Aula 10: Vetores aleatórios

Disciplina: PGE950 - Probabilidade

Programa de Pós-graduação em Estatística da UFPE

Prof. Pablo M. Rodriguez

http://www.pablo-rodriguez.org

Universidade Federal de Pernambuco

CCEN Centro de Ciências Exatas e da Natureza

Conteúdo desta aula

- Vetores aleatórios discretos
- Vetores aleatórios continuos.

Vetores aleatórios discretos

X e Y variáveis aleatórias discretas com função de probabilidade conjunta:

$$p(x, y) := P(X = x, Y = y).$$

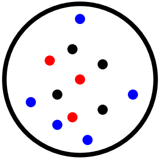
Lembre que,

$$p_X(x) = \sum_y p(x, y) \qquad \text{(função de probabilidade marginal de } X\text{),}$$

$$p_Y(y) = \sum_x p(x, y) \qquad \text{(função de probabilidade marginal de } Y\text{).}$$

Exemplo 10.1

Escolha três bolas ao acaso e sem reposição de uma urna contendo



3 bolas vermelhas, 4 bolas pretas e 5 bolas azuis

Se

$$X = \textit{número de bolas vermelhas escolhidas, e}$$

$$Y = \textit{número de bolas brancas escolhidas,}$$

então

$$p(i,j) = P(X = i,\, Y = j), \qquad \textit{para } i,j \in \{0,1,2,3\}.$$

... continuação do Exemplo 10.1. Por exemplo, como

$$p(0,0) = \frac{\#\text{possíveis retiradas que somente resultam em bolas azuis}}{\#\text{total de possíveis retiradas}}$$

então

$$p(0,0) = \binom{5}{3} \Big/ \binom{12}{3}.$$

Analogamente,

$$p(1,1) = \frac{\#\text{possíveis retiradas que resultam com uma bola de cada cor}}{\#\text{total de possíveis retiradas}}$$

então

$$p(1,1) = \left(\binom{3}{1} \cdot \binom{4}{1} \cdot \binom{5}{1} \right) \Big/ \binom{12}{3}.$$

... continuação do Exemplo 10.1. Em geral,

$$p(i,j) = \frac{\left(\binom{3}{i} \cdot \binom{4}{j} \cdot \binom{5}{3-(i+j)}\right)}{\binom{12}{3}}$$

se $i + j \leq 3$, e $p(i,j) = 0$ se $i + j > 3$.

... continuação do Exemplo 10.1. As probabilidades podem ser representadas na forma de uma tabela:

(i,j)	0	1	2	3	$P(X=i)$
0	$10/220$	$40/220$	$30/220$	$4/220$	$84/220$
1	$30/220$	$60/220$	$18/220$	0	$108/220$
2	$15/220$	$12/220$	0	0	$27/220$
3	$1/220$	0	0	0	$1/220$
$P(Y=j)$	$56/220$	$112/220$	$48/220$	$4/220$	

Exemplo 10.2

Suponha que uma urna contem três bolas vermelhas e cinco bolas brancas. Retire, sucessivamente e com reposição, três bolas da urna, e conside as variáveis aleatórias

$$X_i = \begin{cases} 1, & \text{se a } i\text{-ésima bola retirada for vermelha,} \\ 0, & \text{caso contrário.} \end{cases}$$

para $i \in \{1, 2, 3\}$. *A função de probabilidades conjunta de* X_1, X_2 *e* X_3 *é dada por*

$$p(i_1, i_2, i_3) = P(X_1 = i_1, X_2 = i_2, X_3 = i_3)$$

com $i_1, i_2, i_3 \in \{0, 1\}$. *Como as v.a. são independentes temos que*

$$p(i_1, i_2, i_3) = \prod_{j=1}^{3} P(X_j = i_j)$$

... e portanto

$$p(i_1, i_2, i_3) = (3/8)^{(i_1+i_2+i_3)} \cdot (5/8)^{3-(i_1+i_2+i_3)}.$$

Em geral

Em geral, dadas $X_1, X_2, \ldots, X_n$, variáveis aleatórias discretas, a função de probabilidades conjunta é dada por

$$p(x_1, x_2, x_3, \ldots, x_n) := P(X_1 = x_1, X_2 = x_2, X_3 = x_3, \ldots, X_n = x_n),$$

e a função de probabilidade marginal de X_i, para $i \in \{1, 2, \ldots, n\}$, é obtida fazendo

$$p_{X_i}(x_i) := P(X_i = x_i) = \underbrace{\sum_{x_1} \sum_{x_2} \cdots \sum_{x_n}}_{x_j : j \neq i} p(x_1, x_2, \ldots, x_i, \ldots, x_n)$$

Distribuição Multinomial

Considere um experimento que satisfaz as seguintes condições:

1. O experimento consiste em n ensaios idênticos;
2. cada ensaio pode levar a qualquer um de r resultados possíveis;
3. as respectivas probabilidades dos r resultados são dadas por

$$p_1, p_2, \ldots p_r$$

   e sãs as mesmas nos n ensaios. Note que $\sum_{i=1}^{r} p_i = 1$.
4. Os resultados dos n ensaios são independentes.

