$	$(1, 0)$ 01	$(0, 0)$ 065

Se observa que el referendo *gana* en la región 2. Cuál es la probabilidad de ganar en la región 1?

$$p_{Y_1|Y_2}(1 \mid 1) = \frac{p_{Y_1 Y_2}(1, 1)}{p_{Y_2}(1)}$$
$$= \frac{02}{025} \approx 08$$

Compárese:

- $p_{Y_1}(1) = 03$
- $p_{Y_1|Y_2}(1 \mid 1) = 08$

Resultados de un referendo en dos regiones

		Región 1	
		Gana $Y_1 = 1$	Pierde $Y_1 = 0$
Región 2	Gana $Y_2 = 1$	$(1,1)$ 02	$(0,1)$ 005
	Pierde $Y_2 = 0$	$(1,0)$ 01	$(0,0)$ 065

Se observa que el referendo *gana* en la región 2. Cuál es la probabilidad de ganar en la región 1?

$$p_{Y_1|Y_2}(1 \mid 1) = \frac{p_{Y_1 Y_2}(1,1)}{p_{Y_2}(1)}$$
$$= \frac{02}{025} \approx 08$$

Compárese:

- $p_{Y_1}(1) = 03$
- $p_{Y_1|Y_2}(1 \mid 1) = 08$

Observar Y_2 provee información acerca de Y_1

Independencia

Para dos variables aleatorias independientes,

$$p_{XY}(x, y) = p_X(x)p_Y(y)$$

Por lo tanto,

$$p_{X|Y}(x \mid y) = \frac{p_{XY}(x, y)}{p_Y(y)} = \frac{p_X(x)p_Y(y)}{p_Y(y)} = p_X(x)$$

i.e, la probabilidad de $X = x$ es independiente del valor de Y

Considérese dos variables aleatorias contínuas

$$X : \Omega \to C_X, \quad Y : \Omega \to C_Y$$

- **pdf conjunta** $p_{XY}(x, y)$
  Generaliza el concepto de pmf conjunta a variables contínuas
- **pdf marginales**

$$p_X(x) = \int_{C_Y} p_{XY}(x, y)\, dy, \quad p_Y(y) = \int_{C_X} p_{XY}(x, y)\, dx$$

Ejemplo: Variable aleatoria normal multivariante (n-variante)

$$Y = [Y_1, Y_2, \ldots, Y_n]^\top, \quad Y \sim \mathcal{N}(\mu, C),$$

donde

- $\mu = [\mu_1, \mu_2, \ldots, \mu_n]^\top$ es el vector de medias
- C es la matrix $n \times n$ de *covarianza*

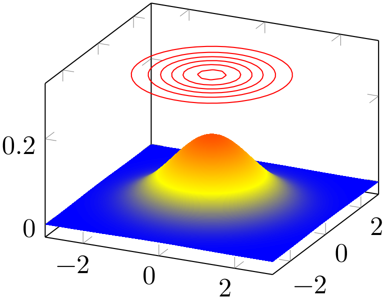
En el caso bivariante,

$$C = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2, \end{bmatrix}$$

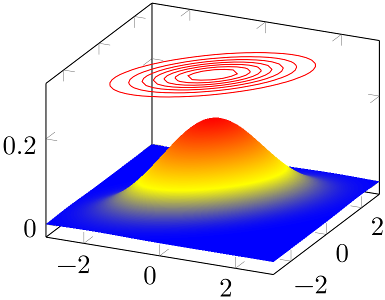
donde ρ denota la correlación entre Y_1 y Y_2

