Marco Bayesiano para el análisis de datos, calibración de parámetros y modelamiento inverso

Probabilidad conjunta y condicional

04

Universidad Industrial de Santander
U18 Fest

23–26 de septiembre de 2019

Resultados de un referendo en dos regiones

		Región 1	
		Gana G_1	Pierde P_1
Región 2	Gana G_2	A 0.2	B 0.1
	Pierde P_2	C 0.4	D 0.3

Podemos pensar en términos de eventos

- $P(G_1) = P(A) + P(C) = 0.6$
- $P(P_1) = P(B) + P(D) = 0.4$
- Etc.

Resultados de un referendo en dos regiones

		Región 1	
		Gana $Y_1 = 1$	Pierde $Y_1 = 0$
Región 2	Gana $Y_2 = 1$	$(1, 1)$ 0.2	$(0, 1)$ 0.1
	Pierde $Y_2 = 0$	$(1, 0)$ 0.4	$(0, 0)$ 0.3

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

Resultados de un referendo en dos regiones

		Región 1	
		Gana $Y_1 = 1$	Pierde $Y_1 = 0$
Región 2	Gana $Y_2 = 1$	$(1, 1)$ 0.2	$(0, 1)$ 0.1
	Pierde $Y_2 = 0$	$(1, 0)$ 0.4	$(0, 0)$ 0.3

Introducimos la *pmf conjunta*

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

- $p_{Y_1 Y_2}(1, 0) = 0.4$
- $p_{Y_1 Y_2}(0, 0) = 0.3$
- Etc.

Probabilidad conjunta y marginal

Considérese las dos variables aleatorias

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

- **pmf conjunta**

$$p_{XY}(x, y) = P(\{X = x\} \cup \{Y = y\})$$

Probabilidad conjunta y marginal

Considérese las dos variables aleatorias

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

- **pmf conjunta**

$$p_{XY}(x, y) = P(\{X = x\} \cup \{Y = y\})$$

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

Probabilidad conjunta y marginal

Considérese las dos variables aleatorias

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

- **pmf conjunta**

$$p_{XY}(x,y) = P(\{X = x\} \cup \{Y = y\})$$

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

		Región 1	
		Gana $Y_1 = 1$	Pierde $Y_1 = 0$
Región 2	Gana $Y_2 = 1$	$(1,1)$ 0.2	$(0,1)$ 0.1
	Pierde $Y_2 = 0$	$(1,0)$ 0.4	$(0,0)$ 0.3

pmf marginal de Y_1

$$
\begin{aligned}
p_{Y_1}(1) &= \sum_{y_2} p_{Y_1 Y_2}(1, y_2) \\
&= p_{Y_1 Y_2}(1,1) + p_{Y_1 Y_2}(1,0) \\
&= 0.2 + 0.4 = 0.6 \\
p_{Y_1}(0) &= \sum_{y_2} p_{Y_1 Y_2}(0, y_2) \\
&= p_{Y_1 Y_2}(0,1) + p_{Y_1 Y_2}(0,0) \\
&= 0.1 + 0.3 = 0.4
\end{aligned}
$$

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

Resultados de un referendo en dos regiones

		Región 1	
		Gana $Y_1 = 1$	Pierde $Y_1 = 0$
Región 2	Gana $Y_2 = 1$	$(1, 1)$ 0.2	$(0, 1)$ 0.05
	Pierde $Y_2 = 0$	$(1, 0)$ 0.1	$(0, 0)$ 0.65

- $p_{Y_1}(1) = 0.3$
- $p_{Y_2}(1) = 0.25$
- $p_{Y_1 Y_2}(1, 1) = 0.2 \neq 0.3 \times 0.25$

... 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 $Y_1 = 1$	Región 1 Pierde $Y_1 = 0$
Región 2	Gana $Y_2 = 1$	$(1,1)$ 0.2	$(0,1)$ 0.05
Región 2	Pierde $Y_2 = 0$	$(1,0)$ 0.1	$(0,0)$ 0.65

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

$$
\begin{aligned}
p_{Y_1|Y_2}(1 \mid 1) &= \frac{p_{Y_1Y_2}(1,1)}{p_{Y_2}(1)} \\
&= \frac{0.2}{0.25} \approx 0.8
\end{aligned}
$$

Resultados de un referendo en dos regiones

		Región 1	
		Gana $Y_1 = 1$	Pierde $Y_1 = 0$
Región 2	Gana $Y_2 = 1$	$(1, 1)$ 0.2	$(0, 1)$ 0.05
	Pierde $Y_2 = 0$	$(1, 0)$ 0.1	$(0, 0)$ 0.65

