Universidade Federal do Rio Grande do Norte
Instituto Metrópole Digital
IMD0601 - Bioestatística

Distribuição multivariada

Prof. Dr. Tetsu Sakamoto
Instituto Metrópole Digital - UFRN
Sala A224, ramal 182
Email: tetsu@imd.ufrn.br

Baixe a aula (e os arquivos)

- Para aqueles que não clonaram o repositório:

```
> git clone https://github.com/tetsufmbio/IMD0601.git
```

- Para aqueles que já tem o repositório local:

```
> cd /path/to/IMD0601

> git pull
```

Revisão

Distribuição contínua univariada

- **Normal**
- **Gama**
- **Beta**

Funções no R relacionadas a distribuições comuns:

- **rnorm, rgamma, rbeta**
- **dnorm, dgamma, dbeta**
- **pnorm, pgamma, pbeta**
- **qnorm, qgamma, qbeta**

Distribuição multivariada

Probabilidade envolvendo mais de uma variável aleatório

Modelar fenômenos reais pode requer um modelo de distribuição com mais variáveis;

Estudos ambientais: temperatura, umidade, concentração de gases, ...;

Genômica: Cada posição do nucleotídeo pode ser uma variável aleatória.

Probabilidade condicional

Conceitos expandidos de probabilidade

Cálculo da probabilidade de um evento dado que algum outro evento a priori aconteceu e que temos a sua probabilidade.

$P(A|B)$ → Probabilidade do evento A dado evento B.

Probabilidade condicional

Exemplo:

Coloração das ervilhas de Mendel.

Característica cor: codificado por um gene com dois alelos (**Y** e **y**);

Y: codifica para fenótipo de cor **amarela**;

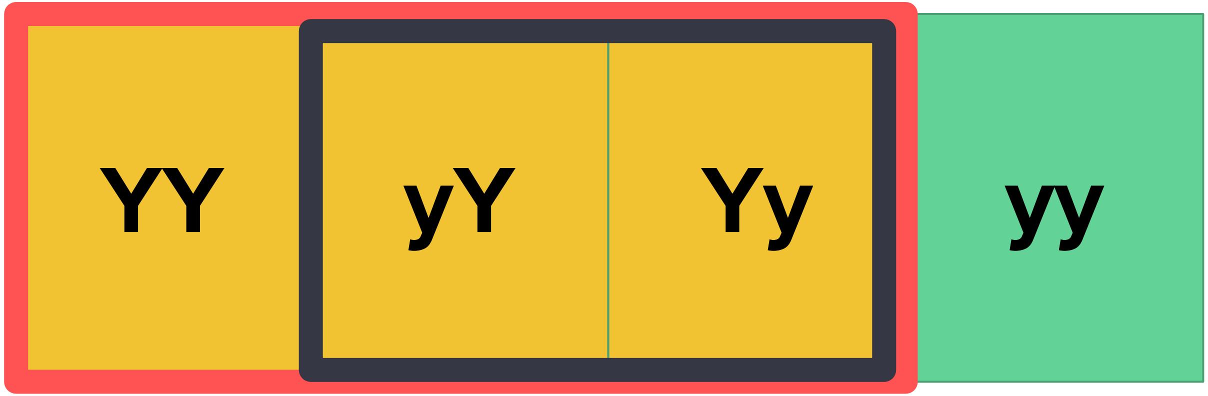
y: codifica para fenótipo de cor **verde**.

Suponha que temos uma ervilha de cor amarela.
Qual a probabilidade dele ser heterozigoto?

