Pares Aleatórios

Introdução às Probabilidades e Estatística

Ciências ULisboa

Os pares aleatórios de variáveis aleatórias são uma ferramenta poderosa para entender como diferentes fenómenos estão interligados. Imagine ter dados sobre a quantidade de chuva e o nível de água dos rios em determinada região. Isoladamente, cada variável pode nos fornecer informações úteis, mas ao analisá-las juntas como um par aleatório, podemos obter uma compreensão mais profunda das suas interações e prever com maior precisão o impacto da precipitação nos níveis dos rios. Em termos práticos, a análise de pares aleatórios permite identificar correlações, prever resultados conjuntos e ajustar modelos matemáticos para refletir a realidade com maior precisão.

Definição (informal) de Par Aleatório

Considere-se uma experiência aleatória e o seu espaço de resultados Ω. Um par aleatório (X, Y) é uma função de $\Omega \to \mathbb{R}^2$.

(X, Y) é um par (ou vetor) aleatório:

- **discreto**: se X e Y forem variáveis aleatórias discretas;
- **contínuo**: se X e Y forem variáveis aleatórias contínuas.

Nota: O contradomínio do par aleatório (X, Y), i.e., o conjunto de valores possíveis de (X, Y), será representado por $\mathbb{R}_{X,Y}$ ($\mathbb{R}_{X,Y} \subseteq \mathbb{R}^2$).

Obs: Neste curso vamos trabalhar somente com pares aleatórios discretos.

Pares aleatórios (cont.)

Definição (Função (massa) de probabilidade conjunta)

Sejam X e Y duas variáveis aleatórias discretas. Então, a **função (massa) de probabilidade conjunta** do par aleatório discreto (X, Y) é definida por

$$f_{X,Y}(x, y) = P(X = x, Y = y), \quad \forall (x, y) \in \mathbb{R}^2$$

e satisfaz

1. $f_{X,Y}(x, y) \geq 0, \quad \forall (x, y) \in \mathbb{R}^2$
2. $\sum_x \sum_y f_{X,Y}(x, y) = 1$

Pares aleatórios (cont.)

A função massa de probabilidade conjunta em tabela de dupla entrada

$X \backslash Y$	y_1	y_2	$\cdots$	y_j	$\cdots$	y_m
x_1	p_{11}	p_{12}	$\cdots$	p_{1j}	$\cdots$	p_{1m}
x_2	p_{21}	p_{22}	$\cdots$	p_{2j}	$\cdots$	p_{2m}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
x_r	p_{r1}	p_{r2}	$\cdots$	p_{rj}	$\cdots$	p_{rm}
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$	$\ddots$	$\vdots$
x_n	p_{n1}	p_{n2}	$\cdots$	p_{nj}	$\cdots$	p_{nm}

- $p_{ij} = P(X = x_i, Y = y_j), \quad i = 1, \ldots, n, \ j = 1, \ldots, m$
- $p_{ij} \geq 0$
- $\sum_i \sum_j p_{ij} = 1$

Pares aleatórios (cont.)

Definição (Função de distribuição conjunta (caso discreto))

A **função de distribuição conjunta** do par aleatório discreto (X, Y) pode ser obtida à custa da função (massa) de probabilidade conjunta:

$$\begin{aligned} F_{X,Y}(x, y) &= P(X \leq x, Y \leq y) \\ &= \sum_{x_i \leq x} \sum_{y_j \leq y} P(X = x_i, Y = y_j), \quad \forall (x, y) \in \mathbb{R}^2. \end{aligned}$$

Pares aleatórios (cont.)

Propriedades da função de distribuição conjunta

1. $0 \leq F_{X,Y}(x,y) \leq 1, \quad \forall (x,y) \in \mathbb{R}^2$
2. $F_{X,Y}(x+\Delta_x, y+\Delta_y) \geq F_{X,Y}(x,y), \quad \forall \Delta_x, \Delta_y \geq 0$
3. $F_{X,Y}(+\infty,+\infty) = \lim_{x,y \to +\infty} F_{X,Y}(x,y) = 1$
4. $F_{X,Y}(-\infty,y) = \lim_{x \to -\infty} F_{X,Y}(x,y) = 0$ e
   $F_{X,Y}(x,-\infty) = \lim_{y \to -\infty} F_{X,Y}(x,y) = 0$
5. $F_{X,Y}(+\infty,y) = \lim_{x \to +\infty} F_{X,Y}(x,y) = F_Y(y)$ e
   $F_{X,Y}(x,+\infty) = \lim_{y \to +\infty} F_{X,Y}(x,y) = F_X(x)$.

