

# LGN5830 - BIOMETRIA DE MARCADORES GENÉTICOS

## AULA 3: MAPAS GENÉTICOS II

Antonio Augusto Franco Garcia  
Roland Vencovsky

Departamento de Genética  
ESALQ/USP  
2007

# CONTEÚDO

## 1 ANÁLISE DE LIGAÇÃO (CONT.)

- Revisão
- Teste de Dois Pontos ( $F_2$ )
- Grupos de Ligação

## 2 ORDENAÇÃO DOS LOCOS

- Fundamentos
- Estatísticas (Critérios)
- Algoritmos de Ordenação

# CONTEÚDO

## 1 ANÁLISE DE LIGAÇÃO (CONT.)

- Revisão
- Teste de Dois Pontos ( $F_2$ )
- Grupos de Ligação

## 2 ORDENAÇÃO DOS LOCOS

- Fundamentos
- Estatísticas (Critérios)
- Algoritmos de Ordenação











# TESTE DA RAZÃO DE VEROSSIMILHANÇA

- Princípio: comparar valores da verossimilhança considerando diferentes valores dos parâmetros
- $\theta$ : valor do parâmetro sob  $H_0$
- $\hat{\theta}$ : valor do parâmetro sob  $H_1$
- Estatística razão de verossimilhanças:

$$LRT = -2 \log \frac{L(\theta)}{L(\hat{\theta})} = -2[l(\theta) - l(\hat{\theta})]$$

- LOD (*log of the odds*):

$$LOD = \log_{10} \frac{L(\hat{\theta})}{L(\theta)}$$

## TESTE DA RAZÃO DE VEROSSIMILHANÇA

- Princípio: comparar valores da verossimilhança considerando diferentes valores dos parâmetros
- $\theta$ : valor do parâmetro sob  $H_0$
- $\hat{\theta}$ : valor do parâmetro sob  $H_1$
- Estatística razão de verossimilhanças:

$$LRT = -2 \log \frac{L(\theta)}{L(\hat{\theta})} = -2[l(\theta) - l(\hat{\theta})]$$

- LOD (*log of the odds*):

$$LOD = \log_{10} \frac{L(\hat{\theta})}{L(\theta)}$$

# TESTE DA RAZÃO DE VEROSSIMILHANÇA

- Princípio: comparar valores da verossimilhança considerando diferentes valores dos parâmetros
- $\theta$ : valor do parâmetro sob  $H_0$
- $\hat{\theta}$ : valor do parâmetro sob  $H_1$
- Estatística razão de verossimilhanças:

$$LRT = -2 \log \frac{L(\theta)}{L(\hat{\theta})} = -2[l(\theta) - l(\hat{\theta})]$$

- LOD (*log of the odds*):

$$LOD = \log_{10} \frac{L(\hat{\theta})}{L(\theta)}$$

# TESTE DA RAZÃO DE VEROSSIMILHANÇA

- Princípio: comparar valores da verossimilhança considerando diferentes valores dos parâmetros
- $\theta$ : valor do parâmetro sob  $H_0$
- $\hat{\theta}$ : valor do parâmetro sob  $H_1$
- Estatística razão de verossimilhanças:

$$LRT = -2 \log \frac{L(\theta)}{L(\hat{\theta})} = -2[l(\theta) - l(\hat{\theta})]$$

- LOD (*log of the odds*):

$$LOD = \log_{10} \frac{L(\hat{\theta})}{L(\theta)}$$

## TESTE DA RAZÃO DE VEROSSIMILHANÇA

- Princípio: comparar valores da verossimilhança considerando diferentes valores dos parâmetros
- $\theta$ : valor do parâmetro sob  $H_0$
- $\hat{\theta}$ : valor do parâmetro sob  $H_1$
- Estatística razão de verossimilhanças:

$$LRT = -2 \log \frac{L(\theta)}{L(\hat{\theta})} = -2[l(\theta) - l(\hat{\theta})]$$

- LOD (*log of the odds*):

$$LOD = \log_{10} \frac{L(\hat{\theta})}{L(\theta)}$$

## EXEMPLO

- Uso de um Sistema Algébrico Computacional (MAXIMA)

# CONTEÚDO

## 1 ANÁLISE DE LIGAÇÃO (CONT.)

- Revisão
- **Teste de Dois Pontos ( $F_2$ )**
- Grupos de Ligação

## 2 ORDENAÇÃO DOS LOCOS

- Fundamentos
- Estatísticas (Critérios)
- Algoritmos de Ordenação

POPULAÇÃO  $F_2$ 

Genótipo	Código	freq. esp. ( $p_i$ )	freq. obs. ( $f_i$ )
$\frac{AB}{AB}$	22	$\frac{(1-r)^2}{4}$	$n_1$
$\frac{AB}{Ab}$	21	$\frac{r(1-r)}{2}$	$n_2$
$\frac{Ab}{Ab}$	20	$\frac{r^2}{4}$	$n_3$
$\frac{AB}{aB}$	12	$\frac{r(1-r)}{2}$	$n_4$
$\frac{Ab}{aB}$	11	$\frac{r^2}{2}$	$n_5$
$\frac{AB}{ab}$	11	$\frac{(1-r)^2}{2}$	$n_5$
$\frac{Ab}{ab}$	10	$\frac{r(1-r)}{2}$	$n_6$
$\frac{aB}{aB}$	02	$\frac{r^2}{4}$	$n_7$
$\frac{aB}{ab}$	01	$\frac{r(1-r)}{2}$	$n_8$
$\frac{ab}{ab}$	00	$\frac{(1-r)^2}{4}$	$n_9$

$F_2$ 

- Função de verossimilhança para  $r$ :

$$L(r) = \left[ \frac{(1-r)^2}{4} \right]^{n_1+n_9} \left[ \frac{r(1-r)}{2} \right]^{n_2+n_4+n_6+n_8} \left[ \frac{r^2}{4} \right]^{n_3+n_7} \left[ \frac{(1-r)^2}{2} + \frac{r^2}{2} \right]^{n_5}$$

- Estimador de máxima verossimilhança para  $r$ :
  - Uso do MAXIMA

$F_2$ 

- Função de verossimilhança para  $r$ :

$$L(r) = \left[ \frac{(1-r)^2}{4} \right]^{n_1+n_9} \left[ \frac{r(1-r)}{2} \right]^{n_2+n_4+n_6+n_8} \left[ \frac{r^2}{4} \right]^{n_3+n_7} \left[ \frac{(1-r)^2}{2} + \frac{r^2}{2} \right]^{n_5}$$

- Estimador de máxima verossimilhança para  $r$ :
  - Uso do MAXIMA

$F_2$ 

- Função de verossimilhança para  $r$ :

$$L(r) = \left[ \frac{(1-r)^2}{4} \right]^{n_1+n_9} \left[ \frac{r(1-r)}{2} \right]^{n_2+n_4+n_6+n_8} \left[ \frac{r^2}{4} \right]^{n_3+n_7} \left[ \frac{(1-r)^2}{2} + \frac{r^2}{2} \right]^{n_5}$$

- Estimador de máxima verossimilhança para  $r$ :
  - Uso do MAXIMA

Maxima encountered a Lisp error:  
 Error in PROGN [or a callee]:  
 The storage for CONS is exhausted.  
 Currently, **48273 pages are allocated**.  
 Use ALLOCATE to expand the space.

# MÉTODOS NUMÉRICOS

- Em várias situações, não é possível obter formas explícitas para os MLE's
- Nesses casos, é comum a utilização de métodos numéricos (algoritmos)
- Um algoritmo muito usado é o **algoritmo EM**
  - E: Expectation; M: Maximization (Esperança e Maximização)

# MÉTODOS NUMÉRICOS

- Em várias situações, não é possível obter formas explícitas para os MLE's
- Nesses casos, é comum a utilização de métodos numéricos (algoritmos)
- Um algoritmo muito usado é o **algoritmo EM**
  - E: Expectation; M: Maximization (Esperança e Maximização)

# MÉTODOS NUMÉRICOS

- Em várias situações, não é possível obter formas explícitas para os MLE's
- Nesses casos, é comum a utilização de métodos numéricos (algoritmos)
- Um algoritmo muito usado é o **algoritmo EM**
  - E: Expectation; M: Maximization (Esperança e Maximização)

# ALGORITMO EM

- EM: método iterativo para obter estimativas de máxima verossimilhança
- Muito usado em análises de mapeamento genético
- Mostrou-se muito poderoso na prática, principalmente quando as observações possuem dados incompletos quanto à informação

## EXEMPLOS

- 1 Numa população  $F_2$ , não é possível separar as classes  $Ab/aB$  e  $AB/ab$
- 2 Para marcadores dominantes,  $AA = Aa$
- 3 No mapeamento de QTL's, não é possível separar  $QQ$ ,  $Qq$  e  $qq$ , já que os genótipos dos QTL's não são observáveis