Genótipo	Fenótipo	Prob
YY	amarelo	¼
Yy, yY	amarelo	½
yy	verde	¼

Probabilidade condicional

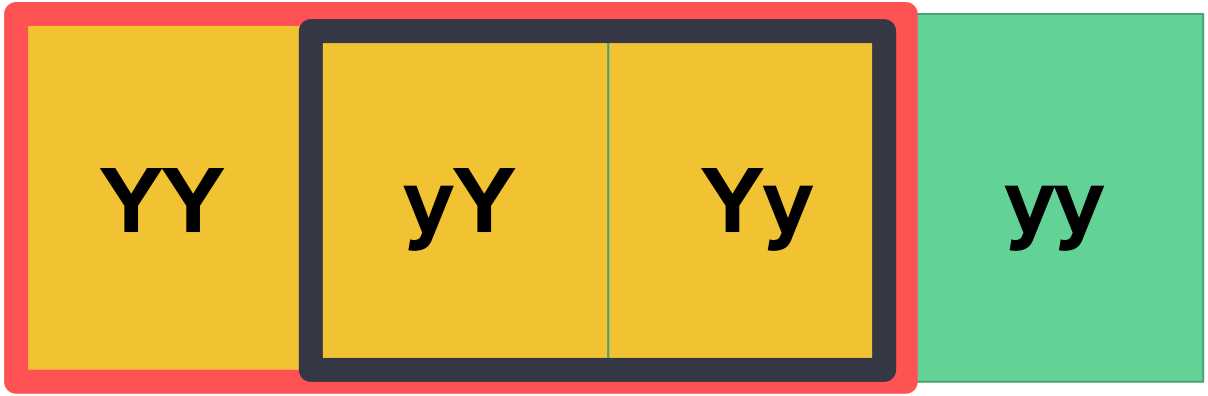
Suponha que temos uma ervilha de cor amarela. Qual a probabilidade dele ser heterozigoto?



P(yY ou Yy | amarelo) = 2/3

Probabilidade condicional

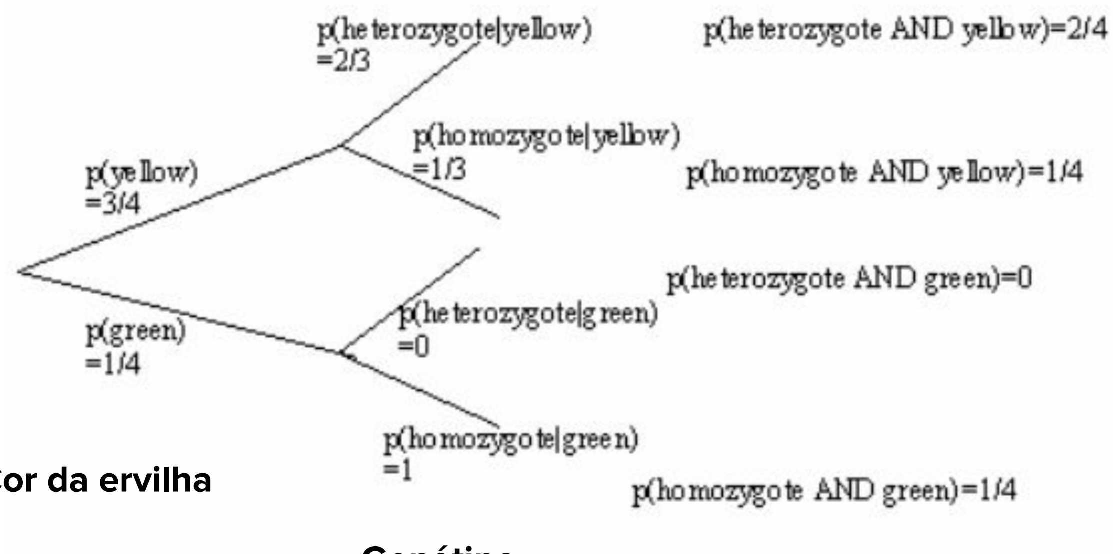
Suponha que temos uma ervilha de cor amarela. Qual a probabilidade dele ser heterozigoto?



P(amarelo) = 3/4 **P(yY ou Yy ∩ amarelo) = 2/4**

P(yY ou Yy | amarelo) = P(yY ou Yy ∩ amarelo) / P(amarelo) = 2/3

Probabilidade condicional



Cor da ervilha

Genótipo

Probabilidades em cada conjunto de ramos soma 1

Soma das interseções é igual a 1

Independência

Conceitos expandidos de probabilidade

Eventos independentes → Quando a ocorrência de um não altera a probabilidade da ocorrência do outro.

Independência entre eventos

Eventos independentes ➔ Quando a ocorrência de um não altera a probabilidade da ocorrência do outro.