Ciências ULisboa

Pares aleatórios (cont.)

Exemplo 1: Considere um teste de escolha múltipla com três questões, cujas respostas são dadas de forma independente. A resposta a cada questão pode estar correta (C) com probabilidade $P(C) = 0.5$ ou incorreta $\overline{C}$ com probabilidade $P(\overline{C}) = 0.5$. Considere, agora, o par aleatório (X, Y), onde X e Y representam o número de respostas corretas e incorretas neste teste, respetivamente.

(a) Confirme que o conjunto de valores possíveis de (X, Y) é $\{(0, 3), (1, 2), (2, 1), (3, 0)\}$.

(b) Defina a f.p. conjunta de (X, Y).

Resolução:

Par aleatório

X = número de respostas corretas neste teste
Y = número de respostas incorretas neste mesmo teste

Contradomínio de (X, Y)

Eventos elementares	X	Y
$\bar{C}\bar{C}\bar{C} = \bar{C}_1 \cap \bar{C}_2 \cap \bar{C}_3$	0	3
$C\bar{C}\bar{C}$	1	2
$\bar{C}C\bar{C}$	1	2
$\bar{C}\bar{C}C$	1	2
$CC\bar{C}$	2	1
$C\bar{C}C$	2	1
$\bar{C}CC$	2	1
CCC	3	0

Portanto, $\mathbb{R}_{X,Y} = \{(0,3), (1,2), (2,1), (3,0)\}$

Pares aleatórios (cont.)

F.p. conjunta de (X, Y)

$$\begin{aligned} P(X=0, Y=3) &= P(\bar{C}\bar{C}\bar{C}) \\ &= P(X=0) \\ &= \frac{1}{8} \end{aligned}$$

$$\begin{aligned} P(X=1, Y=2) &= P(\bar{C}\bar{C}C) + P(\bar{C}C\bar{C}) + P(C\bar{C}\bar{C}) \\ &= P(X=1) \\ &= \frac{3}{8} \end{aligned}$$

$$\begin{aligned} P(X=2, Y=1) &= P(\bar{C}CC) + P(C\bar{C}C) + P(CC\bar{C}) \\ &= P(X=2) \\ &= \frac{3}{8} \end{aligned}$$

Pares aleatórios (cont.)

$$\begin{aligned} P(X=3, Y=0) &= P(CCC) \\ &= P(X=3) \\ &= \frac{1}{8} \end{aligned}$$

Portanto,

$$P(X=x, Y=y) = \begin{cases} \frac{1}{8}, & (x,y) = (0,3), (3,0) \\ \frac{3}{8}, & (x,y) = (1,2), (2,1) \\ 0, & \text{outros pares } (x,y) \end{cases}$$

X \ Y	0	1	2	3
0	0	0	0	$\frac{1}{8}$
1	0	0	$\frac{3}{8}$	0
2	0	$\frac{3}{8}$	0	0
3	$\frac{1}{8}$	0	0	0

Definição (Funções de probabilidade marginais de X e Y)

A **f.p. marginal de X** define-se à custa da f.p. conjunta do par aleatório (X, Y):

$$f_X(x) = P(X = x) = \sum_y f_{X,Y}(x, y) = \sum_y P(X = x, Y = y), \quad \forall x \in \mathbb{R}.$$

A **f.p. marginal de Y** obtém-se de modo análogo:

$$f_Y(y) = P(Y = y) = \sum_x f_{X,Y}(x, y) = \sum_x P(X = x, Y = y), \quad \forall y \in \mathbb{R}.$$

$X \backslash Y$	y_1	y_2	$\ldots$	y_m	**Total**
x_1	$P(X = x_1, Y = y_1)$	$P(X = x_1, Y = y_2)$	$\ldots$	$P(X = x_1, Y = y_m)$	$P(X = x_1)$
x_2	$P(X = x_2, Y = y_1)$	$P(X = x_2, Y = y_2)$	$\ldots$	$P(X = x_2, Y = y_m)$	$P(X = x_2)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_n	$P(X = x_n, Y = y_1)$	$P(X = x_n, Y = y_2)$	$\ldots$	$P(X = x_n, Y = y_m)$	$P(X = x_n)$
Total	$P(Y = y_1)$	$P(Y = y_2)$	$\ldots$	$P(Y = y_m)$	**1**

