7600054 — Sistemas Complexos

Gonzalo Travieso

2020-06-10

Outline

Função de distribuição conjunta

- Até agora, trabalhamos apenas com uma variável aleatória, mas muitas vezes temos múltiplas variáveis aleatórias.
- Vamos estudar um pouco o caso de duas variáveis aleatórias, que chamaremos de X e Y. Os conceitos e resultados se generalizam para mais variáveis.
- Cada uma dessas variáveis possui sua própria função de distribuição cumulativa, $F_X(x)$ e $F_Y(y)$.
- No entanto, o comportamento das duas variáveis pode ser relacionado, caso em que precisamos descrever essa interrelação.
- Introduzimos a função de distribuição cumulativa conjunta $F_{XY}(x,y)$ que é a probabilidade associada ao evento

  $$\{X \leq x, Y \leq y\} = \{X \leq x\} \cap \{Y \leq y\}$$

  isto é, que a variável X tenha valor menor do que x simultaneamente à variável Y tendo valor menor do que y.

Densidade conjunta

- Para variáveis contínuas, podemos descrever a probabilidade de o valor das variáveis X e Y estarem numa região infinitesimal $dx\,dy$ através da função de densidade de probabilidade conjunta $\rho(x,y)$ tal que

$$P(\{x < X \leq x + dx, y < Y \leq y + dy\}) = \rho(x,y)\,dx\,dy.$$

- A probabilidade de os valores de X e Y estarem numa dada região D do plano x,y é então dada por:

$$P((x,y) \in D) = \iint_D \rho(x,y)\,dx\,dy.$$

Densidade conjunta (cont)

- Disso concluímos que

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

ou em outra forma:

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

- Também vemos que

$$\rho(x, y) \geq 0, \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \rho(x, y)\, dx\, dy = 1.$$

Distribuições marginais

- Podemos definir distribuições marginais, que são as distribuições de apenas uma variável, considerando todos os possíveis valores da outra:

$$F_X(x) = F_{XY}(x, \infty), \quad \rho(x) = \int_{-\infty}^{\infty} \rho(x, y)\, dy.$$

$$F_Y(y) = F_{XY}(\infty, y), \quad \rho(y) = \int_{-\infty}^{\infty} \rho(x, y)\, dx.$$

Variáveis discretas

- No caso de variáveis discretas, precisamos apenas definir as probabilidades de cada par de valores:

$$P(\{X = x_i, Y = y_j\}) = p_{ij}.$$

- As probabilidades marginais serão

$$q_i \equiv P(\{X = x_i\}) = \sum_j p_{ij}$$

e

$$r_j \equiv P(\{Y = y_j\}) = \sum_i p_{ij}.$$

- Também:

$$\sum_i q_i = \sum_j r_j = \sum_{ij} p_{ij} = 1.$$

Variáveis independentes

- Dizemos que as variáveis aleatórias X e Y são independente se os eventos $\{X \leq x\}$ e $\{Y \leq y\}$ são independentes.
- Isto é, se
$$P(\{X \leq x, Y \leq y\}) = P(\{X \leq x\})P(\{Y \leq y\}),$$
o que resulta em
$$F_{XY}(x,y) = F_X(x)F_Y(y).$$
- Diferenciando a expressão anterior em relação a x e y encontramos
$$\rho(x,y) = \rho(x)\rho(y).$$
- Para variáveis discretas teremos:
$$p_{ij} = q_i r_j.$$

Exercício

A variável X tem uma distribuição uniforme no intervalo $[0, a]$ e a variável Y tem uma distribuição uniforme no intervalo $[0, b]$. Ademais, as variáveis são independentes.

Exercício

Encontre a densidade de probabilidade conjunta de X e Y, $\rho(x, y)$.

Gaussianas

- Considere duas variáveis gausssianas X e Y independentes, as duas com média zero e desvio padrão σ. Portanto

$$\rho(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$
$$\rho(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{y^2}{2\sigma^2}\right).$$

- A densidade conjunta será então:

$$\rho(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right).$$

- Note que o fator constante é apenas um fator de normalização, para garantir que a integral de $\rho(x, y)$ em todo o plano seja 1.
- Portanto (x, y) estão distribuídas de tal forma que a distância à origem segue uma gaussiana com desvio padrão σ.