Se A e B são eventos independentes, então:

$P(A|B) = P(A)$, e **$P(B|A) = P(B)$**.

Considerando que **$P(A|B) = P(A \cap B)/P(B)$**

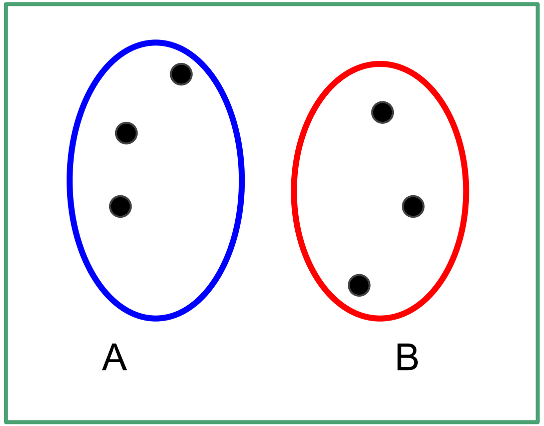
Podemos reescrever a fórmula:

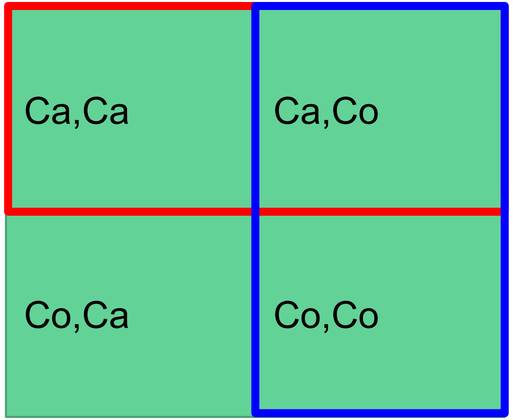
$P(A \cap B) = P(A)P(B)$

Mutuamente exclusivo ≠ independente

Eventos mutuamente exclusivo ➔ interseção é vazio;

Independente ➔ A ocorrência de um evento não interfere no outro;





Independência em mais de dois eventos

$$P(A \cap B \cap C) = P(A)*P(B)*P(C)$$

Se a expressão acima for verdadeira, então os pares de eventos também serão independentes entre eles.

Distribuição multivariada

Probabilidade envolvendo mais de uma variável aleatório

Variáveis:

- Discreta
- Contínua

Distribuição

- Conjunta;
- Marginal;
- Condicional.

Probabilidade e distribuição multivariada discreta

		Nucleotide at position 1			
		A	T	C	G
Nucleotide at position 2	A	$\boxed{0.2}$	0.1	0	0.1
	T	0	0.1	0.1	0.1
	C	0.1	0	0.1	0
	G	0	0.1	0	0

P(A1 ∩ A2)

Probabilidade e distribuição conjunta discreta

		Nucleotide at position 1			
		A	T	C	G
Nucleotide at position 2	A	0.2	0.1	0	0.1
	T	0	0.1	0.1	0.1
	C	0.1	0	0.1	0
	G	0	0.1	0	0

P(A1 ∩ A2)

Distribuição de probabilidade multivariada

Função massa de probabilidade

Tabela de probabilidades conjunta → Probabilidades devem somar 1

Probabilidade e distribuição conjunta

		X= Nucleotide at position 1			
Y=Nucleotide at position 2		A	T	C	G
	A	0.2	0.1	0	0.1
	T	0	0.1	0.1	0.1
	C	0.1	0	0.1	0
	G	0	0.1	0	0

$P(X=xi, Y=yi) = P(X=xi \cap Y=yi)$

Probabilidade e distribuição marginal

		Nucleotide at position 1				Marginal probabilities for rows
		A	T	C	G	
Nucleotide at position 2	A	0.2	0.1	0	0.1	0.4
	T	0	0.1	0.1	0.1	0.3
	C	0.1	0	0.1	0	0.2
	G	0	0.1	0	0	0.1
Marginal probabilities for columns		0.3	0.3	0.2	0.2	1

Probabilidade marginal $P(X=A)$

Lei da probabilidade total

Distribuição de probabilidade multivariada

Probabilidade e distribuição condicional

		Nucleotide at position 1 (X)				Marginal probabilities for rows
		A	T	C	G	
Nucleotide at position 2 (Y)	A	0.2	0.1	0	0.1	0.4
	T	0	0.1	0.1	0.1	0.3
	C	0.1	0	0.1	0	0.2
	G	0	0.1	0	0	0.1
Marginal probabilities for columns		0.3	0.3	0.2	0.2	1

P(X=A | Y=A) = ?