Definição (Funções de distribuição marginais de X e Y)

As **f.d. marginais de X e de Y** podem obter-se à custa das respetivas f.p. marginais, ou então à f.p. conjunta ou à f.d. conjunta:

$$\begin{aligned} F_X(x) &= P(X \leq x) = \sum_{x_i \leq x} P(X = x_i) = \sum_{x_i \leq x} \left[\sum_{y} P(X = x_i, Y = y) \right] \\ &= F_{X,Y}(x, +\infty), \quad \forall x \in \mathbb{R}; \end{aligned}$$

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = \sum_{y_i \leq y} P(Y = y_i) = \sum_{y_i \leq y} \left[\sum_{x} P(X = x, Y = y_i) \right] \\ &= F_{X,Y}(+\infty, y), \quad \forall y \in \mathbb{R}. \end{aligned}$$

Exemplo 2: Seja X (resp. Y) o lucro do jogador A (resp. B) num dado jogo. As variáveis aleatórias X e Y são identicamente distribuídas e o par aleatório (X, Y) possui f.m.p. conjunta:

X\Y	-2	0	2
-2	0	1/6	1/6
0	1/6	0	1/6
2	1/6	1/6	0

Determine o valor esperado e a variância do lucro de cada um dos jogadores no jogo.

Resolução:

Par aleatório

$X =$ lucro do jogador A
$Y =$ lucro do jogador B

F.p. conjunta e f.p. marginais

$X \backslash Y$	-2	0	2	$P(X=x)$
-2	0	1/6	1/6	1/3
0	1/6	0	1/6	1/3
2	1/6	1/6	0	1/3
$P(Y=y)$	1/3	1/3	1/3	1

Valor esperado comum de X e Y

$$\begin{aligned} E(Y) = E(X) &= \sum_x x \times P(X=x) \\ &= [-2+0+2] \times \frac{1}{3} \\ &= 0 \end{aligned}$$

Variância comum de X e Y

$$\begin{aligned} V(Y) &= V(X) = E(X^2) - E(X)^2 \\ &= E(X^2) \\ &= \sum_x x^2 \times P(X=x) = \frac{8}{3} \end{aligned}$$

Pares aleatórios (cont.)

Definição (Funções massa de probabilidade condicionais)

A f.m.p. condicional de X dado $Y = y$ é

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{f_{X,Y}(x,y)}{f_Y(y)}, \quad \text{se} \quad P(Y = y) > 0.$$

A f.m.p. condicional de Y dado $X = x$ é

$$P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)} = \frac{f_{X,Y}(x,y)}{f_X(x)}, \quad \text{se} \quad P(X = x) > 0.$$

Pares aleatórios (cont.)

Definição (Valores esperados e variâncias condicionais)

Valor esperado de $X|Y = y$

$$E(X|Y = y) = \sum_x x\, P(X = x|Y = y), \quad \text{se} \quad P(Y = y) > 0.$$

Valor esperado de $Y|X = x$

$$E(Y|X = x) = \sum_y y\, P(Y = y|X = x), \quad \text{se} \quad P(X = x) > 0.$$

Definição (Independência)

As v.a. X e Y dizem-se **independentes** se, para todos os x e y, for válido

$$P(X = x, Y = y) = P(X = x)P(Y = y),$$

ou

$$f_{X,Y}(x, y) = f_X(x)f_Y(y),$$

ou

$$F_{X,Y}(x, y) = F_X(x)F_Y(y).$$

Ciências ULisboa

Exemplo 3: Considere as variáveis aleatórias X e Y representando, respetivamente, o número de minerais encontrados numa amostra de solo e o número de anos de experiência formal de geólogos que trabalham em campo num determinado projeto. A tabela abaixo fornece a correspondente função massa de probabilidade conjunta:

	Número de minerais (X)			
Experiência (Y)	0 minerais	1 mineral	2 minerais	3 minerais
5 anos	0.10	0.10	0.05	0.02
10 anos	0.12	0.15	0.07	0.03
15 anos	0.05	0.18	0.08	0.05

Supondo que um geólogo é selecionado aleatoriamente:

(a) Qual é a probabilidade de ele identificar 2 minerais e ter 10 anos de experiência?