Simetria circular

- Uma distribuição tem simetria circular se a densidade de probabilidade em um ponto (x, y) depende apenas da sua distância à origem $r^2 = x^2 + y^2$:

$$\rho(x, y) = \phi(r).$$

- Um teorema estabele que, se X e Y são independentes e a sua densidade de probabilidade conjunta tem simetria circular, então as variáveis são normais, com média 0 e variâncias iguais.

Momentos

- Como no caso de uma única variável, podemos definir momentos de variáveis no caso de mais do que uma variável aleatória.
- Neste caso, temos os momentos de cada variável, e precisamos considerar a distribuição conjunta no seu cálculo.
- Os momentos de ordem k serão dados por

$$\langle X^k \rangle = \int_{-\infty}^{\infty} x^k \rho(x)\, dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^k \rho(x, y)\, dx\, dy$$

e

$$\langle Y^k \rangle = \int_{-\infty}^{\infty} y^k \rho(y)\, dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y^k \rho(x, y)\, dx\, dy.$$

Momentos (cont)

- No caso de variáveis discretas:

$$\langle X^k \rangle = \sum_i \sum_j x_i^k p_{ij},$$

e

$$\langle Y^k \rangle = \sum_i \sum_j y_j^k p_{ij}.$$

Médias e variâncias

- As médias serão

$$\langle X \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x\rho(x, y)\, dx\, dy$$

e

$$\langle Y \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y\rho(x, y)\, dx\, dy.$$

- E as variâncias

$$\sigma_X^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \langle X \rangle)^2 \rho(x, y)\, dx\, dy$$

e

$$\sigma_Y^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - \langle Y \rangle)^2 \rho(x, y)\, dx\, dy.$$

Média de uma função

- Dada uma função de duas variáveis $g(x, y)$, podemos definir uma função de duas variáveis aleatórias X e Y, que é em si uma variável aleatória:

$$g(X, Y)$$

.

- A média dessa variável aleatória é dada por

$$\langle g(X, Y) \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \rho(x, y)\, dx\, dy.$$

- Isto é, para avaliar a média de uma função precisamos ponderar com a probabilidade de cada um dos valores de seus parâmetros.

Covariância

- O fenômeno novo para o caso multivariado é que podemos definir momentos associados com as relações entre as variáveis.
- A covariância entre X e Y é definida como

$$\operatorname{cov}(X, Y) = \langle (X - \langle X \rangle)\,(Y - \langle Y \rangle) \rangle.$$

- Isto pode ser reescrito como:

$$\operatorname{cov}(X, Y) = \langle XY \rangle - \langle X \rangle \langle Y \rangle,$$

pois

$$\begin{aligned} \langle (X - \langle X \rangle)\,(Y - \langle Y \rangle) \rangle &= \langle XY - X\langle Y \rangle - Y\langle X \rangle + \langle X \rangle \langle X \rangle \rangle \\ &= \langle XY \rangle - 2\langle X \rangle \langle Y \rangle + \langle X \rangle \langle Y \rangle \\ &= \langle XY \rangle - \langle X \rangle \langle Y \rangle. \end{aligned}$$

Covariância e variância

- Note que a covariância de uma variável com ela mesma é dada por:

$$\operatorname{cov}(X, X) = \langle X^2 \rangle - \langle X \rangle^2.$$

- Mas isto é a definição da variância de X.

Correlação

- A expressão

$$r_{XY} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

  é denominada a correlação (mais precisamente, o coeficiente de correlação de Pearson, ou simplesmente coeficiente de correlação) entre X e Y.

- Uma outra forma de definir o coeficiente de correlação é como a covariância entre os escores-z de x e y:

$$r_{XY} = \text{cov}(\hat{X}, \hat{Y}),$$

  onde

$$\hat{X} = \frac{X - \langle X \rangle}{\sigma_X}, \quad \hat{Y} = \frac{Y - \langle Y \rangle}{\sigma_Y}.$$

Correlação (cont)

- De fato, sabemos que
$$\langle\hat{X}\rangle = \langle\hat{Y}\rangle = 0.$$
- Portanto

$$\begin{aligned}
\operatorname{cov}(\hat{X},\hat{Y}) &= \langle\hat{X}\hat{Y}\rangle - \langle\hat{X}\rangle\langle\hat{Y}\rangle \\
&= \left\langle \frac{X-\langle X\rangle}{\sigma_X}\frac{Y-\langle Y\rangle}{\sigma_Y} \right\rangle \\
&= \frac{\langle (X-\langle X\rangle)(Y-\langle Y\rangle)\rangle}{\sigma_X\sigma_Y} \\
&= \frac{\operatorname{cov}(X,Y)}{\sigma_X\sigma_Y} \\
&= r_{XY}.
\end{aligned}$$

