Variáveis aleatórias bidimensionais

Prof. Wagner Hugo Bonat

Departamento de Estatística
Universidade Federal do Paraná

Motivação

Múltiplas características

- Nosso estudo até aqui se restringiu a apenas uma variável de interesse.
- Interesse no comportamento **conjunto** de mais de uma variável de interesse.
- Explorar relações e determinar a existência de **associação** entre as variáveis de interesse.

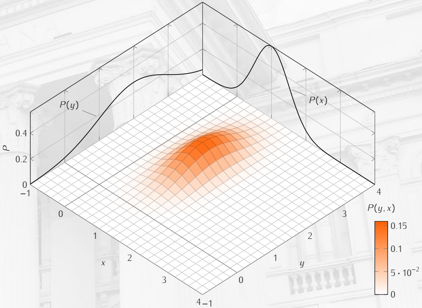

Figura 1. Exemplo de distribuição conjunta contínua.

Exemplo: Número de vestibulares

Uma amostra de 20 alunos do primeiro ano de uma faculdade foi escolhida. Perguntou-se aos alunos se trabalhavam, variável que foi representada por X, e o número de vestibulares prestados, variável representada por Y. Os dados obtidos estão na tabela abaixo.

X	Y	X	Y
Não	1	Não	2
Sim	1	Não	2
Não	2	Sim	1
Não	1	Não	3
Não	1	Sim	2
Sim	2	Não	2
Sim	3	Não	2
Não	1	Não	1
Sim	1	Sim	3
Sim	1	Não	2

Exemplo: Número de vestibulares (cont.)

- Comportamento conjunto.

X/Y	1	2	3	Total
Não	5	6	1	12
Sim	4	2	2	8
Total	9	8	3	20

- Comportamento marginal de X.

X	Não	Sim	Total
Freq.	12	8	20

- Comportamento marginal de Y.

Y	1	2	3	Total
Freq.	9	8	3	20

Exemplo: Pacientes HIV positivo

Um estudo envolveu 345 pacientes HIV positivos, acompanhados durante um ano, pelo setor de doenças infecciosas de um grande hospital público. Os dados apresentados contêm as ocorrências relacionadas às variáveis número de internações (I) e número de crises com infecções oportunistas (C).

I/C	0	1	2	3	4
0	84	21	8	2	0
1	20	59	35	14	2
2	6	11	43	28	12

Obtenha as marginais de I e C.

Exemplo: Pacientes HIV positivo (cont.)

- Marginal de I.

I	0	1	2	Total
Freq.	115	130	100	345

- Marginal de C.

C	0	1	2	3	4	Total
Freq.	110	91	86	44	14	345

Distribuição de probabilidade conjunta e marginal

Distribuição conjunta e marginal

- Sejam X e Y duas v.a.'s discretas originárias do mesmo fenômeno aleatório, com valores atribuídos a partir do mesmo espaço amostral.
- A **função de probabilidade conjunta** é definida, para todos os possíveis pares de valores (X,Y), da seguinte forma:

$$\mathrm{p}(x,y) = \mathrm{P}[(X = x) \cap (Y = y)] = \mathrm{P}(X = x, Y = y).$$

- A função de probabilidade conjunta também pode ser chamada de **distribuição conjunta** ou simplesmente **conjunta** das v.a.'s.
- Definição: As **distribuição marginal** de X e Y são dadas por

$$\mathrm{p}(x) = \sum_{y} \mathrm{p}(x,y) \quad \text{e} \quad \mathrm{p}(y) = \sum_{x} \mathrm{p}(x,y).$$

Exemplo: Empresa de encomendas

Uma empresa atende encomendas de supermercados dividindo os pedidos em duas partes de modo a serem atendidos, de forma independente, pelas suas duas fábricas. Devido à grande demanda, pode haver atraso no cronograma de entrega, sendo que a fábrica 1 atrasa com probabilidade 0.1 e a 2 com 0.2. Sejam A_1 e A_2 os eventos correspondentes a ocorrência de atraso nas fábricas 1 e 2, respectivamente.