(b) Qual é a probabilidade de ele ter 10 ou mais anos de experiência?

(c) Calcule a probabilidade de ele identificar menos de 3 minerais.

(d) Calcule a probabilidade de ele identificar mais de 2 minerais se tiver 15 anos de experiência.

(e) Verifique se as variáveis X e Y (número de minerais identificados e anos de experiência) são independentes.

Par aleatório (X, Y)
X = "Número de minerais encontrados numa amostra de solo"
Y = "Número de anos de experiência formal"

(a) Probabilidade de identificar 2 minerais e ter 10 anos de experiência:

$$P(X = 2, Y = 10) = f_{X,Y}(2, 10) = 0.07$$

(b) Probabilidade de ter 10 ou mais anos de experiência:

		Número de minerais (X)			
Experiência (Y)	0 minerais	1 mineral	2 minerais	3 minerais	$P(Y = y)$
5 anos	0.10	0.10	0.05	0.02	0.27
10 anos	0.12	0.15	0.07	0.03	0.37
15 anos	0.05	0.18	0.08	0.05	0.36
$P(X = x)$	0.27	0.43	0.2	0.1	1

$$P(Y \geq 10) = P(Y = 10) + P(Y = 15) = 0.37 + 0.36 = 0.73$$

(c) Probabilidade de ele identificar menos de 3 minerais.

$P(X < 3) = 1 - P(X = 3) = 1 - 0.10 = 0.9$

(d) Probabilidade de identificar mais de 2 minerais se tiver 15 anos de experiência.

$P(X > 2|Y = 15) = \frac{P(X>2, Y=15)}{P(Y=15)} = \frac{f_{X,Y}(3,15)}{f_Y(15)} = \frac{0.05}{0.36} = 0.138$

Se X e Y fossem independentes, teríamos $f_{X,Y}(x, y) = f_X(x) f_Y(y)$, para todo $(x, y) \in \mathbb{R}^2$. Porém,

$$f_{X,Y}(0, 5) = 0.1 \neq 0.0729 = f_X(0) f_Y(5),$$

pelo que se conclui que X e Y são dependentes.

Pares aleatórios (cont.)

Definição (Covariância)

Sejam X e Y duas v.a. discretas, a **covariância** entre X e Y, denotada por $cov(X, Y)$ é definida por

$$\begin{aligned} cov(X, Y) &= E[(X - E(X))(Y - E(Y))] \\ &= E(XY) - E(X)E(Y) \\ &= \sum_x \sum_y x\, y \times P(X = x, Y = y) - \sum_x x\, P(X = x) \sum_y y\, P(Y = y) \end{aligned}$$

- A covariância mede o sentido da associação linear entre as variáveis.

Ciências ULisboa

Pares aleatórios (cont.)

Propriedades da covariância

Sejam X e Y v.a. e a, b números reais. Então:

1. $V(X+Y) = V(X) + V(Y) + 2cov(X,Y)$;
2. $V(X-Y) = V(X) + V(Y) - 2cov(X,Y)$;
3. Se X e Y forem independentes $\Rightarrow cov(X,Y) = 0$;
4. $cov(X,Y) = 0 \not\Rightarrow X$ e Y independentes;
5. $cov(X,Y) = cov(Y,X)$;
6. $cov(X,X) = V(X)$;
7. $cov(aX+b, Y) = a\,cov(X,Y)$;
8. $cov(X+Z, Y) = cov(X,Y) + cov(Z,Y)$.

Pares aleatórios (cont.)

Definição (Correlação)

Sejam X e Y duas v.a. discretas, a **correlação** entre X e Y, denotada por $corr(X, Y)$ é definida por

$$\rho_{X,Y} = corr(X, Y) = \frac{cov(X, Y)}{\sqrt{V(X)V(Y)}}.$$

- A correlação entre duas v.a. toma valores entre -1 e 1;
- A correlação quantifica a associação linear entre X e Y;
- Quanto mais próximo de 1 for o valor absoluto de $corr(X, Y)$ mais forte é a intensidade da correlação;
- Se $|corr(X, Y)| \approx 1$, podemos afirmar que a associação entre X e Y é praticamente linear;
- Se $corr(X, Y) > 0$, podemos afirmar que as v.a. X e Y tenderão a variar no mesmo sentido;
- Se $corr(X, Y) < 0$, podemos afirmar que as v.a. X e Y tenderão a variar em sentidos opostos.

