Variables aleatorias bidimensionales: distribución condicional

Antonio Sala

Dept. Ing. Sistemas y Automatica (DISA)
Universitat Politècnica de València (UPV)

Video-presentación disponible en:

http://personales.upv.es/asala/YT/V/**prcond**.html

Presentación

Motivación:

El objetivo del análisis multivariante es describir la relación entre múltiples variables aleatorias. El caso más sencillo es 2 variables. En muchos casos, una de las dos variables es “medible”, y otra no, con lo que deseamos “estimarla” a partir de la primera.

Objetivos:

Comprender los conceptos asociados a la probabilidad condicional en variables continuas y discretas.

Contenidos:

Revisión de conceptos. Probabilidad condicional. Ejemplos. Conclusiones.

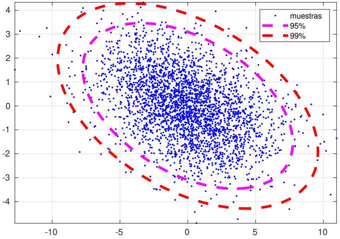
Variable bidimensional $\zeta \in \mathbb{R}^2$

Consideremos una variable aleatoria bidimensional

$$\zeta := (x, y) \in \Omega := \mathbb{X} \times \mathbb{Y} \subseteq \mathbb{R}^2$$

con una función de densidad $f : \mathbb{R}^2 \mapsto \mathbb{R}^+$ tal que $\int_{\mathbb{X}} \int_{\mathbb{Y}} f(x, y)\, dxdy = 1$.

Ejemplo: distribución normal bidimensional



$$f(x, y) = A e^{-\frac{1}{2}\left(a(x-x_0)^2+2b(x-x_0)(y-y_0)+c(y-y_0)^2\right)}$$

$$f(\zeta) = \frac{1}{2\pi\sqrt{|\Sigma|}} e^{-\frac{1}{2}(\zeta-\zeta_0)^T \Sigma^{-1}(\zeta-\zeta_0)}$$

Ejemplo discreto:

Una empresa fabrica balones con la siguiente distribución de "color" (vble. categórica) y "tamaño" (vble. numérica):

Tam. \ Col.	Rojo	Verde
10 cm	10%	45%
15 cm	30%	15%

El par (tamaño, color) de un balón al azar es una vble. aleatoria bidimensional discreta.

La distribución marginal de la variable aleatoria 1-D "$x :=$*tamaño*" es:

$$p_x(10cm) = 0.1 + 0.45 = 0.55, \quad p_x(15cm) = 0.3 + 0.15 = 0.45$$

La distribución marginal de la variable aleatoria 1-D "$y :=$*color*" es:

$$p_y(Rojo) = 0.1 + 0.3 = 0.4, \quad p_y(Verde) = 0.45 + 0.15 = 0.6$$

Probabilidad condicional (caso discreto)

Probabilidad condicional de y dado x.

$$p(y|x) := \frac{p(x,y)}{p_x(x)} = \frac{p(x,y)}{\sum_y p(x,y)}$$

Fórmula de Bayes

$$p(y|x)p_x(x) = p(x,y) = p(x|y)p_y(y)$$

Ejemplo: En el caso de los balones (tamaño, color),

Tam. \ Col.	Rojo	Verde
10 cm	10%	45%
15 cm	30%	15%

Tendríamos que, sabiendo que el color es "Rojo" (prob. marginal 0.4), el tamaño verifica:

$$f(10cm|Rojo) = \frac{0.1}{0.4} = 0.25, \quad f(15cm|Rojo) = \frac{0.3}{0.4} = \underbrace{0.75}_{\text{Mejor predicción condicional (moda)}}$$

Del mismo modo, sabiendo que el balón es de 15cm, podríamos decir:

$$f(Rojo|10cm) = \frac{0.3}{0.45} = \underbrace{0.67}_{\text{Mejor predicción condicional (moda)}}, \qquad f(Verde|10cm) = \frac{0.15}{0.45} = 0.33$$

Densidad condicional

Densidad condicional de y dado x.

$$f(y|x) := \frac{f(x,y)}{f_x(x)} = \frac{f(x,y)}{\int_{\mathbb{Y}} f(x,y)dy}$$

Fórmula de Bayes

$$f(y|x)f_x(x) = f(x,y) = f(x|y)f_y(y)$$

Ejemplo: Si $f(x,y) = \frac{3}{20}(x+2y)^2$ definida en $-1 \le x \le 1$, $-1 \le y \le 1$, las marginales son:

$$f_x(x) = \int_{-1}^{1} f(x,y)\,dy = 0.4 + 0.3x^2, \qquad f_y(y) = \int_{-1}^{1} f(x,y)\,dx = 0.1 + 1.2y^2$$

Las distrib. condicionales son:

$$f(y|x) = \frac{f(x,y)}{f_x(x)} = \frac{3(x+2y)^2}{2(3x^2+4)}, \qquad f(x|y) = \frac{f(x,y)}{f_y(y)} = \frac{3(x+2y)^2}{2(12y^2+1)}$$

Las medias de las condicionales:

$$E(y|x) := \int_{-1}^{1} y \cdot f(y|x)dy = \frac{4x}{3x^2+4}, \qquad E(x|y) = \int_{-1}^{1} x \cdot f(x|y)dx = \frac{4y}{12y^2+1}$$

Conclusiones

- Distribución "condicional": probabilidad de una variable aleatoria conocido (medido) el valor de otra.
- Es el concepto básico en "predicción".
- Su cálculo necesita "integrales" (un modelo). A veces el resultado es "complicado" o el modelo, llanamente, no está disponible... existen otras posibilidades de predicción, como por ejemplo regresión lineal dado un conjunto de datos.