Observação!

Como exemplo da situação: lançamentos de um dado, coleção de figurinhas ...

Distribuição Multinomial

Seja

$$X_i = \# \text{ de ensaios que levam ao } i-\text{ésimo resultado},$$

para $i \in \{1, 2, \ldots, r\}$. Então, a função de probabilidades do vetor

$$\mathbf{X} := (X_1, X_2, \ldots, X_r)$$

é dada por

$$p(i_1, i_2, \ldots, i_r) = \frac{n!}{i_1! \cdot i_2! \cdots i_r!} \cdot p_1^{i_1} \cdot p_2^{i_2} \cdots p_r^{i_r},$$

sempre que $\sum_{j=1}^{r} i_j = n$.

No cálculo de $p(i_1, i_2, \ldots, i_r)$ usamos que existem

$$\binom{n}{i_1} \cdot \binom{n-i_1}{i_2} \cdots \binom{n-i_1-i_2-\ldots-i_{r-1}}{i_r} = \frac{n!}{i_1! \cdot i_2! \cdots i_r!}$$

configurações possíveis dos resultados desejados.

Distribuição Multinomial

O vetor $\mathbf{X} = (X_1, X_2, \ldots, X_r)$ com as características mencionadas, n ensaios, r resultados por ensaio e probabilidades $p_1, p_2, \ldots, p_r$, tem distribuição multinomial.
Notação:

$$\mathbf{X} \sim M(n, r, p_1, p_2, \ldots, p_r),$$

ou por

$$\mathbf{X} \sim M_r(n, p_1, p_2, \ldots, p_r).$$

Observação!

- *Note que* $X_i \sim B(n, p_i)$, *para* $i \in \{1, 2, \ldots, r\}$.
- *Quando* $r = 2$, *temos uma distribuição binomial.*

Exemplo 10.3

Acidentes de trânsito. *A seguinte tabela mostra a distribuição de acidentes de trânsito, por dias da semana, envolvendo pedestres na Dinamarca em 1981:*

Dia da semana	*seg*	*ter*	*qua*	*qui*	*sex*	*sab*	*dom*	*TOTAL*
# de acidentes	279	256	230	304	330	210	130	1739

- *Se os acidentes ocorrem independentemente um do outro, temos que o vetor de acidentes observados* $(X_1, X_2, \ldots, X_7)$ *segue uma distribuição multinomial.*

Pergunta de interesse: um acidente escolhido ao acaso tem a mesma chance de ter ocorrido em qualquer um dos 7 dias?

Referência!

Exemplo 2.1 (pág. 18) do livro Introduction to the Statistical Analysis of Categorical Data de Erling B. Andersen, Springer, 1997.

Vetores aleatórios continuos

Se X e Y são variáveis aleatórias continuas, chama-se função densidade conjunta de X e Y à função $f(x,y) \geq 0$, definida para $x, y \in \mathbb{R}$, satisfazendo

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

e tal que para todo $\mathcal{C} \in \mathcal{B}^2$ vale

$$P\left((X,Y) \in \mathcal{C}\right) = \iint\limits_{\mathcal{C}} f(x,y)\,dx\,dy.$$

Lembre que:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy \qquad \text{(densidade marginal de } X\text{)},$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx \qquad \text{(densidade marginal de } Y\text{)}.$$

Exemplo 10.4

Sejam X e Y com função densidade conjunta dada por

$$f(x,y) = \begin{cases} \frac{3}{80}(x^2 + x\,y), & \textit{se } 0 \leq x \leq 2, \quad 0 \leq y \leq 4, \\ \\ 0, & \textit{caso contrário.} \end{cases}$$

Calcule:

- *as densidades marginais de X e Y;*
- $P(X + Y \leq 3)$.