Para uma entrega, a indústria recebe 200 u.m., mas paga 20 para cada fábrica que atrasar. Considere que o supermercado que recebe a encomenda fez um índice relacionado à pontualidade de entrega. Este índice, atribuiu 10 pontos para cada entrega dentro do cronograma previsto. Denote por X o valor recebido pelo pedido e Y o índice obtido. Obtenha a distribuição conjunta de Y e X e as distribuições marginais de Y e X.

Exemplo: Empresa de encomendas (cont.)

- Espaço amostral : $\Omega = \{A_1A_2, A_1^cA_2, A_1A_2^c, A_1^cA_2^c\}$.

Eventos	Probabilidades	X	Y
A_1A_2	$0.1 \cdot 0.2$	160	0
$A_1A_2^c$	$0.1 \cdot 0.8$	180	10
$A_1^cA_2$	$0.9 \cdot 0.2$	180	10
$A_1^cA_2^c$	$0.9 \cdot 0.8$	200	20

- Distribuição conjunta.

X/Y	0	10	20
160	0.02	0	0
180	0	0.26	0
200	0	0	0.72

Exemplo: Poços e riachos

Uma região foi dividida em 10 sub-regiões. Em cada uma delas, foram observadas duas variáveis: número de poços artesianos (X) e número de riachos ou rios presentes na sub-região (Y). Os resultados são apresentados na tabela a seguir:

Sub-região	X	Y
1	0	1
2	0	2
3	0	1
4	0	0
5	1	1
6	2	0
7	1	0
8	2	1
9	2	2
10	0	2

- Construa a distribuição conjunta e marginais de X e Y.

Exemplo: Poços e riachos (cont.)

Consideramos que cada região tem a mesma probabilidade 1/10 de ser escolhida. Assim, a distribuição conjunta é:

X/Y	0	1	2
0	0.10	0.20	0.20
1	0.10	0.10	0.00
2	0.10	0.10	0.10

Para obter as marginais, efetuamos a soma nas linhas para obter a marginal de X, e nas colunas para obter a marginal de Y. Por exemplo, $\mathrm{P}(X = 0)$ é obtida através de:

$$\begin{aligned}\mathrm{P}(X = 0) &= \mathrm{P}(X = 0, Y = 0) + \mathrm{P}(X = 0, Y = 1) + \mathrm{P}(X = 0, Y = 2)\\ &= 0.1 + 0.2 + 0.2 = 0.5.\end{aligned}$$

Exemplo: Poços e riachos (cont.)

Repetindo os cálculos para todos os valores de X e Y, obtemos as marginais:

X/Y	0	1	2	$P(X = x)$
0	0.10	0.20	0.20	0.50
1	0.10	0.10	0.00	0.20
2	0.10	0.10	0.10	0.30
$P(Y = y)$	0.30	0.40	0.30	1.00

Distribuição condicional (caso discreto)

Probabilidade condicional para v.a.'s discretas

- A **probabilidade condicional** de $X = x$, dado que $Y = y$ ocorreu, é dada pela expressão:

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

- Duas v.a.'s discretas são **independentes**, se a ocorrência de qualquer valor de uma delas não altera a probabilidade de valores da outra. Em termos de probabilidade

$$P(X = x|Y = y) = P(X = x).$$

- Definição alternativa

$$P(X = x, Y = y) = P(X = x)P(Y = y), \quad \forall (x, y).$$

Exemplo: Salários de estagiários

- O Centro Acadêmico de uma faculdade de administração fez um levantamento da remuneração dos estágios dos alunos, em salários mínimos (Sal), com relação ao ano que estão cursando (Ano). As probabilidades de cada caso são apresentadas na próxima tabela, incluindo as distribuições marginais.

Sal/Ano	2	3	4	5	$P(\text{Sal} = x)$
2	2/25	2/25	1/25	0	5/25
3	2/25	5/25	2/25	2/25	11/25
4	1/25	2/25	2/25	4/25	9/25
$P(\text{Ano} = y)$	5/25	9/25	5/25	6/25	1

- Sal e Ano são independentes?

$$P(\text{Sal} = 3) \cap (\text{Ano} = 4) = 2/25 \neq P(\text{Sal} = 3)P(\text{Ano} = 4) = \frac{11}{25}\frac{5}{25} = \frac{11}{125}.$$

Não são independentes.