# ALGORITMO EM

- EM: método iterativo para obter estimativas de máxima verossimilhança
- Muito usado em análises de mapeamento genético
- Mostrou-se muito poderoso na prática, principalmente quando as observações possuem dados incompletos quanto à informação

## EXEMPLOS

- 1 Numa população  $F_2$ , não é possível separar as classes  $Ab/aB$  e  $AB/ab$
- 2 Para marcadores dominantes,  $AA = Aa$
- 3 No mapeamento de QTL's, não é possível separar  $QQ$ ,  $Qq$  e  $qq$ , já que os genótipos dos QTL's não são observáveis

# ALGORITMO EM

- EM: método iterativo para obter estimativas de máxima verossimilhança
- Muito usado em análises de mapeamento genético
- Mostrou-se muito poderoso na prática, principalmente quando as observações possuem dados incompletos quanto à informação

## EXEMPLOS

- 1 Numa população  $F_2$ , não é possível separar as classes  $Ab/aB$  e  $AB/ab$
- 2 Para marcadores dominantes,  $AA = Aa$
- 3 No mapeamento de QTL's, não é possível separar  $QQ$ ,  $Qq$  e  $qq$ , já que os genótipos dos QTL's não são observáveis

# ALGORITMO EM

- EM: método iterativo para obter estimativas de máxima verossimilhança
- Muito usado em análises de mapeamento genético
- Mostrou-se muito poderoso na prática, principalmente quando as observações possuem dados incompletos quanto à informação

## EXEMPLOS

- 1 Numa população  $F_2$ , não é possível separar as classes  $Ab/aB$  e  $AB/ab$
- 2 Para marcadores dominantes,  $AA = Aa$
- 3 No mapeamento de QTL's, não é possível separar  $QQ$ ,  $Qq$  e  $qq$ , já que os genótipos dos QTL's não são observáveis

# ALGORITMO EM

- EM: método iterativo para obter estimativas de máxima verossimilhança
- Muito usado em análises de mapeamento genético
- Mostrou-se muito poderoso na prática, principalmente quando as observações possuem dados incompletos quanto à informação

## EXEMPLOS

- 1 Numa população  $F_2$ , não é possível separar as classes  $Ab/aB$  e  $AB/ab$
- 2 Para marcadores dominantes,  $AA = Aa$
- 3 No mapeamento de QTL's, não é possível separar  $QQ$ ,  $Qq$  e  $qq$ , já que os genótipos dos QTL's não são observáveis

# ALGORITMO EM

- EM: método iterativo para obter estimativas de máxima verossimilhança
- Muito usado em análises de mapeamento genético
- Mostrou-se muito poderoso na prática, principalmente quando as observações possuem dados incompletos quanto à informação

## EXEMPLOS

- 1 Numa população  $F_2$ , não é possível separar as classes  $Ab/aB$  e  $AB/ab$
- 2 Para marcadores dominantes,  $AA = Aa$
- 3 No mapeamento de QTL's, não é possível separar  $QQ$ ,  $Qq$  e  $qq$ , já que os genótipos dos QTL's não são observáveis

# ALGORITMO EM

- O EM faz uma clara distinção entre os **dados observados** ( $Y$ ), que são incompletos, e os **dados completos** ( $X$ ), não observáveis
- Alguma função  $t(X) = Y$  associa  $X$  e  $Y$
- Idéia básica: tomar  $X$  tal que a obtenção de estimativas de máxima verossimilhança torne-se trivial para os dados completos

# ALGORITMO EM

- O EM faz uma clara distinção entre os **dados observados** ( $Y$ ), que são incompletos, e os **dados completos** ( $X$ ), não observáveis
- Alguma função  $t(X) = Y$  associa  $X$  e  $Y$
- Idéia básica: tomar  $X$  tal que a obtenção de estimativas de máxima verossimilhança torne-se trivial para os dados completos

# ALGORITMO EM

- O EM faz uma clara distinção entre os **dados observados** ( $Y$ ), que são incompletos, e os **dados completos** ( $X$ ), não observáveis
- Alguma função  $t(X) = Y$  associa  $X$  e  $Y$
- Idéia básica: tomar  $X$  tal que a obtenção de estimativas de máxima verossimilhança torne-se trivial para os dados completos

# ALGORITMO EM

- Assume-se que os dados completos tenham função de densidade  $f(X/\theta)$
- **Passo E:** calcula-se a esperança condicional

$$Q(\theta/\theta_n) = E[\log f(X/\theta)|Y, \theta_n]$$

( $\theta_n$  é o valor estimado atual de  $\theta$ )

- **Passo M:** maximiza-se  $Q(\theta/\theta_n)$  com respeito à  $\theta$ , fornecendo uma nova estimativa para  $\theta_n$  (denominada  $\theta_{n+1}$ )
- Os dois passos são repetidos até haver convergência

# ALGORITMO EM

- Assume-se que os dados completos tenham função de densidade  $f(X/\theta)$
- **Passo E:** calcula-se a esperança condicional

$$Q(\theta/\theta_n) = E [\log f(X/\theta)|Y, \theta_n]$$

( $\theta_n$  é o valor estimado atual de  $\theta$ )

- **Passo M:** maximiza-se  $Q(\theta/\theta_n)$  com respeito à  $\theta$ , fornecendo uma nova estimativa para  $\theta_n$  (denominada  $\theta_{n+1}$ )
- Os dois passos são repetidos até haver convergência

# ALGORITMO EM

- Assume-se que os dados completos tenham função de densidade  $f(X/\theta)$
- **Passo E:** calcula-se a esperança condicional

$$Q(\theta/\theta_n) = E[\log f(X/\theta)|Y, \theta_n]$$

( $\theta_n$  é o valor estimado atual de  $\theta$ )

- **Passo M:** maximiza-se  $Q(\theta/\theta_n)$  com respeito à  $\theta$ , fornecendo uma nova estimativa para  $\theta_n$  (denominada  $\theta_{n+1}$ )
- Os dois passos são repetidos até haver convergência

# ALGORITMO EM

- Assume-se que os dados completos tenham função de densidade  $f(X/\theta)$
- **Passo E:** calcula-se a esperança condicional

$$Q(\theta/\theta_n) = E [\log f(X/\theta)|Y, \theta_n]$$

( $\theta_n$  é o valor estimado atual de  $\theta$ )

- **Passo M:** maximiza-se  $Q(\theta/\theta_n)$  com respeito à  $\theta$ , fornecendo uma nova estimativa para  $\theta_n$  (denominada  $\theta_{n+1}$ )
- Os dois passos são repetidos até haver convergência

# ALGORITMO EM

- Maximizando-se  $Q(\theta/\theta_n)$ , há aumento em  $\log g(Y/\theta)$  (verossimilhança dos dados observados)
- No passo E, estimam-se as estatísticas para os dados completos, condicionados aos dados incompletos observados
- No passo M, os dados completos estimados são usados para obter as estimativas de máxima verossimilhança

# ALGORITMO EM

- Maximizando-se  $Q(\theta/\theta_n)$ , há aumento em  $\log g(Y/\theta)$  (verossimilhança dos dados observados)
- No passo E, estimam-se as estatísticas para os dados completos, condicionados aos dados incompletos observados
- No passo M, os dados completos estimados são usados para obter as estimativas de máxima verossimilhança

# ALGORITMO EM

- Maximizando-se  $Q(\theta/\theta_n)$ , há aumento em  $\log g(Y/\theta)$  (verossimilhança dos dados observados)
- No passo E, estimam-se as estatísticas para os dados completos, condicionados aos dados incompletos observados
- No passo M, os dados completos estimados são usados para obter as estimativas de máxima verossimilhança

ESTIMANDO  $r$  NUM  $F_2$ 

- Vimos que  $f(X/\theta) = L(r) = \left[ \frac{(1-r)^2}{4} \right]^{n_1+n_9} \left[ \frac{r(1-r)}{2} \right]^{n_2+n_4+n_6+n_8} \left[ \frac{r^2}{4} \right]^{n_3+n_7} \left[ \frac{(1-r)^2}{2} + \frac{r^2}{2} \right]^{n_5}$
- Logo,
 
$$\log f(X/\theta) = n_1 \log \frac{(1-r)^2}{4} + n_2 \log \frac{r(1-r)}{2} + n_3 \log \frac{r^2}{4} + n_4 \log \frac{r(1-r)}{2} + n_{51} \log \frac{(1-r)^2}{2} + n_{52} \log \frac{r^2}{2} + n_6 \log \frac{r(1-r)}{2} + n_7 \log \frac{r^2}{4} + n_8 \log \frac{r(1-r)}{2} + n_9 \log \frac{(1-r)^2}{4}$$
- E portanto
 
$$E[\log f(X/\theta)] = n_1 \log \frac{(1-r)^2}{4} + n_2 \log \frac{r(1-r)}{2} + n_3 \log \frac{r^2}{4} + n_4 \log \frac{r(1-r)}{2} + \log \frac{(1-r)^2}{2} E(n_{51}) + \log \frac{r^2}{2} E(n_{52}) + n_6 \log \frac{r(1-r)}{2} + n_7 \log \frac{r^2}{4} + n_8 \log \frac{r(1-r)}{2} + n_9 \log \frac{(1-r)^2}{4}$$