Propriedades da correlação

Sejam X e Y v.a. e a, b números reis. Então:

1. X e Y independentes $\Rightarrow corr(X, Y) = 0$;
2. $corr(X, Y) = 0 \not\Rightarrow X$ e Y independentes;
3. $corr(X, Y) = corr(Y, X)$;
4. $corr(X, X) = 1$;
5. $corr(aX + b, Y) = \begin{cases} corr(X, Y), & \text{se} \quad a > 0 \\ -corr(X, Y), & \text{se} \quad a < 0; \end{cases}$
6. $-1 \leq corr(X, Y) \leq 1$, para qualquer par de v.a..

Exemplo 4: Considere as variáveis aleatórias X e Y representando, respetivamente, o número de minerais encontrados numa amostra de solo e o número de anos de experiência formal de geólogos que trabalham em campo num determinado projeto. A tabela abaixo fornece a correspondente função massa de probabilidade conjunta:

	Número de minerais (X)			
Experiência (Y)	0 minerais	1 mineral	2 minerais	3 minerais
5 anos	0.10	0.10	0.05	0.02
10 anos	0.12	0.15	0.07	0.03
15 anos	0.05	0.18	0.08	0.05

Determine e interprete a covariância e a correlação entre as variáveis X e Y.

Par aleatório (X, Y)
$X =$ "Número de minerais encontrados numa amostra de solo"
$Y =$ "Número de anos de experiência formal"

		Número de minerais (X)			
Experiência (Y)	0 minerais	1 mineral	2 minerais	3 minerais	$P(Y = y)$
5 anos	0.10	0.10	0.05	0.02	0.27
10 anos	0.12	0.15	0.07	0.03	0.37
15 anos	0.05	0.18	0.08	0.05	0.36
$P(X = x)$	0.27	0.43	0.2	0.1	1

Pares aleatórios (cont.)

$$E(X) = 0 \times (0.27) + 1 \times (0.43) + 2 \times (0.20) + 3 \times (0.10) = 1.13$$

$$E(Y) = 5 \times (0.27) + 10 \times (0.37) + 15 \times (0.36) = 10.45$$

$$E(X^2) = 0^2 \times (0.27) + 1^2 \times (0.43) + 2^2 \times (0.20) + 3^2 \times (0.10) = 2.13$$

$$E(Y^2) = 5^2 \times (0.27) + 10^2 \times (0.37) + 15^2 \times (0.36) = 124.75$$

$$V(X) = E(X^2) - E(X)^2 = 2.13 - 1.13^2 = 0.8531$$

$$V(Y) = E(Y^2) - E(Y)^2 = 124.75 - 10.45^2 = 15.5475$$

$$E(XY) = \sum_x \sum_y xy \times P(X = x, Y = y) = 12.45$$

$$\sigma_X = \sqrt{0.8531} = 0.9236341$$

$$\sigma_Y = \sqrt{15.5475} = 3.943032$$

Pares aleatórios (cont.)

A covariância é dada por:

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$$

Substituindo os valores calculados:

$$\text{Cov}(X, Y) = 12.45 - (1.13 \times 10.45) = 0.6415$$

- A covariância $\text{Cov}(X, Y) = 0.6415$ indica uma leve relação positiva entre o número de minerais encontrados e os anos de experiência.

Pares aleatórios (cont.)

A correlação é definida como:

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

Onde $\sigma_X = \sqrt{V(X)}$ e $\sigma_Y = \sqrt{V(Y)}$.
Substituindo os valores:

$$\rho_{X,Y} = \frac{0.6415}{(0.9236341 \times 3.943032)} = 0.1761434$$

- O coeficiente de correlação $\rho_{X,Y} = 0.176$ mostra que a relação entre X e Y é fraca.
- Portanto, apesar de haver uma tendência de que mais experiência pode levar a mais descobertas, a correlação é pequena.

Referências

- Manuel Cabral Morais (2020): Probabilidade e Estatística: Teoria, Exemplos & Exercícios. IST Press, 1a edição.

Ciências ULisboa