Variáveis não-correlacionadas

- Dizemos que X e Y são não-correlacionadas se

$$r_{XY} = 0.$$

- Equivalentemente:

$$\langle XY \rangle = \langle X \rangle \langle Y \rangle.$$

Combinações lineares

- Seja Z uma variável aleatória que é uma combinação linear de X e Y:

$$Z = aX + bY.$$

- Então a sua média vale

$$\langle Z\rangle = a\langle X\rangle + b\langle Y\rangle.$$

- A sua variância será:

$$\begin{aligned}\sigma_Z^2 &= \langle (Z - \langle Z\rangle)^2\rangle \\ &= \langle (a(X - \langle X\rangle) + b(Y - \langle Y\rangle))^2\rangle \\ &= a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\,\mathrm{cov}(X, Y).\end{aligned}$$

- No caso em que $\mathrm{cov}(X, Y) = 0$ (variáveis não-correlacionadas), então

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2.$$

Ortogonalidade

- Duas variáveis aleatórias X e Y são ortogonais se

$$\langle XY \rangle = 0.$$

- Neste caso

$$\langle (X+Y)^2 \rangle = \langle X^2 \rangle + \langle Y^2 \rangle.$$

Correlação, Ortogonalidade e Independência

- Se as variáveis X e Y são não-correlacionadas, então as variáveis $X - \langle X \rangle$ e $Y - \langle Y \rangle$ são ortogonais.
- Também, se X e Y são independentes, então elas são não-correlacionadas, pois

$$
\begin{aligned}
\langle XY \rangle &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy\rho(x, y)\, dx\, dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy\rho(x)\rho(y)\, dx\, dy \\
&= \int_{-\infty}^{\infty} x\rho(x)\, dx \int_{-\infty}^{\infty} y\rho(y)\, dy \\
&= \langle X \rangle \langle Y \rangle.
\end{aligned}
$$

- Já variáveis não-correlacionadas não são necessariamente independentes!

Limites do coeficiente de correlação

- A desigualdade de Schwarz estabele que, para quaisquer variáveis aleatórias X e Y
$$\langle XY\rangle^2 \leq \langle X^2\rangle\langle Y^2\rangle.$$
Além disso, a igualdade só ocorre se Y é linearmente proporcional a X:
$$Y = \beta X.$$
- Portanto
$$\langle(X-\langle X\rangle)(Y-\langle Y\rangle)\rangle^2 \leq \langle(X-\langle X\rangle)^2\rangle\langle(Y-\langle Y\rangle)^2\rangle,$$
e temos:
$$\mathrm{cov}(X,Y)^2 \leq \sigma_X^2\sigma_Y^2.$$
- Daí concluímos que
$$|r_{XY}| \leq 1,$$
e $|r_{XY}| = 1$ apenas quando as duas variáveis são linearmente relacionadas.

Problemas com o coeficiente de correlação de Pearson

- O coeficiente r_{XY} diz quanto as variáveis aleatórias X e Y são relacionadas:
  - Se $r_{XY} = 0$ elas são não relacionadas.
  - Se $r_{XY} = 1$ elas são linearmente relacionadas, com ambas crescendo ou decrescendo simultaneamente.
  - Se $r_{XY} = -1$ elas são linearmente relacionadas, com uma crescendo quando a outra decresce.
  - Valores interemediários indicam diversos graus de relacionamento entre as variáveis.
- No entanto ele tem dois problemas:
  1. A relação avaliada é linear: o coeficiente só é alto quando a relação é linear. Variáveis fortemente relacionadas de forma não linear podem ter r_{XY} relativamente baixo.
  2. Para o seu cálculo, usamos o z-score das variáveis, que é uma normalização que só é realmente adequada para variáveis com distribuição gaussiana, ou não muito diferente de gaussiana. Portanto, o uso de r_{XY} para variáveis não-gaussianas pode ser inadequado.