Variables aleatorias contínuas

Ejemplo: Variable aleatoria normal bivariante

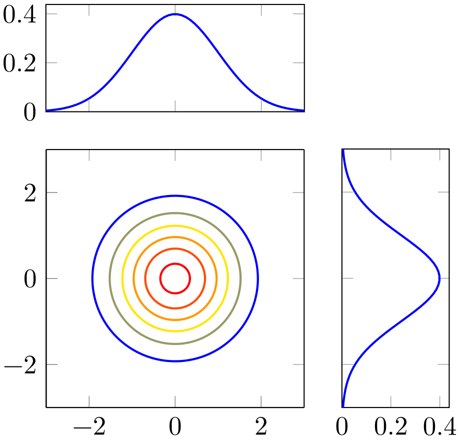


$\rho = 00$



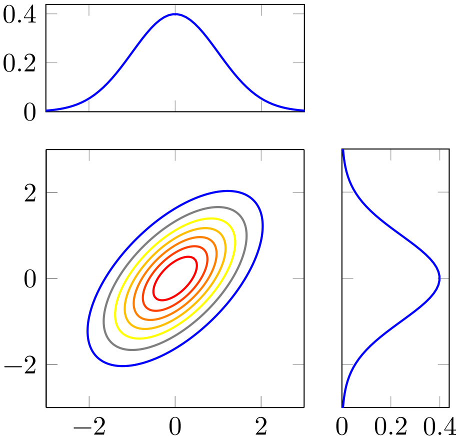
$\rho = 06$

Ejemplo: Variable aleatoria normal bivariante con $\rho = 00$



Variables aleatorias contínuas

Ejemplo: Variable aleatoria normal bivariante con $\rho = 06$



Lanzar una moneda cargada

- Se lanza una moneda cargada n (e.g., $n = 100$) veces
- Se observa un número y (e.g., $y = 43$) de caras
- Cuál es la probabilidad θ de obtener cara?

Lanzar una moneda cargada

- Se lanza una moneda cargada n (e.g., $n = 100$) veces
- Se observa un número y (e.g., $y = 43$) de caras
- Cuál es la probabilidad θ de obtener cara?

Cómo podemos resolver éste problema?

- *Frecuentista*: Estimador de máxima verosimilitud

$$\hat{\theta} = \frac{y}{n} = \frac{43}{100} = 043$$

Lanzar una moneda cargada

- Se lanza una moneda cargada n (e.g., $n = 100$) veces
- Se observa un número y (e.g., $y = 43$) de caras
- Cuál es la probabilidad θ de obtener cara?

Cómo podemos resolver éste problema?

- *Frecuentista*: Estimador de máxima verosimilitud

$$\hat{\theta} = \frac{y}{n} = \frac{43}{100} = 043$$

- *Bayesiano*

Lanzar una moneda cargada

Análisis Bayesiano

1. Tratamos a y y θ como variables aleatorias Y y Θ
2. Observamos Y y buscamos calcular la distribución de P condicional a la observación de Y

$$p_{\Theta|Y}(\theta \mid y) = \frac{p_{Y\Theta}(y,\theta)}{p_Y(y)} = \frac{p_{Y\Theta}(y,\theta)}{\int p_{Y\Theta}(y,\theta)\,dy}$$

Análisis Bayesiano

1. Tratamos a y y θ como variables aleatorias Y y Θ
2. Observamos Y y buscamos calcular la distribución de P condicional a la observación de Y

$$p_{\Theta|Y}(\theta \mid y) = \frac{p_{Y\Theta}(y,\theta)}{p_Y(y)} = \frac{p_{Y\Theta}(y,\theta)}{\int p_{Y\Theta}(y,\theta)\,dy}$$

Necesitamos especificar la densidad conjunta $p_{Y\Theta}(y,\theta)$

Lanzar una moneda cargada

Cómo especificar la densidad conjunta $p_{Y\Theta}(y, \theta)$?

- Si se conoce θ, podemos modelar la densidad de Y

$$p_{Y|\Theta}(y \mid \theta) = \text{Binomial}(y \mid \theta)$$

  En otras palabras, podemos modelar la densidad condicional $p_{Y|\Theta}(y \mid \theta)$

Lanzar una moneda cargada

Cómo especificar la densidad conjunta $p_{Y\Theta}(y, \theta)$?

- Si se conoce θ, podemos modelar la densidad de Y

$$p_{Y|\Theta}(y \mid \theta) = \text{Binomial}(y \mid \theta)$$

  En otras palabras, podemos modelar la densidad condicional $p_{Y|\Theta}(y \mid \theta)$

- Dada la definición de probabilidad condicional,

$$p_{Y\Theta}(y, \theta) = p_{Y|\Theta}(y \mid \theta) p_{\Theta}(\theta)$$

  En otras palabras, hace falta especificar $p_{\Theta}(\theta)$

Lanzar una moneda cargada

Cómo especificar $p_\Theta(\theta)$?

- Antes de observar Y no sabemos mucho de la distribución de los valores de Θ. Es decir, no sabemos qué valores de Θ son más probables que otros...
- ...pero sabemos que $\Theta \in [0, 1]$

Lanzar una moneda cargada

Cómo especificar $p_\Theta(\theta)$?

- Antes de observar Y no sabemos mucho de la distribución de los valores de Θ. Es decir, no sabemos qué valores de Θ son más probables que otros...
- ...pero sabemos que $\Theta \in [0, 1]$
- Entonces podemos asumir que $\Theta \sim U(0, 1)$.

