Repaso de Probabilidad (2)

Lección 06

Dr. Pablo Alvarado Moya

CE5506 Introducción al reconocimiento de patrones
Área de Ingeniería en Computadores
Tecnológico de Costa Rica

II Semestre, 2019

Contenido

Funciones de distribución acumulada conjunta y marginal

- Sean X e Y variables aleatorias.
- Las variables por separado tienen distribuciones acumuladas $F_X(x)$ y $F_Y(y)$
- **Distribución acumulada conjunta** representa relación entre las dos variables:

$$F_{XY}(x,y) = P(X \leq x, Y \leq y)$$

- La relación entre la CDF acumulada y las separadas es

$$F_X(x) = \lim_{y \to \infty} F_{XY}(x,y)$$
$$F_Y(y) = \lim_{x \to \infty} F_{XY}(x,y)$$

- En este contexto $F_X(x)$ y $F_Y(y)$ son las funciones de distribución acumuladas **marginales** de $F_{XY}(x,y)$

Propiedades de CDF conjunta

- $0 \leq F_{XY}(x,y) \leq 1$
- $\lim_{x,y\to\infty} F_{XY}(x,y) = 1$
- $\lim_{x,y\to-\infty} F_{XY}(x,y) = 0$
- $F_X(x) = \lim_{y\to\infty} F_{XY}(x,y)$

Funciones de masa de probabilidad conjunta y marginal (1)

- Sean X e Y dos variables aleatorias discretas
- La función de masa de probabilidad conjunta $p_{XY} : \mathbb{R} \times \mathbb{R} \to [0,1]$ es:

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

- $0 \leq P_{XY}(x,y) \leq 1$ para todo x e y
- $\sum_{x \in \mathrm{val}(X)} \sum_{y \in \mathrm{val}(Y)} P_{XY}(x,y) = 1$
- La función de masa de probabilidad marginal de X es

$$p_X(x) = \sum_{y \in \mathrm{val}(Y)} p_{XY}(x,y)$$

Funciones de masa de probabilidad conjunta y marginal (2)

- La función de masa de probabilidad marginal de Y es

$$p_Y(y) = \sum_{x \in \mathrm{val}(X)} p_{XY}(x,y)$$

- **Marginalización**: formar distribución marginal con respecto a una variable, eliminando la otra variable por medio de la suma

Funciones de densidad de probabilidad conjunta y marginal (1)

- Sean X e Y variables aleatorias continuas con función de distribución acumulada conjunta F_{XY} (continua y suave)
- La función de densidad de probabilidad conjunta es:

$$f_{XY}(x,y) = \frac{\partial^2 F_{XY}(x,y)}{\partial x \partial y}$$

- Se cumple que

$$\iint_{(x,y)\in A} f_{XY}(x,y)\, dx\, dy = P((X,Y) \in A)$$

- **¡Cuidado!** $f_{XY}(x,y) \neq P(X = x, Y = y)$

Funciones de densidad de probabilidad conjunta y marginal (2)

- $f_{XY}(x,y) \geq 0$. Puede ser mayor que uno (¡densidad!)
- Se debe cumplir

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x,y)\, dx\, dy = 1$$

- Marginalización en el caso continuo se realiza con

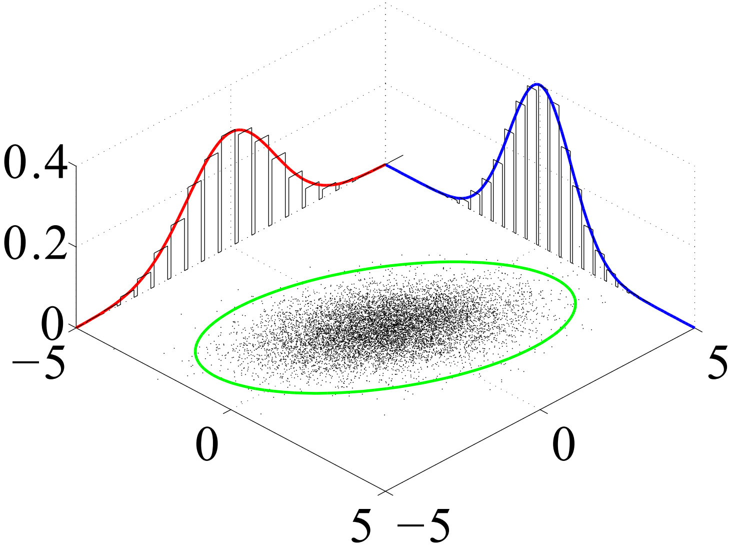
$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\, dy$$

que es la función de densidad de probabilidad **marginal** (o **densidad marginal**) de X.