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

$$p_{Y_1|Y_2}(1 \mid 1) = \frac{p_{Y_1 Y_2}(1, 1)}{p_{Y_2}(1)} = \frac{0.2}{0.25} \approx 0.8$$

Compárese:

- $p_{Y_1}(1) = 0.3$
- $p_{Y_1|Y_2}(1 \mid 1) = 0.8$

Resultados de un referendo en dos regiones

Región 2		Región 1: Gana $Y_1 = 1$	Región 1: Pierde $Y_1 = 0$
	Gana $Y_2 = 1$	$(1, 1)$ 0.2	$(0, 1)$ 0.05
	Pierde $Y_2 = 0$	$(1, 0)$ 0.1	$(0, 0)$ 0.65

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

$$p_{Y_1|Y_2}(1 \mid 1) = \frac{p_{Y_1 Y_2}(1, 1)}{p_{Y_2}(1)} = \frac{0.2}{0.25} \approx 0.8$$

Compárese:

- $p_{Y_1}(1) = 0.3$
- $p_{Y_1|Y_2}(1 \mid 1) = 0.8$

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

Variables aleatorias contínuas

Considérese dos variables aleatorias contínuas

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

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

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

Variables aleatorias contínuas

Ejemplo: Variable aleatoria normal multivariante (n-variante)

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

donde

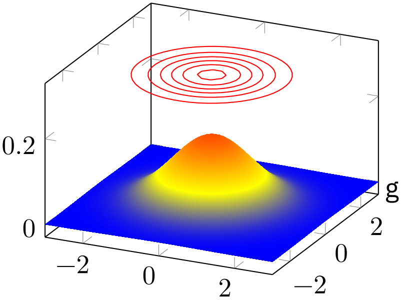
- $\mu = [\mu_1, \mu_2, \dots, \mu_n]^\top$ es el vector de medias
- C es la matrix $n \times n$ de *covarianza*

En el caso bivariante,

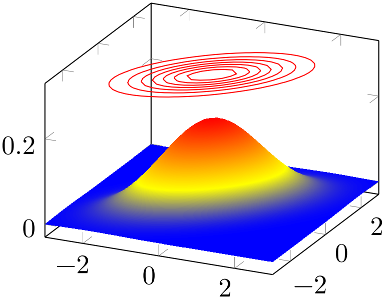
$$C = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2, \end{bmatrix}$$

donde ρ denota la correlación entre Y_1 y Y_2

Ejemplo: Variable aleatoria normal bivariante



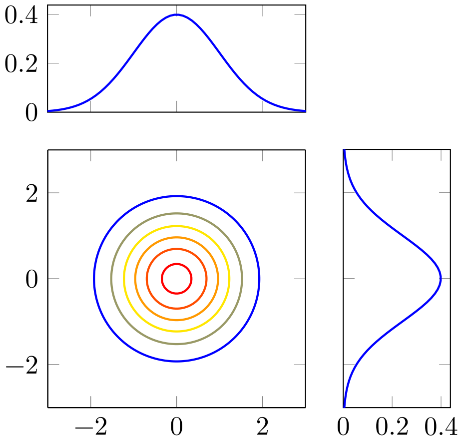
$\rho = 0.0$



$\rho = 0.6$

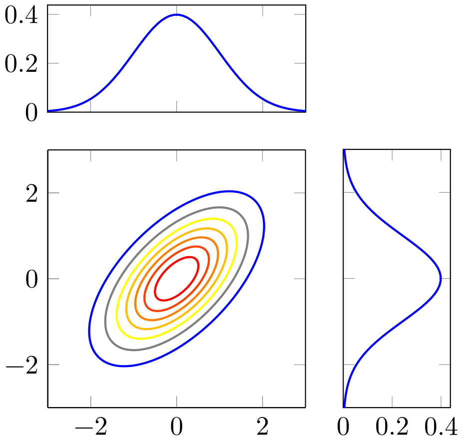
Variables aleatorias contínuas

Ejemplo: Variable aleatoria normal bivariante con $\rho = 0.0$



Variables aleatorias contínuas

Ejemplo: Variable aleatoria normal bivariante con $\rho = 0.6$



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

$$\hat{\theta} = \frac{y}{n} = \frac{43}{100} = 0.43$$

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

$$\hat{\theta} = \frac{y}{n} = \frac{43}{100} = 0.43$$

- *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

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

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

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

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 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 $\Theta \sim U(0, 1)$

Acá 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

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

- 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

Cálculos Bayesianos

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

- 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