Coeficiente de correlação de Spearman

- Para lidar com o primeiro problema apresentado, podemos usar o denominado coeficiente de correlação de classificação de Spearman (*Spearman's rank correlation coefficient*), ou simplesmente coeficiente de Spearman ou correlação de Spearman.
- Neste coeficiente, não nos preocupamos com os valores numéricos das variáveis, mas apenas com a sua ordem na classificação crescente de valores.
- Este coeficiente é calculado para uma amostragem especifica, com n amostras x_i e y_i.

Coeficiente de Spearman (cont)

- Começamos por ordenar os valores x_i entre si, e os valores y_i entre si.
- Definimos agora $r_{X,i}$ como o classificação do valor x_i na ordem das amostras de X, e $r_{Y,i}$ como a classificação de y_i na ordem das amostras de Y.
- O coeficiente de Spearman ρ_{XY} é o coeficiente de Pearson das classificações de X e Y:

$$\rho_{XY} = \frac{\operatorname{cov}(r_X, r_Y)}{\sigma_{r_X}\sigma_{r_Y}}.$$

Propriedades

- Se as variáveis forem não-correlacionadas, a ordem dos X e dos Y vai ser não-correlacionada, e portanto $\rho_{XY} = 0$.
- Se Y crescer monotonicamente com X, não importa de que forma (linear, quadrática, logaritmica, exponencial, etc.), então as classificações dos x_i e y_i vão ser as mesmas, e portanto os $r_{X,i}$ e $r_{Y,i}$ estão linearmente relacionados, resultando em $\rho_{XY} = 1$.
- Similarmente, se Y decrescer monotonicamente com X então $\rho_{XY} = -1$.

Entropia

- Dada uma partição de Ω, $\mathcal{A} = \{A_1, A_2, \ldots, A_m\}$ e usando $p_i = P(A_i)$, definimos:
$$H(\mathcal{A}) = -\sum_{i=1}^{m} p_i \log_2 p_i.$$
- Esta função é denominada a entropia da partição $\mathcal{A}$.
- A base do logaritmo usada na definição pode ser qualquer, resultando apenas em um fator multiplicativo de diferença, que indica a unidade usada. No caso da base 2, a unidade da entropia é bits.
- Por essa definição, quando um evento é de probabilidade 0, precisamos avaliar $\log_2 p$ que envolve uma divergência, mas como ela aparece multiplicada por p, o resultado é que $p \log_2 p = 0$ (verifique por limite), e esse evento não contribui para a entropia.

Exemplos

- Considere uma moeda justa e a partição $\mathcal{A} = \{\{\text{CARA}\}, \{\text{COROA}\}\}$. Neste caso $p_{\text{CARA}} = p_{\text{COROA}} = \frac{1}{2}$ e portanto

$$H(\mathcal{A}) = -2\tfrac{1}{2}\log_2 \tfrac{1}{2} = 1,$$

  isto é, esta partição tem 1 bit de entropia.

- Considere um dado justo e a partição $\mathcal{A} = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \}$. Neste caso, $p_i = \frac{1}{6}$ e

$$H(\mathcal{A}) = -\sum_{i=1}^{6} \tfrac{1}{6}\log_2 \tfrac{1}{6} = -6\tfrac{1}{6}\log_2 \tfrac{1}{6} = \log_2 6 \approx 2.58,$$

  isto é, esta partição tem pouco mais de 2 bits de entropia.

Exemplos (cont)

- Considere novamente um dado justo, mas com a partição $\mathcal{A} = \{\{1,3,5\},\{2,4,6\}\}$. Neste caso, $p_1 = p_2 = \frac{1}{2}$ e portanto

$$H(\mathcal{A}) = 1.$$

- Considere ainda um dado justo agora com a partição $\mathcal{A} = \{\{1\},\{2,3\},\{4,5,6\}\}$. Neste caso, $p_1 = \frac{1}{6}$, $p_2 = \frac{1}{3}$ e $p_3 = \frac{1}{2}$, resultando em

$$\begin{aligned} H(\mathcal{A}) &= -\tfrac{1}{6}\log_2\tfrac{1}{6} - \tfrac{1}{3}\log_2\tfrac{1}{3} - \tfrac{1}{2}\log_2\tfrac{1}{2} \\ &= \tfrac{1}{6}\log_2 6 + \tfrac{1}{3}\log_2 3 + \tfrac{1}{2} \\ &\approx 1.46. \end{aligned}$$