Lanzar una moneda cargada

Cómo especificar $p_\Theta(\theta)$?

- Antes de observar Y no sabemos mucho de la distribución de los valores de Θ. Es decir, no sabemos qué valores de Θ son más probables que otros...
- ...pero sabemos que $\Theta \in [0, 1]$
- Entonces podemos asumir que $\Theta \sim U(0, 1)$.

Utilizamos el concepto de probabilidad como medida de **(in)certidumbre**

- La suposición $\Theta \sim U(0, 1)$ codifica nuestro nivel de certidumbre acerca del valor de de Θ

Lanzar una moneda cargada

$$p_{\Theta|Y}(\theta \mid y) = \frac{P_{Y|\Theta}(y \mid \theta) p_{\Theta}(\theta)}{\int p_{Y|\Theta}(y \mid \theta) p_{\Theta}(\theta)\, d\theta}$$

- Tenemos $p_{Y|\Theta}(y \mid \theta) = \text{Binomial}(y \mid \theta)$
- Tenemos $p_{\Theta}(\theta) = U(\theta \mid 0, 1)$
- Listo! Podemos evaluar la integral del denominador y ya...
- ...pero no vamos a hacer eso

$$p_{\Theta|Y}(\theta \mid y) = \frac{P_{Y|\Theta}(y \mid \theta) p_{\Theta}(\theta)}{\int p_{Y|\Theta}(y \mid \theta) p_{\Theta}(\theta)\, d\theta}$$

- Calcular la integral en el denominador (*función de partición*) es sencillo cuando son pocos parámetros pero es muy costoso computacionalmente cuando son muchos parámetros
- Utilizaremos una familia de algoritmos, **Monte Carlo de cadenas de Markov (MCMC)** para generar directamente muestras de la distribución $p_{\Theta|Y}(\theta \mid y)$ sin tener qué calcular la integral

Monte Carlo de cadenas de Markov (MCMC)

- Se generan una o varias “cadenas” de números
- En el límite de convergencia, las cadenas corresponden a conjuntos de muestras de la distribución condicional $p_{\Theta|Y}(\theta \mid y)$
- Es el algoritmo de cálculos Bayesianos más costoso y más general

Ejemplo: Binomial

Revisemos el *Notebook* llamado "*binomial.ipynb*".

- Acceda a `GoogleColaboratory`.
- Cargue el archivo `binomial.ipynb`.
- Siga paso a paso la guía.

Outline

5 Modelamiento probabilístico

- **Objetivo**: Modelar datos
- Si tenemos observaciones y de un observable Y y un modelo de las observaciones parametrizado por Θ, podemos calcular la distribución de los parámetros dadas las observaciones usando

$$p(\theta \mid y) = \frac{p(y \mid \theta)p(\theta)}{p(y)} = \frac{p(y \mid \theta)p(\theta)}{\int p(y \mid \theta)p(\theta)\, d\theta}$$

$$p(\theta \mid y) = \frac{p(y \mid \theta)p(\theta)}{p(y)} = \frac{p(y \mid \theta)p(\theta)}{\int p(y \mid \theta)p(\theta)\, d\theta}$$

- $p(y \mid \theta)$: **Verosimilitud**
  El valor de la pdf de las observaciones dado el valor específico $\Theta = \theta$ de los parámetros. La verosimilitud indica qué tan posible es observar las (uh) observaciones dado θ

- $p(\theta)$: **(Distribución) Anterior**
  Codifica la información disponible o suposiciones acerca de la distribución de probabibilidad de Θ, i.e, indica qué valores de Θ son más o menos probables de acuerdo a la información *a priori*

Nomenclatura

$$p(\theta \mid y) = \frac{p(y \mid \theta)p(\theta)}{p(y)} = \frac{p(y \mid \theta)p(\theta)}{\int p(y \mid \theta)p(\theta)\, d\theta}$$

- $p(\theta \mid y)$: **(Distribución) Posterior**
  pdf de Θ dadas las observaciones, i.e., indica qué valores de Θ son más o menos probables de acuerdo a las observaciones y la información *a priori*

- $p(y) = \int p(y \mid \theta)p(\theta)\, d\theta$: **Verosimilitud marginal**
  Indica qué valores del observable Y son más o menos posibles dada la información *a priori* acerca de los parámetros del modelo

- Para calcular la distribución posterior sólo hace falta especificar las verosimilitud $p(y \mid \theta)$ y la distribución anterior, i.e, la distribución conjunta
$$p(y,\theta) = p(y \mid \theta)p(\theta)$$
- Ésta distribución conjunta se conoce como el *modelo probabilístico*