Funciones de densidad de probabilidad conjunta y marginal (3)

- La marginalización de Y resulta en

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\, dx$$


0.4
0.2
0
−5
0
5
−5
0
5

Distribuciones condicionales

- Distribución condicional busca cuál es la distribución de probabilidad de Y, cuando sabemos que X debe tomar un cierto valor.
- El caso discreto, la función de masa de probabilidad condicional de Y dado X es (se asume $p_X(x) \neq 0$)

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

- El caso continuo, la función de densidad de probabilidad condicional de Y dado X es (por analogía)

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)}$$

con $f_X(x) \neq 0$

Regla de Bayes

- Caso de variables aleatorias discretas

$$P_{Y|X}(y|x) = \frac{P_{XY}(x,y)}{P_X(x)} = \frac{P_{X|Y}(x|y)P_Y(y)}{\sum_{y'\in \mathrm{val}(Y)} P_{X|Y}(x|y')P_Y(y')}$$

- Caso de variables aleatorias continuas

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{f_{X|Y}(x|y)f_Y(y)}{\int_{-\infty}^{\infty} f_{X|Y}(x|y')f_Y(y')\,dy'}$$

Regla de Bayes
Nombres de las partes

- En general, forma de la regla de Bayes es:

$$p(a \mid b) = \frac{p(b \mid a)p(a)}{p(b)}$$

- $p(a \mid b)$ es la probabilidad **a posteriori**,
- $p(b \mid a)$ es valor de **verosimilitud** (o *likelihood*)
- $p(a)$ es la probabilidad **a priori**
- $p(b)$ constante de normalización.

Ejemplo de aplicación de regla de Bayes

Ejemplo: mapas de probabilidad de color

Mapas de probabilidad de color

- **Problema**: encontrar en imagen lugares con objetos de color específico (p. ej. piel).
- Proceso de formación: color depende de iluminación, objeto y sensor
  $\Rightarrow$ variación impredecible de colores
- **Solución**: usar entrenamiento de un "modelo de color" que cubra cambios esperados.
- Defínanse dos clases: *objeto* y $\neg$*objeto*.
- Generar canal con probabilidad de que cada píxel pertenezca a *objeto* o a $\neg$*objeto*, de acuerdo únicamente a su color.

Modelos de probabilidad

- Se estima la probabilidad de que un determinado color c forma parte del objeto con dos **histogramas** de datos de entrenamiento:

$$p(c \mid \text{objeto}) \qquad\qquad p(c \mid \neg\text{objeto})$$

- Se busca $p(\text{objeto} \mid c)$ con regla de Bayes

Probabilidad de ser objeto

- Para caso de probabilidad de ser objeto por su color:

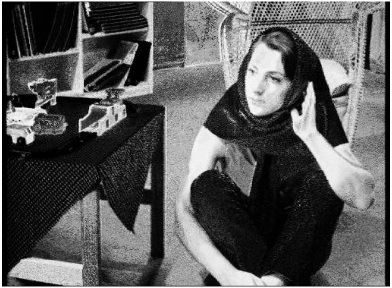
$$p(\text{objeto} \mid c) = \frac{p(c \mid \text{objeto})p(\text{objeto})}{p(c)}$$

y utilizando las reglas de probabilidad se conoce que

$$p(c) = p(c \mid \text{objeto})p(\text{objeto}) + p(c \mid \neg\text{objeto})p(\neg\text{objeto})$$

donde además $p(\neg\text{objeto}) = 1 - p(\text{objeto})$.

Ejemplo



Independencia (1)

- Dos variables aleatorias X e Y son **independientes** si

$$F_{XY}(x,y) = F_X(x)F_Y(y)$$

  para todo valor de x e y
- Para variables discretas
  - $p_{XY}(x,y) = p_X(x)p_Y(y)\ \forall x \in \mathrm{val}(X), \forall y \in \mathrm{val}(Y)$
  - $p_{Y|X}(y|x) = p_Y(y)$ si $p_X(x) \neq 0$, $\forall y \in \mathrm{val}(Y)$
- Para variables continuas
  - $f_{XY}(x,y) = f_X(x)f_Y(y)\ \forall x \in \mathbb{R}, \forall y \in \mathbb{R}$
  - $f_{Y|X}(y|x) = f_Y(y)$ si $f_X(x) \neq 0$, $\forall y \in \mathbb{R}$

Independencia (2)

- Si X e Y son independientes entonces para cualesquiera subconjuntos $A,B \subseteq \mathbb{R}$

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

- Si X e Y son independientes, entonces también lo serán funciones de X e Y

Esperanza y covarianza (1)

- Esperanza de X, Y discretas:

$$\mathrm{E}[g(X,Y)] \stackrel{!}{=} \sum_{x \in \mathrm{val}(X)} \sum_{y \in \mathrm{val}(Y)} g(x,y) p_{XY}(x,y)$$

- Esperanza de X, Y continuas:

$$\mathrm{E}[g(X,Y)] \stackrel{!}{=} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{XY}(x,y)\, dx\, dy$$

- La **covarianza** expresa la relación de dos variables aleatorias:

$$\mathrm{Cov}[X,Y] \stackrel{!}{=} \mathrm{E}[(X - E[X])(Y - E[Y])]$$

- Observe que $\mathrm{Cov}[X,X] = \mathrm{Var}[X]$

Esperanza y covarianza (2)