ESTIMANDO  $r$  NUM  $F_2$ 

- Vimos que  $f(X/\theta) = L(r) = \left[ \frac{(1-r)^2}{4} \right]^{n_1+n_9} \left[ \frac{r(1-r)}{2} \right]^{n_2+n_4+n_6+n_8} \left[ \frac{r^2}{4} \right]^{n_3+n_7} \left[ \frac{(1-r)^2}{2} + \frac{r^2}{2} \right]^{n_5}$
- Logo,
 
$$\log f(X/\theta) = n_1 \log \frac{(1-r)^2}{4} + n_2 \log \frac{r(1-r)}{2} + n_3 \log \frac{r^2}{4} + n_4 \log \frac{r(1-r)}{2} + n_{51} \log \frac{(1-r)^2}{2} + n_{52} \log \frac{r^2}{2} + n_6 \log \frac{r(1-r)}{2} + n_7 \log \frac{r^2}{4} + n_8 \log \frac{r(1-r)}{2} + n_9 \log \frac{(1-r)^2}{4}$$
- E portanto
 
$$E[\log f(X/\theta)] = n_1 \log \frac{(1-r)^2}{4} + n_2 \log \frac{r(1-r)}{2} + n_3 \log \frac{r^2}{4} + n_4 \log \frac{r(1-r)}{2} + \log \frac{(1-r)^2}{2} E(n_{51}) + \log \frac{r^2}{2} E(n_{52}) + n_6 \log \frac{r(1-r)}{2} + n_7 \log \frac{r^2}{4} + n_8 \log \frac{r(1-r)}{2} + n_9 \log \frac{(1-r)^2}{4}$$

ESTIMANDO  $r$  NUM  $F_2$ 

- Vimos que  $f(X/\theta) = L(r) = \left[ \frac{(1-r)^2}{4} \right]^{n_1+n_9} \left[ \frac{r(1-r)}{2} \right]^{n_2+n_4+n_6+n_8} \left[ \frac{r^2}{4} \right]^{n_3+n_7} \left[ \frac{(1-r)^2}{2} + \frac{r^2}{2} \right]^{n_5}$
- Logo,
 
$$\log f(X/\theta) = n_1 \log \frac{(1-r)^2}{4} + n_2 \log \frac{r(1-r)}{2} + n_3 \log \frac{r^2}{4} + n_4 \log \frac{r(1-r)}{2} + n_{51} \log \frac{(1-r)^2}{2} + n_{52} \log \frac{r^2}{2} + n_6 \log \frac{r(1-r)}{2} + n_7 \log \frac{r^2}{4} + n_8 \log \frac{r(1-r)}{2} + n_9 \log \frac{(1-r)^2}{4}$$
- E portanto
 
$$E[\log f(X/\theta)] = n_1 \log \frac{(1-r)^2}{4} + n_2 \log \frac{r(1-r)}{2} + n_3 \log \frac{r^2}{4} + n_4 \log \frac{r(1-r)}{2} + \log \frac{(1-r)^2}{2} E(n_{51}) + \log \frac{r^2}{2} E(n_{52}) + n_6 \log \frac{r(1-r)}{2} + n_7 \log \frac{r^2}{4} + n_8 \log \frac{r(1-r)}{2} + n_9 \log \frac{(1-r)^2}{4}$$

ESTIMANDO  $r$  NUM  $F_2$ 

- Note que a Esperança foi calculada em relação à  $n_{51}$  e  $n_{52}$  que, se fossem conhecidos, tornariam o problema bem mais simples (essa é a idéia do EM!)
- $n_{51}$  e  $n_{52}$ : podem ser modelados usando a distribuição binomial
- Lembrete:  $X \sim B(n, p)$ ,  $E(X) = np$

$$P(n_{51}) = \frac{\frac{(1-r)^2}{2}}{\frac{(1-r)^2}{2} + \frac{r^2}{2}} = 1 - q; \quad P(n_{52}) = \frac{\frac{r^2}{2}}{\frac{(1-r)^2}{2} + \frac{r^2}{2}} = q$$

$$E(n_{51}) = n_5(1 - q);$$

$$E(n_{52}) = n_5q$$

ESTIMANDO  $r$  NUM  $F_2$ 

- Note que a Esperança foi calculada em relação à  $n_{51}$  e  $n_{52}$  que, se fossem conhecidos, tornariam o problema bem mais simples (essa é a idéia do EM!)
- $n_{51}$  e  $n_{52}$ : podem ser modelados usando a distribuição binomial
- Lembrete:  $X \sim B(n, p)$ ,  $E(X) = np$

$$P(n_{51}) = \frac{\frac{(1-r)^2}{2}}{\frac{(1-r)^2}{2} + \frac{r^2}{2}} = 1 - q; \quad P(n_{52}) = \frac{\frac{r^2}{2}}{\frac{(1-r)^2}{2} + \frac{r^2}{2}} = q$$

$$E(n_{51}) = n_5(1 - q);$$

$$E(n_{52}) = n_5q$$

ESTIMANDO  $r$  NUM  $F_2$ 

- Note que a Esperança foi calculada em relação à  $n_{51}$  e  $n_{52}$  que, se fossem conhecidos, tornariam o problema bem mais simples (essa é a idéia do EM!)
- $n_{51}$  e  $n_{52}$ : podem ser modelados usando a distribuição binomial
- Lembrete:  $X \sim B(n, p)$ ,  $E(X) = np$

$$P(n_{51}) = \frac{\frac{(1-r)^2}{2}}{\frac{(1-r)^2}{2} + \frac{r^2}{2}} = 1 - q; \quad P(n_{52}) = \frac{\frac{r^2}{2}}{\frac{(1-r)^2}{2} + \frac{r^2}{2}} = q$$

$$E(n_{51}) = n_5(1 - q);$$

$$E(n_{52}) = n_5q$$

ESTIMANDO  $r$  NUM  $F_2$ 

- Note que a Esperança foi calculada em relação à  $n_{51}$  e  $n_{52}$  que, se fossem conhecidos, tornariam o problema bem mais simples (essa é a idéia do EM!)
- $n_{51}$  e  $n_{52}$ : podem ser modelados usando a distribuição binomial
- Lembrete:  $X \sim B(n, p)$ ,  $E(X) = np$

$$P(n_{51}) = \frac{\frac{(1-r)^2}{2}}{\frac{(1-r)^2}{2} + \frac{r^2}{2}} = 1 - q; \quad P(n_{52}) = \frac{\frac{r^2}{2}}{\frac{(1-r)^2}{2} + \frac{r^2}{2}} = q$$

$$E(n_{51}) = n_5(1 - q);$$

$$E(n_{52}) = n_5q$$

ESTIMANDO  $r$  NUM  $F_2$ 

- Note que a Esperança foi calculada em relação à  $n_{51}$  e  $n_{52}$  que, se fossem conhecidos, tornariam o problema bem mais simples (essa é a idéia do EM!)
- $n_{51}$  e  $n_{52}$ : podem ser modelados usando a distribuição binomial
- Lembrete:  $X \sim B(n, p)$ ,  $E(X) = np$

$$P(n_{51}) = \frac{\frac{(1-r)^2}{2}}{\frac{(1-r)^2}{2} + \frac{r^2}{2}} = 1 - q; \quad P(n_{52}) = \frac{\frac{r^2}{2}}{\frac{(1-r)^2}{2} + \frac{r^2}{2}} = q$$

$$E(n_{51}) = n_5(1 - q);$$

$$E(n_{52}) = n_5q$$

ESTIMANDO  $r$  NUM  $F_2$ 

- Finalmente,

$$Q(\theta/\theta_n) = n_1 \log \frac{(1-r)^2}{4} + n_2 \log \frac{r(1-r)}{2} + n_3 \log \frac{r^2}{4} +$$

$$n_4 \log \frac{r(1-r)}{2} + \log \frac{(1-r)^2}{2} n_5(1-q) + \log \frac{r^2}{2} n_5q +$$

$$n_6 \log \frac{r(1-r)}{2} + n_7 \log \frac{r^2}{4} + n_8 \log \frac{r(1-r)}{2} + n_9 \log \frac{(1-r)^2}{4}$$