... continuação do Exemplo 10.4. Notemos que para x fixado, com $0 \leq x \leq 2$, os valores de y tais que $f(x, y) > 0$ estão no intervalo $[0, 4]$. Então,

$$f_X(x) = \int_0^4 \frac{3}{80}(x^2 + x\,y)\,dy = \frac{3}{20}x^2 + \frac{3}{10}x.$$

Da mesma forma, para y fixado, com $0 \leq y \leq 4$, temos que

$$f_Y(y) = \int_0^2 \frac{3}{80}(x^2 + x\,y)\,dx = \frac{1}{10} + \frac{3}{40}y.$$

... continuação do Exemplo 10.4. Finalmente, as densidades marginais são

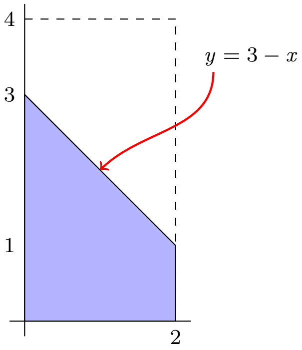
$$f_X(x) = \begin{cases} \frac{3}{20}x^2 + \frac{3}{10}x, & 0 \leq x \leq 2, \\ \\ 0, & \text{caso contrário.} \end{cases}$$

e

$$f_Y(y) = \begin{cases} \frac{1}{10} + \frac{3}{40}y, & 0 \leq y \leq 4, \\ \\ 0, & \text{caso contrário.} \end{cases}$$

... continuação do Exemplo 10.4. Para calcular $P(X + Y \leq 3)$ trabalhamos na região

$$\mathcal{C} = \{(x, y) \in \mathbb{R}^2 : 0 \leq x \leq 2, 0 \leq y \leq 4, x + y \leq 3\}.$$



... continuação do Exemplo 10.4. Então,

$$P(X + Y \leq 3) = \int_0^2 \left(\int_0^{3-x} \frac{3}{80}(x^2 + x\,y)\,dy \right)\,dx = \frac{21}{80}$$

(verificar!)

Em geral

Dadas $X_1, \ldots, X_n$ variáveis aleatórias continuas, a função densidade conjunta destas v.a. é a função não-negativa $f(x_1, x_2, \ldots, x_n)$ tal que para todo $\mathcal{C} \in \mathcal{B}^n$

$$P((X_1, X_2, \ldots, X_n) \in \mathcal{C}) = \underset{\mathcal{C}}{\int \cdots \int} f(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2\, \ldots\, dx_n.$$

A densidade marginal de X_i, para $i \in \{1, 2, \ldots, n\}$, é obtida fazendo

$$f_{X_i}(x_i) = \underset{\mathbb{R}^{n-1}}{\int \cdots \int} f(x_1, x_2, \ldots, x_n)\, \underbrace{dx_1\, dx_2\, \ldots\, dx_n}_{dx_j : j \neq i}.$$

Distribuição normal bi-variada

O vetor aleatório (X, Y) tem distribuição normal bi-variada se a função densidade conjunta de X e Y dada por

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2-\frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]\right\}}$$

para $-\infty < x, y < \infty$, onde $\mu_X, \mu_Y, \sigma_X, \sigma_Y$ e ρ são constantes tais que

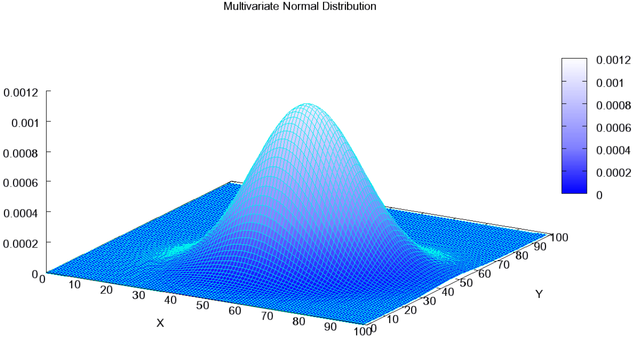
$$\sigma_X > 0, \quad \sigma_Y > 0$$

e

$$-1 < \rho < 1.$$

UNIVERSIDADE FEDERAL DE PERNAMBUCO CCEN CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA

Distribuição normal bi-variada



Fonte: Wikipedia!

Distribuição normal bi-variada: marginais

- $X \sim \mathcal{N}\left(\mu_X, \sigma_X^2\right)$.
- $Y \sim \mathcal{N}\left(\mu_Y, \sigma_Y^2\right)$.

Independência

As variáveis aleatórias X e Y são independentes se, para quaisquer $A, B \subset \mathbb{R}$, vale que

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

Isto é, X e Y são independentes se para quaisquer $A, B \subset \mathcal{B}$ os eventos aleatórios

$$\mathcal{E}_A := \{X \in A\}$$

e

$$\mathcal{E}_B := \{Y \in B\}$$

são independentes.

Independência

As variáveis aleatórias X e Y são independentes se, e somente se,

- $F(a, b) = F_X(a) \cdot F_Y(b)$, para todo a, b;
- $p(x, y) = p_X(x) \cdot p_Y(y)$ para todo x, y (caso discreto);
- $f(x, y) = f_X(x) \cdot f_Y(y)$ para todo x, y (caso continuo).

Bom estudo!