P(X=A ∩ Y=A) = 0,2

P(Y=A) = 0,4

P(X=A | Y=A) = 0,2/0,4

P(X=A | Y=A) = ½

P(X=T | Y=A) = ¼

P(X=C | Y=A) = 0

P(X=G | Y=A) = ¼

Distribuição condicional

Distribuição multivariada contínua

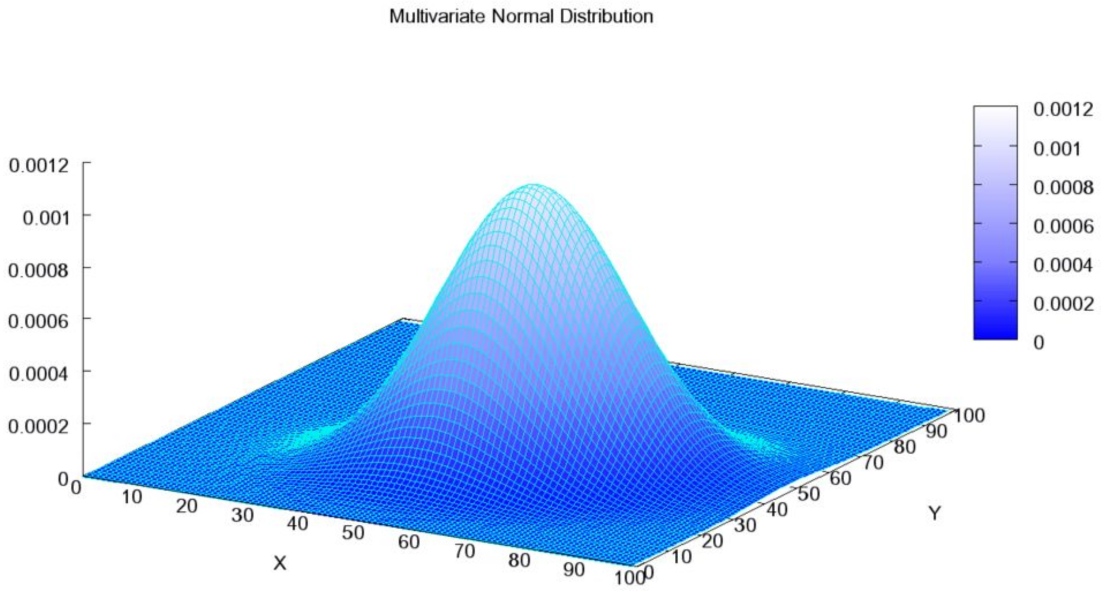
Distribuição conjunta ➜ função densidade de probabilidade conjunta;

Duas variáveis aleatórias (X, Y) ➜ fdp conjunta equivale a uma área (A).

$$P((X,Y) \in A) = \int\int f(x,y)dydx$$

Probabilidade de X e Y estar na área A equivale ao volume sob a área A.