- Agora é possível derivar e encontrar o ponto de máximo
- Com ajuda do MAXIMA:

$$\hat{r} = \frac{(n_2 + n_4 + n_6 + n_8) + 2(n_3 + n_7 + qn_5)}{2n}$$

ESTIMANDO  $r$  NUM  $F_2$ 

- Finalmente,

$$Q(\theta/\theta_n) = n_1 \log \frac{(1-r)^2}{4} + n_2 \log \frac{r(1-r)}{2} + n_3 \log \frac{r^2}{4} +$$

$$n_4 \log \frac{r(1-r)}{2} + \log \frac{(1-r)^2}{2} n_5(1-q) + \log \frac{r^2}{2} n_5q +$$

$$n_6 \log \frac{r(1-r)}{2} + n_7 \log \frac{r^2}{4} + n_8 \log \frac{r(1-r)}{2} + n_9 \log \frac{(1-r)^2}{4}$$

- Agora é possível derivar e encontrar o ponto de máximo
- Com ajuda do MAXIMA:

$$\hat{r} = \frac{(n_2 + n_4 + n_6 + n_8) + 2(n_3 + n_7 + qn_5)}{2n}$$

ESTIMANDO  $r$  NUM  $F_2$ 

- Finalmente,

$$Q(\theta/\theta_n) = n_1 \log \frac{(1-r)^2}{4} + n_2 \log \frac{r(1-r)}{2} + n_3 \log \frac{r^2}{4} +$$

$$n_4 \log \frac{r(1-r)}{2} + \log \frac{(1-r)^2}{2} n_5(1-q) + \log \frac{r^2}{2} n_5q +$$

$$n_6 \log \frac{r(1-r)}{2} + n_7 \log \frac{r^2}{4} + n_8 \log \frac{r(1-r)}{2} + n_9 \log \frac{(1-r)^2}{4}$$

- Agora é possível derivar e encontrar o ponto de máximo
- Com ajuda do MAXIMA:

$$\hat{r} = \frac{(n_2 + n_4 + n_6 + n_8) + 2(n_3 + n_7 + qn_5)}{2n}$$

ESTIMANDO  $r$  NUM  $F_2$ 

- Finalmente,

$$Q(\theta/\theta_n) = n_1 \log \frac{(1-r)^2}{4} + n_2 \log \frac{r(1-r)}{2} + n_3 \log \frac{r^2}{4} +$$

$$n_4 \log \frac{r(1-r)}{2} + \log \frac{(1-r)^2}{2} n_5(1-q) + \log \frac{r^2}{2} n_5 q +$$

$$n_6 \log \frac{r(1-r)}{2} + n_7 \log \frac{r^2}{4} + n_8 \log \frac{r(1-r)}{2} + n_9 \log \frac{(1-r)^2}{4}$$

- Agora é possível derivar e encontrar o ponto de máximo
- Com ajuda do MAXIMA:

$$\hat{r} = \frac{(n_2 + n_4 + n_6 + n_8) + 2(n_3 + n_7 + qn_5)}{2n}$$

## USO DO EM PARA ESTIMAR $r$ NUM $F_2$

- **Passo E:** dado um valor inicial (chute) para  $r$ , obtém-se

$$q = \frac{\frac{r^2}{2}}{\frac{(1-r)^2}{2} + \frac{r^2}{2}}$$

- **Passo M:** usando esse valor de  $q$ ,  $r$  é estimado novamente

$$\hat{r} = \frac{(n_2 + n_4 + n_6 + n_8) + 2(n_3 + n_7 + qn_5)}{2n}$$

- O processo é repetido até a convergência

## USO DO EM PARA ESTIMAR $r$ NUM $F_2$

- **Passo E:** dado um valor inicial (chute) para  $r$ , obtém-se

$$q = \frac{\frac{r^2}{2}}{\frac{(1-r)^2}{2} + \frac{r^2}{2}}$$

- **Passo M:** usando esse valor de  $q$ ,  $r$  é estimado novamente

$$\hat{r} = \frac{(n_2 + n_4 + n_6 + n_8) + 2(n_3 + n_7 + qn_5)}{2n}$$

- O processo é repetido até a convergência

## USO DO EM PARA ESTIMAR $r$ NUM $F_2$

- **Passo E:** dado um valor inicial (chute) para  $r$ , obtém-se

$$q = \frac{\frac{r^2}{2}}{\frac{(1-r)^2}{2} + \frac{r^2}{2}}$$

- **Passo M:** usando esse valor de  $q$ ,  $r$  é estimado novamente

$$\hat{r} = \frac{(n_2 + n_4 + n_6 + n_8) + 2(n_3 + n_7 + qn_5)}{2n}$$

- O processo é repetido até a convergência

## USO DO EM

## EXEMPLO - MAIZE DATA

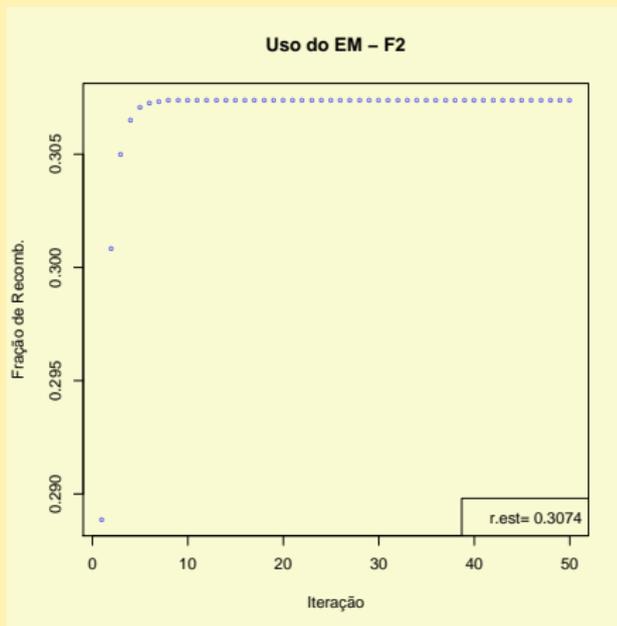
1	2	1	0	0	0	0	0	1	1	1	0	0
2	1	1	1	1	1	2	2	2	2	0	0	1
3	1	2	2	2	2	1	1	1	1	2	2	2
4	1	0	0	0	0	0	0	0	0	1	2	2
5	0	0	1	1	1	1	1	1	1	1	1	1
...												
170	1	0	0	1	1	1	1	1	1	1	1	1
171	2	2	1	1	0	1	1	1	1	0	0	0

- Obter  $\hat{r}$  entre  $M_1$  e  $M_2$  usando o algoritmo EM

# USO DO EM

## EXEMPLO - MAIZE DATA

- Usando o R



# FORMA GERAL PARA $\hat{r}$ USANDO O EM

- Liu (1998):
  - Vimos que

$$\hat{r} = \frac{(n_2 + n_4 + n_6 + n_8) + 2(n_3 + n_7 + n_5)}{2n}$$

- Forma geral:

$$\hat{r} = \frac{1}{n} \sum_i f_i P_i(R/G)$$

- $P_i(R/G)$ : prob. de um dado genótipo possuir um gameta recombinante
- Essa última expressão pode ser usado com combinações de marcadores, p. ex., dominante e co-dominante, etc.

# FORMA GERAL PARA $\hat{r}$ USANDO O EM

- Liu (1998):
  - Vimos que

$$\hat{r} = \frac{(n_2 + n_4 + n_6 + n_8) + 2(n_3 + n_7 + n_5)}{2n}$$

- Forma geral:

$$\hat{r} = \frac{1}{n} \sum_i f_i P_i(R/G)$$

- $P_i(R/G)$ : prob. de um dado genótipo possuir um gameta recombinante
- Essa última expressão pode ser usado com combinações de marcadores, p. ex., dominante e co-dominante, etc.

# FORMA GERAL PARA $\hat{r}$ USANDO O EM

- Liu (1998):
  - Vimos que

$$\hat{r} = \frac{(n_2 + n_4 + n_6 + n_8) + 2(n_3 + n_7 + qn_5)}{2n}$$

- Forma geral:

$$\hat{r} = \frac{1}{n} \sum_i f_i P_i(R/G)$$

- $P_i(R/G)$ : prob. de um dado genótipo possuir um gameta recombinante
- Essa última expressão pode ser usado com combinações de marcadores, p. ex., dominante e co-dominante, etc.