Tareas de regresión

- Queremos analizar la dependencia de un observable Y en una variable explanatoria x utilizando un modelo parametrizado por Θ
- Dadas las observaciones (X, y), podemos calcular la distribución de los parámetros:

$$p(\theta \mid y, X) = \frac{p(y \mid X, \theta)p(\theta)}{p(y \mid X)} = \frac{p(y \mid X, \theta)p(\theta)}{\int p(y \mid X, \theta)p(\theta)\, d\theta}$$

donde

- $p(y \mid X, \theta)$: Verosimilitud
- $p(\theta)$: Anterior
- $p(\theta \mid y, X)$: Posterior

Ejemplo: bioassay

Revisemos el *Notebook* llamado "*bioassay.ipynb*".

- Ingrese a *Google Colaboratory*.
- Siga paso a paso la guía.

Regresión lineal ordinaria

- Tenemos un observable Y, que se asume continua y k variables explanatorias $x = (x_1, x_2, \ldots, x_k)$
- Un número n de observaciones de la forma

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1k} \\ x_{21} & x_{22} & \ldots & x_{2k} \\ \vdots & & & \vdots \\ x_{i1} & x_{i2} & \ldots & x_{ik} \\ \vdots & & & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{nk} \end{bmatrix}$$

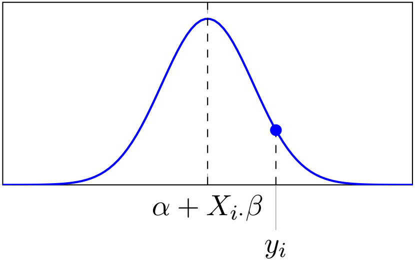
- Modelo lineal para la esperanza de las observaciones

$$\mathbb{E}(y_i \mid X, \theta) = \alpha + x_{i1}\beta_1 + x_{i2}\beta_2 + \ldots x_{ik}\beta_k$$
$$\mathbb{E}(y \mid X, \theta) = \alpha + X\beta$$

Verosimilitud normal

- Para modelar las observaciones, asumimos que cada observación es independiente, y modelada como una variable aleatoria normal con la esperanza dada por el modelo lineal y una cierta varianza

$$p(y_i \mid X, \theta) = \mathcal{N}(y \mid \alpha + X_{i\cdot}\beta, \sigma), \quad \theta = \{\alpha, \beta, \sigma\}$$



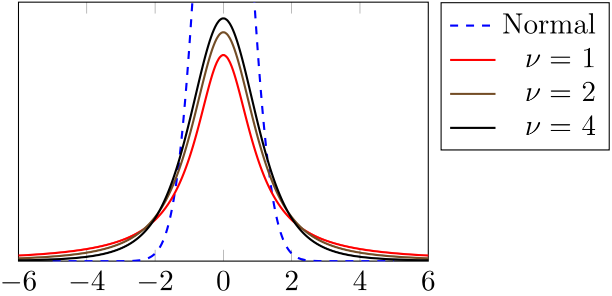
- La distribución conjunta de las observaciones es

$$p(y \mid X, \theta) = \prod_{i=1}^{n} p(y_i \mid X, \theta) = \mathcal{N}(y \mid \alpha + X\beta, \sigma^2 I)$$

Robustez contra valores extremos

- La distribución normal tiene colas “cortas”. Por lo tanto, observaciones lejos del predictor $\alpha + X\beta$ para cierto valor de θ se consideran muy poco probables
- Éstas observaciones “arrastran” el estimador hacia ellas
- Para robustecer la inferencia contra valores extremos es necesario utilizar una verosimilitud con colas más largas

Robustez contra valores extremos



t-Distribución de Student

- Cómo decidir el valor de ν?
- Modelamiento jerárquico.

- Las distribuciones anteriores de los parámetros, θ, se eligen usualmente como distribuciones parametrizadas de la forma $p(\theta \mid \nu)$ con sus propios parámetros ν, o *hiperparámetros*
- Podemos asignar *distribuciones hiperanteriores* $p(\nu)$ a éstos hiperparámetros

$$p(\theta, \nu \mid y, X) = \frac{p(y \mid X, \theta)p(\theta \mid \nu)p(\nu)}{\int p(y \mid X, \theta)p(\theta \mid \nu)p(\nu)\, d\theta d\nu}$$

Ejemplo:Robustez

Revisemos el *Notebook* de Jupyter llamado “*robust.ipynb*".

- Ingrese a *Google Colaboratory*.
- Siga paso a paso la guía.