Distribuição multivariada conjunta contínua



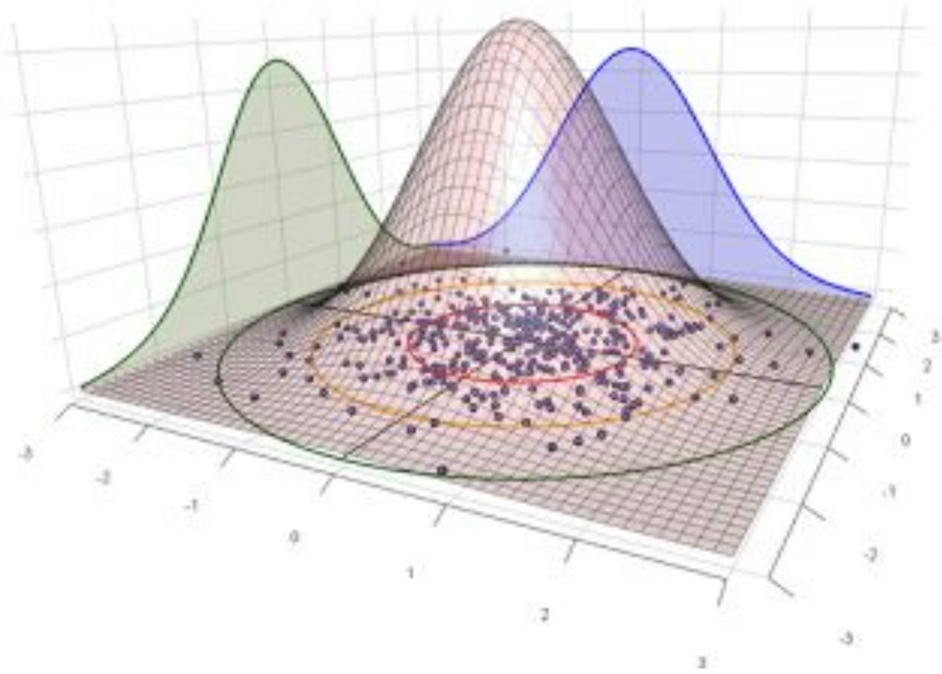
Distribuição multivariada contínua

Distribuição marginal

- **Discreta:** soma das probabilidades ao longo de uma variável;
- **Contínua:** integral da função densidade de probabilidade por uma das variáveis

$$f_x(x) = \int f(x, y)dy$$

Distribuição multivariada contínua marginal



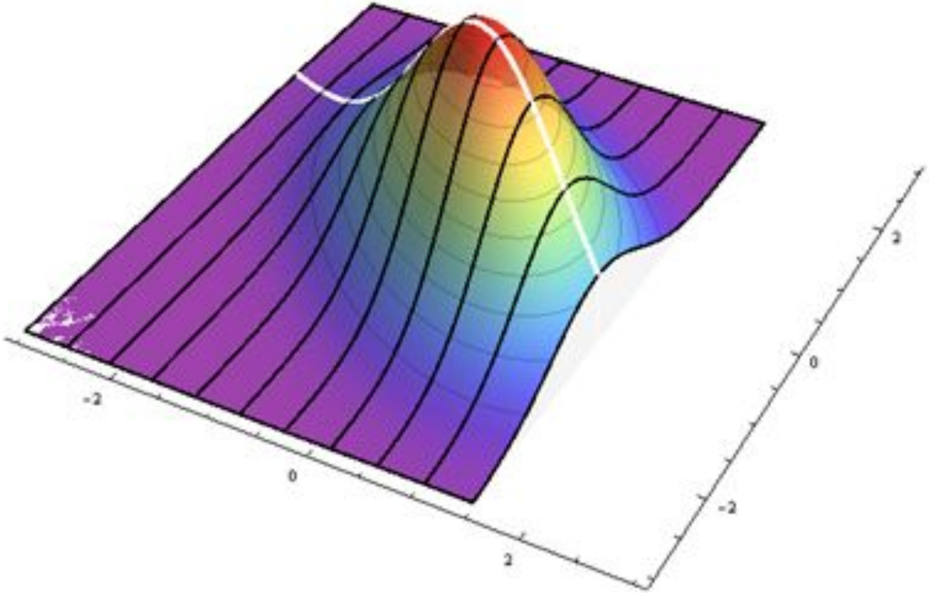
Distribuição multivariada contínua

Distribuição condicional

Se **X** e **Y** possuem uma função densidade de probabilidade conjunta **f(x,y)**, então a função densidade de probabilidade condicional para qualquer intervalo de X, dado que Y=y, é definido como a função conjunta de X e Y dividido pela função marginal de Y=y.

$$f_{x|y}(x|y) = \frac{f(x,y)}{f(y)}$$

Distribuição multivariada contínua condicional



$$f_{x|y}(x|y) = \frac{f(x,y)}{f(y)}$$