FORMA GERAL -  $F_2$ 

Genótipo	freq. esp. ( $p_i$ )	freq. obs. ( $f_i$ )	$P_i(R/G)$	
$\frac{AB}{AB}$	22	$\frac{(1-r)^2}{4}$	$n_1$	0
$\frac{AB}{Ab}$	21	$\frac{r(1-r)}{2}$	$n_2$	1/2
$\frac{Ab}{Ab}$	20	$\frac{r^2}{4}$	$n_3$	1
$\frac{AB}{aB}$	12	$\frac{r(1-r)}{2}$	$n_4$	1/2
$\frac{Ab}{aB}$	11	$\frac{r^2}{2}$	$n_5$	$\frac{(1)r^2/2}{r^2/2+(1-r)^2/2} = q$
$\frac{AB}{ab}$	11	$\frac{(1-r)^2}{2}$	$n_5$	
$\frac{Ab}{ab}$	10	$\frac{r(1-r)}{2}$	$n_6$	1/2
$\frac{aB}{aB}$	02	$\frac{r^2}{4}$	$n_7$	1
$\frac{aB}{ab}$	01	$\frac{r(1-r)}{2}$	$n_8$	1/2
$\frac{ab}{ab}$	00	$\frac{(1-r)^2}{4}$	$n_9$	0

FORMA GERAL -  $L(r)$ 

- Vimos que

$$L(r) = \left[ \frac{(1-r)^2}{4} \right]^{n_1+n_9} \left[ \frac{r(1-r)}{2} \right]^{n_2+n_4+n_6+n_8} \left[ \frac{r^2}{4} \right]^{n_3+n_7} \left[ \frac{(1-r)^2}{2} + \frac{r^2}{2} \right]^{n_5}$$

- Forma geral:

$$L(r) = \prod_i p_i^{f_i}$$

FORMA GERAL -  $L(r)$ 

- Vimos que

$$L(r) = \left[ \frac{(1-r)^2}{4} \right]^{n_1+n_9} \left[ \frac{r(1-r)}{2} \right]^{n_2+n_4+n_6+n_8} \left[ \frac{r^2}{4} \right]^{n_3+n_7} \left[ \frac{(1-r)^2}{2} + \frac{r^2}{2} \right]^{n_5}$$

- Forma geral:

$$L(r) = \prod_i p_i^{f_i}$$

# CONTEÚDO

## 1 ANÁLISE DE LIGAÇÃO (CONT.)

- Revisão
- Teste de Dois Pontos ( $F_2$ )
- Grupos de Ligação

## 2 ORDENAÇÃO DOS LOCOS

- Fundamentos
- Estatísticas (Critérios)
- Algoritmos de Ordenação

# GRUPOS DE LIGAÇÃO

- Biologicamente: grupos de genes no mesmo cromossomo
- Estatisticamente: grupos de locos que segregam conjuntamente
- Critérios:
  - Para um par de locos  $i$  e  $j$ , sejam
    - $r_{ij}$ : estimativa de dois pontos de  $r$
    - $p_{ij}$ :  $p$ -valor relativo à  $H_0 : r_{ij} = 1/2$
    - $z_{ij}$ : LOD Score

# GRUPOS DE LIGAÇÃO

- Biologicamente: grupos de genes no mesmo cromossomo
- Estatisticamente: grupos de locos que segregam conjuntamente
- Critérios:
  - Para um par de locos  $i$  e  $j$ , sejam
    - $r_{ij}$ : estimativa de dois pontos de  $r$
    - $p_{ij}$ :  $p$ -valor relativo à  $H_0 : r_{ij} = 1/2$
    - $z_{ij}$ : LOD Score

# GRUPOS DE LIGAÇÃO

- Biologicamente: grupos de genes no mesmo cromossomo
- Estatisticamente: grupos de locos que segregam conjuntamente
- Critérios:
  - Para um par de locos  $i$  e  $j$ , sejam
    - $r_{ij}$ : estimativa de dois pontos de  $r$
    - $p_{ij}$ :  $p$ -valor relativo à  $H_0 : r_{ij} = 1/2$
    - $z_{ij}$ : LOD Score

# FORMAÇÃO DOS GRUPOS

Usualmente:

- Se  $[r_{ij} \leq c \text{ e } p_{ij} \leq b]$ ,  $i$  e  $j$  pertencem ao mesmo grupo

OU

- Se  $[r_{ij} \leq c \text{ e } z_{ij} \geq a]$ ,  $i$  e  $j$  pertencem ao mesmo grupo

- $c$ : máx. fração recomb. para declarar ligação
- $b$ : máx. valor do  $p$ -valor para declarar ligação
- $a$ : mín. valor do  $LOD$  para declarar ligação
- $a$ ,  $b$  e  $c$ : atribuídos pelo usuário
- Usual:  $c = 0.5$  e  $a = 3$  (como saber se estão **corretos**?)
- Uso da propriedade **transitiva**

# FORMAÇÃO DOS GRUPOS

Usualmente:

- Se  $[r_{ij} \leq c$  e  $p_{ij} \leq b]$ ,  $i$  e  $j$  pertencem ao mesmo grupo

OU

- Se  $[r_{ij} \leq c$  e  $z_{ij} \geq a]$ ,  $i$  e  $j$  pertencem ao mesmo grupo

- $c$ : máx. fração recomb. para declarar ligação
- $b$ : máx. valor do  $p$ -valor para declarar ligação
- $a$ : mín. valor do  $LOD$  para declarar ligação
- $a$ ,  $b$  e  $c$ : atribuídos pelo usuário
- Usual:  $c = 0.5$  e  $a = 3$  (como saber se estão **corretos**?)
- Uso da propriedade **transitiva**

# FORMAÇÃO DOS GRUPOS

Usualmente:

- Se  $[r_{ij} \leq c$  e  $p_{ij} \leq b]$ ,  $i$  e  $j$  pertencem ao mesmo grupo

OU

- Se  $[r_{ij} \leq c$  e  $z_{ij} \geq a]$ ,  $i$  e  $j$  pertencem ao mesmo grupo

- $c$ : máx. fração recomb. para declarar ligação
- $b$ : máx. valor do  $p$ -valor para declarar ligação
- $a$ : mín. valor do  $LOD$  para declarar ligação
- $a$ ,  $b$  e  $c$ : atribuídos pelo usuário
- Usual:  $c = 0.5$  e  $a = 3$  (como saber se estão **corretos**?)
- Uso da propriedade **transitiva**

# FORMAÇÃO DOS GRUPOS

Usualmente:

- Se  $[r_{ij} \leq c \text{ e } p_{ij} \leq b]$ ,  $i$  e  $j$  pertencem ao mesmo grupo

OU

- Se  $[r_{ij} \leq c \text{ e } z_{ij} \geq a]$ ,  $i$  e  $j$  pertencem ao mesmo grupo

- $c$ : máx. fração recomb. para declarar ligação
- $b$ : máx. valor do  $p$ -valor para declarar ligação
- $a$ : mín. valor do  $LOD$  para declarar ligação
- $a$ ,  $b$  e  $c$ : atribuídos pelo usuário
- Usual:  $c = 0.5$  e  $a = 3$  (como saber se estão **corretos**?)
- Uso da propriedade **transitiva**

# FORMAÇÃO DOS GRUPOS

Usualmente:

- Se  $[r_{ij} \leq c \text{ e } p_{ij} \leq b]$ ,  $i$  e  $j$  pertencem ao mesmo grupo

OU

- Se  $[r_{ij} \leq c \text{ e } z_{ij} \geq a]$ ,  $i$  e  $j$  pertencem ao mesmo grupo

- $c$ : máx. fração recomb. para declarar ligação
- $b$ : máx. valor do  $p$ -valor para declarar ligação
- $a$ : mín. valor do  $LOD$  para declarar ligação
- $a$ ,  $b$  e  $c$ : atribuídos pelo usuário
- Usual:  $c = 0.5$  e  $a = 3$  (como saber se estão **corretos**?)
- Uso da propriedade **transitiva**

# FORMAÇÃO DOS GRUPOS

Usualmente:

- Se  $[r_{ij} \leq c \text{ e } p_{ij} \leq b]$ ,  $i$  e  $j$  pertencem ao mesmo grupo

OU

- Se  $[r_{ij} \leq c \text{ e } z_{ij} \geq a]$ ,  $i$  e  $j$  pertencem ao mesmo grupo

- $c$ : máx. fração recomb. para declarar ligação
- $b$ : máx. valor do  $p$ -valor para declarar ligação
- $a$ : mín. valor do  $LOD$  para declarar ligação
- $a$ ,  $b$  e  $c$ : atribuídos pelo usuário
- Usual:  $c = 0.5$  e  $a = 3$  (como saber se estão **corretos**?)
- Uso da propriedade **transitiva**

# FORMAÇÃO DOS GRUPOS

Usualmente:

- Se  $[r_{ij} \leq c$  e  $p_{ij} \leq b]$ ,  $i$  e  $j$  pertencem ao mesmo grupo

OU

- Se  $[r_{ij} \leq c$  e  $z_{ij} \geq a]$ ,  $i$  e  $j$  pertencem ao mesmo grupo

- $c$ : máx. fração recomb. para declarar ligação
- $b$ : máx. valor do  $p$ -valor para declarar ligação
- $a$ : mín. valor do  $LOD$  para declarar ligação
- $a$ ,  $b$  e  $c$ : atribuídos pelo usuário
- Usual:  $c = 0.5$  e  $a = 3$  (como saber se estão **corretos**?)
- Uso da propriedade **transitiva**

# LIGAÇÃO

## EXEMPLO - MAIZE DATA

- Há evidências de que os marcadores  $M_1$  e  $M_2$  estejam no mesmo grupo de ligação?