Função de densidade conjunta e marginal

Densidade conjunta e marginal

- Definição: A função $f(x,y)$ é a **função densidade conjunta** das v.a. contínuas X e Y se

1. $f(x,y) \geq 0, \quad$ para todo $\quad (x,y)$.
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) = 1$.

- $\mathrm{P}[(X,Y) \in A] = \int \int_A f(x,y)dxdy$, para qualquer região A no plano xy.
- Distribuições **marginais**

$$f(x) = \int_{-\infty}^{\infty} f(x,y)dy \quad \text{e} \quad f(y) = \int_{-\infty}^{\infty} f(x,y)dx.$$

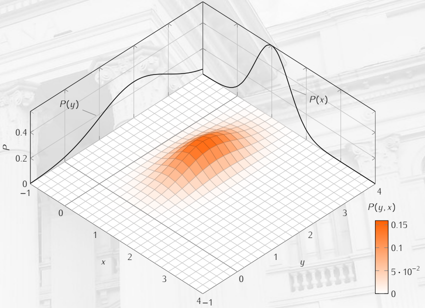

Figura 2. Ilustração fdp conjunta.

Exemplo: Produção

- Uma empresa atua com dois métodos para produzir um certo produto. Para um certo dia de operação, sejam X e Y, as proporções de produtos produzidos pelos métodos A e B, respectivamente. Suponha, que a distribuição conjunta de X e Y é dada por

$$f(x,y) = \begin{cases} \frac{2}{5}(2x+3y), & 0 \leq x \leq 1, 0 \leq y \leq 1, \\ 0 & \text{caso contrário.} \end{cases}$$

1. Verifique se $f(x,y)$ é uma distribuição de probabilidade.
2. Encontre as distribuições marginais de X e Y.
3. Encontre a probabilidade $P[(X,Y) \in A]$, onde $A = \{(x,y) | 0 < x < 0.5, 0.25 < y < 0.5\}$.

Exemplo: Produção (cont.)

Gráfico da função densidade probabilidade bidimensional.

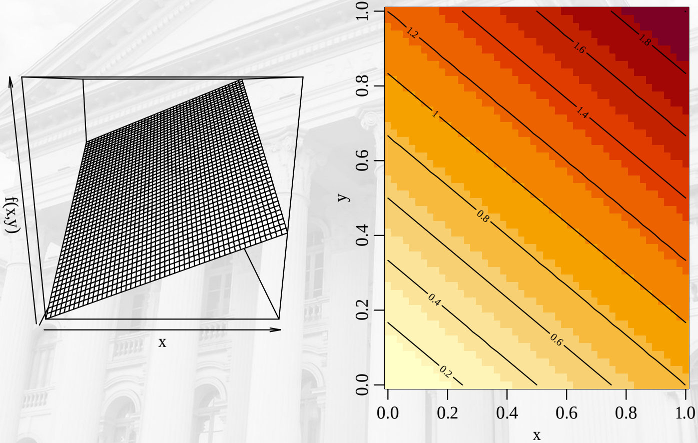

Figura 3. Função densidade probabilidade conjunta.

Exemplo: Produção (cont.)

- Condição (1) é trivial.
- Verificando condição (2)

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{2}{5}(2x+3y)dxdy = \int_0^1\int_0^1\left(\frac{4x}{5}+\frac{6y}{5}\right)dxdy$$

$$\int_0^1 \frac{6y}{5} + \int_0^1 \frac{4x}{5}dxdy = \int_0^1 \frac{6y}{5} + \frac{4}{5}\left[\frac{x^2}{2}\right]_0^1 dy = \int_0^1 \frac{6y}{5} + \frac{4}{10}dy$$

$$\frac{6}{5}\int_0^1 ydy + \frac{4}{10}\int_0^1 1dy = \frac{6}{5}\left[\frac{y^2}{2}\right]_0^1 + \frac{4}{10}[y]_0^1 = \frac{6}{10} + \frac{4}{10} = 1.$$

Exemplo: Produção (cont.)

- Obtendo as marginais

$$f(x) = \int_0^1 \frac{4x}{5} + \frac{6y}{5} dy = \left[\frac{4xy}{5} + \frac{6y^2}{10}\right]_0^1 = \frac{4x+3}{5},$$

para $0 \leq x \leq 1$, e 0 caso contrário.

$$f(y) = \int_0^1 \frac{4x}{5} + \frac{6y}{5} dx = \frac{2+6y}{5},$$

para $0 \leq y \leq 1$, e 0 caso contrário.