- De modo similar que para la varianza se demuestra que

$$\mathrm{Cov}[X,Y] = \mathrm{E}[(X - E[X])(Y - E[Y])] = \mathrm{E}[XY] - E[X]E[Y]$$

- Si $\mathrm{Cov}[X,Y] = 0$ entonces X e Y están **decorreladas**

Propiedades

- Linealidad:
  $\mathrm{E}[af(X,Y) + bg(X,Y)] = a\,\mathrm{E}[f(X,Y)] + b\,\mathrm{E}[g(X,Y)]$
- $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\,\mathrm{Cov}[X,Y]$
- Si X e Y son independientes entonces $\mathrm{Cov}[X,Y] = 0$
- Si X e Y son independientes, entonces
  $\mathrm{E}[f(X)g(Y)] = \mathrm{E}[f(X)]\,\mathrm{E}[f(Y)]$

Múltiples variables

Propiedades básicas (1)
Caso de PDF

- Función de distribución acumulada conjunta de $X_1, X_2, \ldots, X_n$

$$F_{X_1,X_2,\ldots,X_n} = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n)$$

- Función de densidad de probabilidad conjunta de $X_1, \ldots, X_n$

$$f_{X_1,X_2,\ldots,X_n}(x_1, \ldots, x_n) = \frac{\partial^n F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)}{\partial x_1 \ldots \partial x_n}$$

- Función de densidad de probabilidad marginal

$$f_{X_1}(X_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,X_2,\ldots,X_n}(x_1, \ldots, x_n)\, dx_2 \ldots dx_n$$

Propiedades básicas (2)
Caso de PDF

- Función de densidad de probabilidad condicional

$$f_{X_1|X_2,\ldots,X_n}(x_1|x_2,\ldots,x_n) = \frac{f_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n)}{f_{X_2,\ldots,X_n}(x_2,\ldots,x_n)}$$

- Probabilidad de evento $A \subseteq \mathbb{R}^n$

$$P((x_1,\ldots,x_n) \in A) = \int_{(x_1,\ldots,x_n)\in A} f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)\, dx_1 \ldots dx_n$$

Regla de la cadena

- De probabilidades condicionales multivarias se demuestra que:

$$\begin{aligned}
f(x_1 \dots x_n) &= f(x_n|x_1,\dots,x_{n-1})f(x_1,\dots,x_{n-1}) \\
&= f(x_n|x_1,\dots,x_{n-1})f(x_{n-1}|x_1,\dots,x_{n-2})f(x_1,\dots,x_{n-2}) \\
&= f(x_1)\prod_{i=2}^{n} f(x_i|x_1 \dots x_{i-1})
\end{aligned}$$

Independencia

- Para múltiples eventos $A_1 \ldots A_k$ se dicen ser **mutualmente independientes** si para cualquier subconjunto $S \subseteq \{1,2,\ldots,k\}$ se tiene

$$P\left(\cap_{i\in S} A_i\right) = \prod_{i\in S} P(A_i)$$

- Las variables aleatorias $X_1 \ldots X_n$ son independientes si

$$f(X_1,\ldots,X_n) = f(X_1)f(X_2)\ldots f(X_n)$$

Vectores aleatorios (1)

- Todas las derivaciones anteriores se pueden extender a vectores
- Se habla entonces de vectores aleatorios

$$\underline{\mathbf{X}} = \begin{bmatrix} X_1 & X_2 & \dots & X_n \end{bmatrix}^T$$

- Puesto que $X : \Omega \to \mathbb{R}$ entonces $\underline{\mathbf{X}} : \Omega \to \mathbb{R}^n$
- Esperanza con $g : \mathbb{R}^n \to \mathbb{R}$

$$\mathrm{E}[g(\underline{\mathbf{X}})] = \int_{\mathbb{R}^n} g(\underline{\mathbf{x}}) f_{\underline{\mathbf{X}}}(\underline{\mathbf{x}})\, d\underline{\mathbf{x}}$$

Matriz de covarianza

$$\boldsymbol{\Sigma} = \begin{bmatrix} \mathrm{Cov}[X_1,X_1] & \cdots & \mathrm{Cov}[X_1,X_n] \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}[X_n,X_1] & \cdots & \mathrm{Cov}[X_n,X_n] \end{bmatrix}$$
$$= \mathrm{E}[\underline{\mathbf{X}}\underline{\mathbf{X}}^T] - \mathrm{E}[\underline{\mathbf{X}}]\,\mathrm{E}[\underline{\mathbf{X}}]^T$$

- La matriz de covarianza es semidefinida positiva
- La matriz de covarianza es simétrica

Resumen

Este documento ha sido elaborado con software libre incluyendo LATEX, Beamer, GNUPlot, GNU/Octave, XFig, Inkscape, GNU-Make y Subversion en GNU/Linux