# LIGAÇÃO

## EXEMPLO - MAIZE DATA

- Há evidências de que os marcadores  $M_1$  e  $M_2$  estejam no mesmo grupo de ligação?
- Sim, já que  $\hat{r} = 0,3074$ ;  $LRT = 27,974$ ;  $p = 8,4 \times 10^{-7}$  e  $LOD = 6,07$

# CONTEÚDO

## 1 ANÁLISE DE LIGAÇÃO (CONT.)

- Revisão
- Teste de Dois Pontos ( $F_2$ )
- Grupos de Ligação

## 2 ORDENAÇÃO DOS LOCOS

- **Fundamentos**
- Estatísticas (Critérios)
- Algoritmos de Ordenação

# INTRODUÇÃO

- Sturtevant (1913): ordenação é um processo de minimizar o número de crossing-over
- Princípio: locos mais próximos possuem menor probabilidade de ocorrência de c.o.
- As estatísticas usadas para avaliar ordens são extensões das empregadas nos testes de três pontos

# INTRODUÇÃO

- Sturtevant (1913): ordenação é um processo de minimizar o número de crossing-over
- Princípio: locos mais próximos possuem menor probabilidade de ocorrência de c.o.
- As estatísticas usadas para avaliar ordens são extensões das empregadas nos testes de três pontos

# INTRODUÇÃO

- Sturtevant (1913): ordenação é um processo de minimizar o número de crossing-over
- Princípio: locos mais próximos possuem menor probabilidade de ocorrência de c.o.
- As estatísticas usadas para avaliar ordens são extensões das empregadas nos testes de três pontos

# CONTEÚDO

## 1 ANÁLISE DE LIGAÇÃO (CONT.)

- Revisão
- Teste de Dois Pontos ( $F_2$ )
- Grupos de Ligação

## 2 ORDENAÇÃO DOS LOCOS

- Fundamentos
- Estatísticas (Critérios)
- Algoritmos de Ordenação

# TESTE DE TRÊS PONTOS

## EXEMPLO

$$\hat{r}_{AB} = 0,10$$

$$\hat{r}_{AC} = 0,22$$

$$\hat{r}_{BC} = 0,30$$

- Qual a ordem dos locos?
- Difícil de generalizar e implementar. Não há garantia de obter a melhor ordem.

# TESTE DE TRÊS PONTOS

## EXEMPLO

$$\hat{r}_{AB} = 0,10$$

$$\hat{r}_{AC} = 0,22$$

$$\hat{r}_{BC} = 0,30$$

- Qual a ordem dos locos?
- Difícil de generalizar e implementar. Não há garantia de obter a melhor ordem.

# TESTE DE TRÊS PONTOS

## EXEMPLO

$$\hat{r}_{AB} = 0,10$$

$$\hat{r}_{AC} = 0,22$$

$$\hat{r}_{BC} = 0,30$$

- Qual a ordem dos locos?
  - Resp: B-A-C
- 
- Difícil de generalizar e implementar. Não há garantia de obter a melhor ordem.

# TESTE DE TRÊS PONTOS

## EXEMPLO

$$\hat{r}_{AB} = 0,10$$

$$\hat{r}_{AC} = 0,22$$

$$\hat{r}_{BC} = 0,30$$

- Qual a ordem dos locos?
- Resp: B-A-C
- Difícil de generalizar e implementar. Não há garantia de obter a melhor ordem.

# SARF, PARF, SALOD

- *SARF*: Sum of Adjacent Recombination Fraction

$$SARF = \sum_{i=1}^{m-1} \hat{r}_{a_i a_{i+1}}$$

- *PARF*: Product of Adjacent Recombination Fraction

$$PARF = \prod_{i=1}^{m-1} \hat{r}_{a_i a_{i+1}}$$

- *SALOD*: Sum of Adjacent Lod Score

$$SALOD = \sum_{i=1}^{m-1} \hat{z}_{a_i a_{i+1}}$$

OBJETIVO: Menor *SARF*, menor *PARF*, maior *SALOD*

# SARF, PARF, SALOD

- *SARF*: Sum of Adjacent Recombination Fraction

$$SARF = \sum_{i=1}^{m-1} \hat{r}_{a_i a_{i+1}}$$

- *PARF*: Product of Adjacent Recombination Fraction

$$PARF = \prod_{i=1}^{m-1} \hat{r}_{a_i a_{i+1}}$$

- *SALOD*: Sum of Adjacent Lod Score

$$SALOD = \sum_{i=1}^{m-1} \hat{z}_{a_i a_{i+1}}$$

OBJETIVO: Menor *SARF*, menor *PARF*, maior *SALOD*

# SARF, PARF, SALOD

- *SARF*: Sum of Adjacent Recombination Fraction

$$SARF = \sum_{i=1}^{m-1} \hat{r}_{a_i a_{i+1}}$$

- *PARF*: Product of Adjacent Recombination Fraction

$$PARF = \prod_{i=1}^{m-1} \hat{r}_{a_i a_{i+1}}$$

- *SALOD*: Sum of Adjacent Lod Score

$$SALOD = \sum_{i=1}^{m-1} \hat{z}_{a_i a_{i+1}}$$

OBJETIVO: Menor *SARF*, menor *PARF*, maior *SALOD*

# SARF, PARF, SALOD

- *SARF*: Sum of Adjacent Recombination Fraction

$$SARF = \sum_{i=1}^{m-1} \hat{r}_{a_i a_{i+1}}$$

- *PARF*: Product of Adjacent Recombination Fraction

$$PARF = \prod_{i=1}^{m-1} \hat{r}_{a_i a_{i+1}}$$

- *SALOD*: Sum of Adjacent Lod Score

$$SALOD = \sum_{i=1}^{m-1} \hat{z}_{a_i a_{i+1}}$$

**OBJETIVO:** Menor *SARF*, menor *PARF*, maior *SALOD*

# VEROSSIMILHANÇA

- Princípio: comparar a verossimilhança das ordens:

$$L(\theta) = \prod_i p_i^{f_i}$$

(cuidado, essa fórmula não é **multiponto**)

- A ordem com maior verossimilhança é a ordem mais provável para o conjunto de dados
- Dadas as propriedades estatísticas da verossimilhança, é o **melhor critério**
- Principais problemas:
  - 1 Cálculos complexos e demorados
  - 2 Embora forneça valores da verossimilhança, não é claro como esses valores podem ser comparados, já que não há GL's para tanto

# VEROSSIMILHANÇA

- Princípio: comparar a verossimilhança das ordens:

$$L(\theta) = \prod_i p_i^{f_i}$$

(cuidado, essa fórmula não é **multiponto**)

- A ordem com maior verossimilhança é a ordem mais provável para o conjunto de dados
- Dadas as propriedades estatísticas da verossimilhança, é o **melhor critério**
- Principais problemas:
  - 1 Cálculos complexos e demorados
  - 2 Embora forneça valores da verossimilhança, não é claro como esses valores podem ser comparados, já que não há GL's para tanto

# VEROSSIMILHANÇA

- Princípio: comparar a verossimilhança das ordens:

$$L(\theta) = \prod_i p_i^{f_i}$$

(cuidado, essa fórmula não é **multiponto**)

- A ordem com maior verossimilhança é a ordem mais provável para o conjunto de dados
- Dadas as propriedades estatísticas da verossimilhança, é o **melhor critério**
- Principais problemas:
  - 1 Cálculos complexos e demorados
  - 2 Embora forneça valores da verossimilhança, não é claro como esses valores podem ser comparados, já que não há GL's para tanto

# VEROSSIMILHANÇA

- Princípio: comparar a verossimilhança das ordens:

$$L(\theta) = \prod_i p_i^{f_i}$$

(cuidado, essa fórmula não é **multiponto**)

- A ordem com maior verossimilhança é a ordem mais provável para o conjunto de dados
- Dadas as propriedades estatísticas da verossimilhança, é o **melhor critério**
- Principais problemas:
  - 1 Cálculos complexos e demorados
  - 2 Embora forneça valores da verossimilhança, não é claro como esses valores podem ser comparados, já que não há GL's para tanto

# CONTEÚDO

## 1 ANÁLISE DE LIGAÇÃO (CONT.)

- Revisão
- Teste de Dois Pontos ( $F_2$ )
- Grupos de Ligação

## 2 ORDENAÇÃO DOS LOCOS

- Fundamentos
- Estatísticas (Critérios)
- Algoritmos de Ordenação

# BUSCA EXAUSTIVA

- Princípio: comparar todas as ordens possíveis, segundo algum critério (ex: veross.)