- Obtendo a probabilidade

$$P[(X,Y) \in A] = \int_0^{\frac{1}{2}} \int_{\frac{1}{4}}^{\frac{1}{2}} \frac{4x}{5} + \frac{6y}{5} dy dx = \frac{13}{160}.$$

Distribuição condicional

Distribuição condicional: caso discreto ou contínuo

- Sejam X e Y duas v.a. discretas ou contínuas. A **distribuição condicional** de uma v.a. Y dado $X = x$ é

$$f(y|x) = \frac{f(x,y)}{f(x)}, \quad \text{dado que} \quad f(x) > 0.$$

$$f(x|y) = \frac{f(x,y)}{f(y)}, \quad \text{dado que} \quad f(y) > 0.$$

- Probabilidades condicionais

$$P(a < X < b|Y = y) = \sum_{a<x<b} f(x|y).$$

$$P(a < X < b|Y = y) = \int_a^b f(x|y)dx.$$

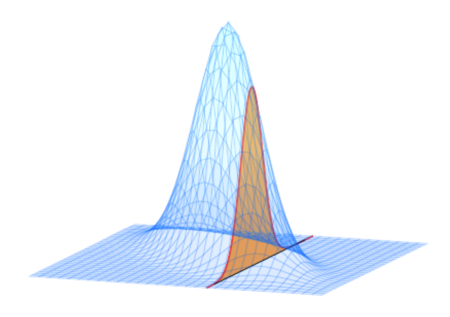

Figura 4. Distribuição condicional.

Independência estatística

Independência estatística

- Sejam X e Y duas v.a. discretas ou contínuas, com distribuição de probabilidade conjunta $f(x,y)$ e distribuições marginais $f(x)$ e $f(y)$. As v.a.'s X e Y são ditas **estatisticamente independentes** se e somente se

$$f(x,y) = f(x)f(y)$$

para todo par (x,y) em seus respectivos domínios.

Exemplo: Tempo de prateleira

- Suponha que o tempo de prateleira, em anos, de um certo produto perecível é uma variável aleatória com distribuição de probabilidade dada por

$$f(y) = \begin{cases} e^{-y}, & y > 0 \\ 0 & \text{caso contrário.} \end{cases}$$

Sejam Y_1, Y_2 e Y_3 o tempo de prateleira de três destes produtos selecionados independentemente. Encontre $\mathrm{P}(Y_1 < 2, 1 < Y_2 < 3, Y_3 > 2)$.

Exemplo: Tempo de prateleira (cont.)

- Usando a independência, tem-se

$$f(y_1, y_2, y_3) = f(y_1)f(y_2)f(y_3) = e^{-y_1}e^{-y_2}e^{-y_3} = e^{-y_1-y_2-y_3}$$

$$\mathrm{P}(Y_1 < 2, 1 < Y_2 < 3, Y_3 > 2) =$$
$$\int_2^{\infty}\int_1^3\int_0^2 e^{-y_1-y_2-y_3}dy_1dy_2dy_3 = 0.0372.$$

Pontos importantes

- Como construímos a distribuição de probabilidade para um fenômeno?
  - Por observação do fenômeno por longos períodos.
  - Natureza do fenômeno sugere uma certa forma de distribuição.
  - Em geral não sabemos qual distribuição de probabilidade pode ser uma boa escolha para um dado fenômeno aleatório.
- Existem muitas distribuições de probabilidade que podem ser avaliadas como uma boa aproximação para o fenômeno em estudo. Vamos ver as mais importantes:
  - Bernoulli, binomial e Poisson (caso discreto).
  - Normal, exponencial e Gama (caso contínuo).
- Mas antes precisamos entender como resumir a informação gerada pela distribuição de probabilidades.

Considerações finais

Considerações finais

- Variáveis aleatórias bidimensionais.
  - Distribuição conjunta.
  - Distribuição marginal.
  - Distribuição condicional.
  - Independência estatística.

Figura 5. Foto de Isabella Mendes no Pexels.