## EXEMPLO - MOUSE DATA (M1, M3 E M14)

- $L_{ABC}(\hat{r}_1, \hat{r}_2) = 1,922 \times 10^{-53}$
- $L_{ACB}(\hat{r}_3, \hat{r}_2) = 4,723 \times 10^{-89}$
- $L_{BAC}(\hat{r}_1, \hat{r}_3) = 6,789 \times 10^{-73}$

## MAPMAKER/EXP

Ordem	LOD
ABC	$\log_{10} \frac{1,922 \times 10^{-53}}{1,922 \times 10^{-53}} = 0$
ACB	$\log_{10} \frac{4,723 \times 10^{-89}}{1,922 \times 10^{-53}} = -35,61$
BAC	$\log_{10} \frac{6,789 \times 10^{-73}}{1,922 \times 10^{-53}} = -19,45$

# BUSCA EXAUSTIVA

- Princípio: comparar todas as ordens possíveis, segundo algum critério (ex: veross.)

## EXEMPLO - MOUSE DATA (M1, M3 E M14)

- $L_{ABC}(\hat{r}_1, \hat{r}_2) = 1,922 \times 10^{-53}$
- $L_{ACB}(\hat{r}_3, \hat{r}_2) = 4,723 \times 10^{-89}$
- $L_{BAC}(\hat{r}_1, \hat{r}_3) = 6,789 \times 10^{-73}$

## MAPMAKER/EXP

Ordem	LOD
ABC	$\log_{10} \frac{1,922 \times 10^{-53}}{1,922 \times 10^{-53}} = 0$
ACB	$\log_{10} \frac{4,723 \times 10^{-89}}{1,922 \times 10^{-53}} = -35,61$
BAC	$\log_{10} \frac{6,789 \times 10^{-73}}{1,922 \times 10^{-53}} = -19,45$

# BUSCA EXAUSTIVA

- Princípio: comparar todas as ordens possíveis, segundo algum critério (ex: veross.)

## EXEMPLO - MOUSE DATA (M1, M3 E M14)

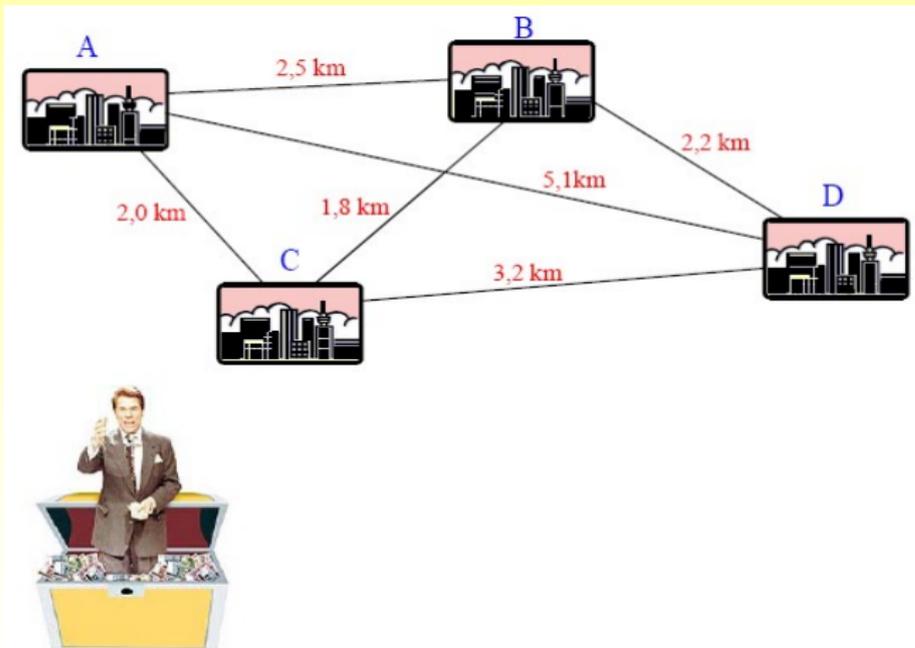
- $L_{ABC}(\hat{r}_1, \hat{r}_2) = 1,922 \times 10^{-53}$
- $L_{ACB}(\hat{r}_3, \hat{r}_2) = 4,723 \times 10^{-89}$
- $L_{BAC}(\hat{r}_1, \hat{r}_3) = 6,789 \times 10^{-73}$

## MAPMAKER/EXP

Ordem	LOD
ABC	$\log_{10} \frac{1,922 \times 10^{-53}}{1,922 \times 10^{-53}} = 0$
ACB	$\log_{10} \frac{4,723 \times 10^{-89}}{1,922 \times 10^{-53}} = -35,61$
BAC	$\log_{10} \frac{6,789 \times 10^{-73}}{1,922 \times 10^{-53}} = -19,45$

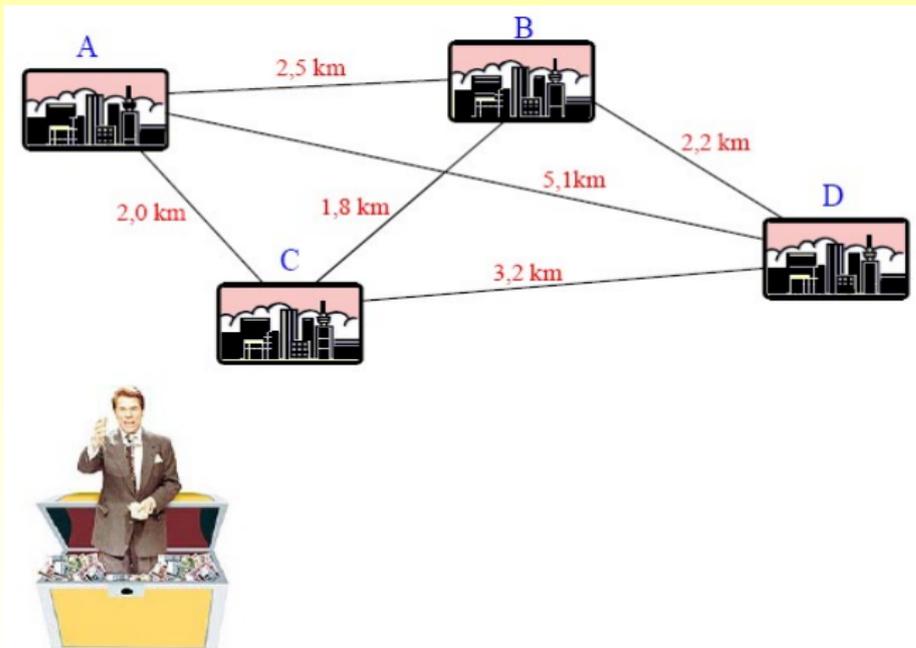
# O PROBLEMA DO CAIXEIRO VIAJANTE

- Problema: Deve visitar  $n$  cidades. Qual a rota mais curta?
- Solução: somar o caminho total percorrido e escolher a menor rota



# O PROBLEMA DO CAIXEIRO VIAJANTE

- Problema: Deve visitar  $n$  cidades. Qual a rota mais curta?
- Solução: somar o caminho total percorrido e escolher a menor rota



# PROBLEMA DO CAIXEIRO VIAJANTE

- **Problema:** Com aumento do número de cidades, o processo torna-se inviável, mesmo com supercomputadores

## EXEMPLO

- Computador que faz 1 bilhão de somas por segundo

Locos	Ordens	Tempos
5	$5!/2$	Insignificante
10	1.814.400	0,0162 seg
15	653.837.184.000	2,5427 horas
20	$1,2165 \times 10^{18}$	732,4 anos
25	$7,7556 \times 10^{24}$	5.898.373.012,27 anos

# PROBLEMA DO CAIXEIRO VIAJANTE

- **Problema:** Com aumento do número de cidades, o processo torna-se inviável, mesmo com supercomputadores

## EXEMPLO

- Computador que faz 1 bilhão de somas por segundo

Locos	Ordens	Tempos
5	$5!/2$	Insignificante
10	1.814.400	0,0162 seg
15	653.837.184.000	2,5427 horas
20	$1,2165 \times 10^{18}$	732,4 anos
25	$7,7556 \times 10^{24}$	5.898.373.012,27 anos

# ALGORITMOS ALTERNATIVOS

- 1 **Comando TRY**
  - Ordena sub-ordem exaustivamente, e depois testa posição dos marcadores remanescentes, um a um
- 2 Comando ORDER
  - Idem, mas de forma automática (cuidado!)
- 3 Comando RIPPLE
  - Verifica possíveis inversões entre marcadores próximos
- 4 Seriation
- 5 Rapid Chain Delineation
- 6 Branch and Bound
- 7 Simulated Annealing
- 8 ...

# ALGORITMOS ALTERNATIVOS

- 1 **Comando TRY**
  - Ordena sub-ordem exaustivamente, e depois testa posição dos marcadores remanescentes, um a um
- 2 **Comando ORDER**
  - Idem, mas de forma automática (cuidado!)
- 3 **Comando RIPPLE**
  - Verifica possíveis inversões entre marcadores próximos
- 4 Seriation
- 5 Rapid Chain Delineation
- 6 Branch and Bound
- 7 Simulated Annealing
- 8 ...

# ALGORITMOS ALTERNATIVOS

- 1 Comando TRY
  - Ordena sub-ordem exaustivamente, e depois testa posição dos marcadores remanescentes, um a um
- 2 Comando ORDER
  - Idem, mas de forma automática (cuidado!)
- 3 Comando RIPPLE
  - Verifica possíveis inversões entre marcadores próximos
- 4 Seriation
- 5 Rapid Chain Delineation
- 6 Branch and Bound
- 7 Simulated Annealing
- 8 ...

# ALGORITMOS ALTERNATIVOS

- 1 Comando TRY
  - Ordena sub-ordem exaustivamente, e depois testa posição dos marcadores remanescentes, um a um
- 2 Comando ORDER
  - Idem, mas de forma automática (cuidado!)
- 3 Comando RIPPLE
  - Verifica possíveis inversões entre marcadores próximos
- 4 Seriation
- 5 Rapid Chain Delineation
- 6 Branch and Bound
- 7 Simulated Annealing
- 8 ...

# ALGORITMOS ALTERNATIVOS

- 1 Comando TRY
  - Ordena sub-ordem exaustivamente, e depois testa posição dos marcadores remanescentes, um a um
- 2 Comando ORDER
  - Idem, mas de forma automática (cuidado!)
- 3 Comando RIPPLE
  - Verifica possíveis inversões entre marcadores próximos
- 4 Seriation
- 5 Rapid Chain Delineation
- 6 Branch and Bound
- 7 Simulated Annealing
- 8 ...

# ALGORITMOS ALTERNATIVOS

- 1 Comando TRY
  - Ordena sub-ordem exaustivamente, e depois testa posição dos marcadores remanescentes, um a um
- 2 Comando ORDER
  - Idem, mas de forma automática (cuidado!)
- 3 Comando RIPPLE
  - Verifica possíveis inversões entre marcadores próximos
- 4 Seriation
- 5 Rapid Chain Delineation
- 6 Branch and Bound
- 7 Simulated Annealing
- 8 ...

# ALGORITMOS ALTERNATIVOS

- 1 Comando TRY
  - Ordena sub-ordem exaustivamente, e depois testa posição dos marcadores remanescentes, um a um
- 2 Comando ORDER
  - Idem, mas de forma automática (cuidado!)
- 3 Comando RIPPLE
  - Verifica possíveis inversões entre marcadores próximos
- 4 Seriation
- 5 Rapid Chain Delineation
- 6 Branch and Bound
- 7 Simulated Annealing
- 8 ...

# RAPID CHAIN DELINEATION (DOERGE E WEIR, 1996)

- A partir da matriz de frações de recombinação:
  - 1 Inicie o primeiro grupo de ligação com o par de marcas com o menor  $\hat{r}$
  - 2 O grupo de ligação é estendido com a adição de marcas remanescentes. Os terminais do grupo de ligação são candidatos à extensão da cadeia. As marcas com os menores  $\hat{r}$  são adicionadas uma a uma (somente ligações significativas).
  - 3 Repetir o passo 2 até que nenhum marcador possa ser adicionado à cadeia.
  - 4 Iniciar o passo 1, com a obtenção de outra cadeia. Repetir os passos 2 e 3.
  - 5 Encerrar o processo quando não restarem marcas remanescentes, ou quando as restantes não estiverem significativamente ligadas a nenhuma outra.

# RAPID CHAIN DELINEATION (DOERGE E WEIR, 1996)

- A partir da matriz de frações de recombinação:
  - 1 Inicie o primeiro grupo de ligação com o par de marcas com o menor  $\hat{r}$
  - 2 O grupo de ligação é estendido com a adição de marcas remanescentes. Os terminais do grupo de ligação são candidatos à extensão da cadeia. As marcas com os menores  $\hat{r}$  são adicionadas uma a uma (somente ligações significativas).
  - 3 Repetir o passo 2 até que nenhum marcador possa ser adicionado à cadeia.
  - 4 Iniciar o passo 1, com a obtenção de outra cadeia. Repetir os passos 2 e 3.
  - 5 Encerrar o processo quando não restarem marcas remanescentes, ou quando as restantes não estiverem significativamente ligadas a nenhuma outra.

# RAPID CHAIN DELINEATION (DOERGE E WEIR, 1996)

- A partir da matriz de frações de recombinação:
  - 1 Inicie o primeiro grupo de ligação com o par de marcas com o menor  $\hat{r}$
  - 2 O grupo de ligação é estendido com a adição de marcas remanescentes. Os terminais do grupo de ligação são candidatos à extensão da cadeia. As marcas com os menores  $\hat{r}$  são adicionadas uma a uma (somente ligações significativas).
  - 3 Repetir o passo 2 até que nenhum marcador possa ser adicionado à cadeia.
  - 4 Iniciar o passo 1, com a obtenção de outra cadeia. Repetir os passos 2 e 3.
  - 5 Encerrar o processo quando não restarem marcas remanescentes, ou quando as restantes não estiverem significativamente ligadas a nenhuma outra.

# RAPID CHAIN DELINEATION (DOERGE E WEIR, 1996)

- A partir da matriz de frações de recombinação:
  - 1 Inicie o primeiro grupo de ligação com o par de marcas com o menor  $\hat{r}$
  - 2 O grupo de ligação é estendido com a adição de marcas remanescentes. Os terminais do grupo de ligação são candidatos à extensão da cadeia. As marcas com os menores  $\hat{r}$  são adicionadas uma a uma (somente ligações significativas).
  - 3 Repetir o passo 2 até que nenhum marcador possa ser adicionado à cadeia.
  - 4 Iniciar o passo 1, com a obtenção de outra cadeia. Repetir os passos 2 e 3.
  - 5 Encerrar o processo quando não restarem marcas remanescentes, ou quando as restantes não estiverem significativamente ligadas a nenhuma outra.

# RAPID CHAIN DELINEATION (DOERGE E WEIR, 1996)

- A partir da matriz de frações de recombinação:
  - 1 Inicie o primeiro grupo de ligação com o par de marcas com o menor  $\hat{r}$
  - 2 O grupo de ligação é estendido com a adição de marcas remanescentes. Os terminais do grupo de ligação são candidatos à extensão da cadeia. As marcas com os menores  $\hat{r}$  são adicionadas uma a uma (somente ligações significativas).
  - 3 Repetir o passo 2 até que nenhum marcador possa ser adicionado à cadeia.
  - 4 Iniciar o passo 1, com a obtenção de outra cadeia. Repetir os passos 2 e 3.
  - 5 Encerrar o processo quando não restarem marcas remanescentes, ou quando as restantes não estiverem significativamente ligadas a nenhuma outra.

# RAPID CHAIN DELINEATION

## EXEMPLO

- Com base na matriz abaixo, ordene os marcadores usando o RCD

	$m_1$	$m_2$	$m_3$	$m_4$
$m_1$		0,030	0,035	0,042
$m_2$			0,100	0,125
$m_3$				0,230
$m_4$				

# RAPID CHAIN DELINEATION

## EXEMPLO

- Com base na matriz abaixo, ordene os marcadores usando o RCD

	$m_1$	$m_2$	$m_3$	$m_4$
$m_1$		0,030	0,035	0,042
$m_2$			0,100	0,125
$m_3$				0,230
$m_4$				

- Resp:  $m_3-m_1-m_2-m_4$

# ANÁLISE DE DADOS

## EXEMPLO

- Uso do R para construção de mapas