

Princípios de Sistemática & Biogeografia

AULA 7: MAXIMUM LIKELIHOOD (ML)

ESTIMATIVA POR MÁXIMA VERROSIMILHANÇA

Darwin
1809-1882



1859

Mundo dinâmico

Resistência e Nova Síntese

Sistemática Evolutiva

1936 - 1947

1960's

Fenética

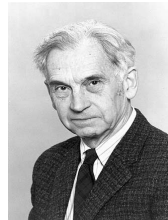
1970's

Cladística

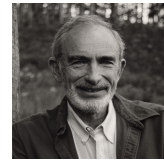
1990's

Probabilismo

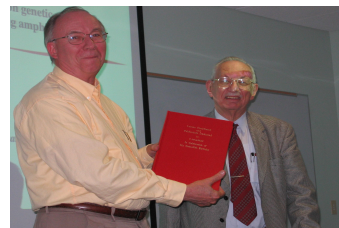
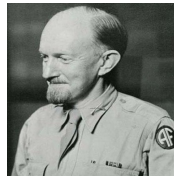
Ernest Mayr
1904 - 2005



Paul Erlich

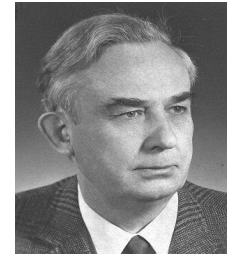


G.G. Simpson
1902 - 1984



James Rohlf

R. Sokal
1926 -



Willi Hennig
1913 - 1976



Joe Felsenstein

Theodosius Dobzhansky
1900 -1975

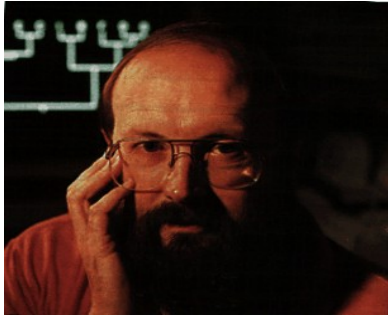


Steve Farris



David Hillis

ML – Motivação:



Joe Felsenstein
University of Washington

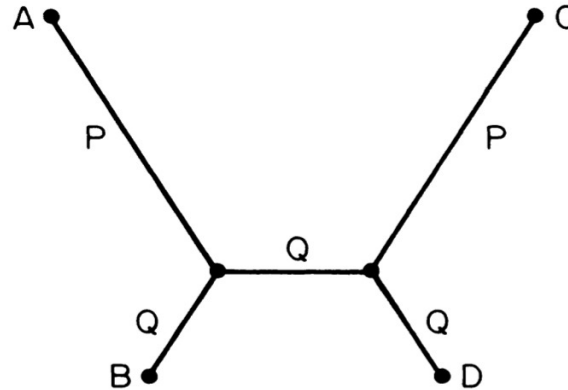


FIG. 3.—True unknown phylogeny (with root omitted) used to find cases in which unrooted Wagner tree parsimony methods will be inconsistent.

CASES IN WHICH PARSIMONY OR COMPATIBILITY METHODS WILL BE POSITIVELY MISLEADING¹

JOSEPH FELSENSTEIN

Abstract

Felsenstein, J. (Department of Genetics, University of Washington, Seattle, WA 98195) 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. 27:401–410.—For some simple three- and four-species cases involving a character with two states, it is determined under what conditions several methods of phylogenetic inference will fail to converge to the true phylogeny as more and more data are accumulated. The methods are the Camin-Sokal parsimony method, the compatibility method, and Farris's unrooted Wagner tree parsimony method. In all cases the conditions for this failure (which is the failure to be statistically consistent) are essentially that parallel changes exceed informative, nonparallel changes. It is possible for these methods to be inconsistent even when change is improbable a priori, provided that evolutionary rates in different lineages are sufficiently unequal. It is by extension of this approach that we may provide a sound methodology for evaluating methods of phylogenetic inference. [Numerical cladistics; phylogenetic inference; maximum likelihood estimation; parsimony; compatibility.]

Parsimony or minimum evolution methods were first introduced into phylogenetic inference by Camin and Sokal (1965). This class of methods for inferring an evolutionary tree from discrete-character data involves making a reconstruction of the changes in a given set of characters on a given tree, counting the smallest number of times that a given kind of event need have happened, and using this as the measure of the adequacy of the evolutionary tree. (Alternatively, one can compute the weighted sum of the numbers of times several different kinds of events have occurred.) One attempts to find that evolutionary tree which requires the fewest of these evolutionary events to explain the observed data. Camin and Sokal treated the case of irreversible changes along a character state tree, minimizing the number of changes

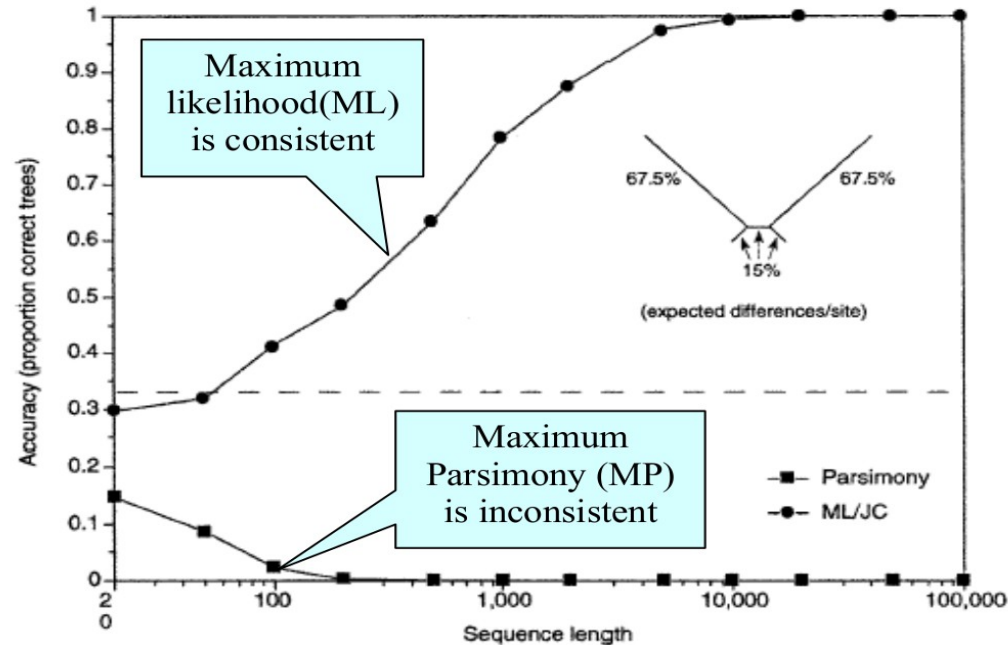
of character states required. A number of other parsimony methods have since appeared in the systematic literature (Kluge and Farris, 1969; Farris, 1969, 1970, 1972, 1977; Farris, Kluge, and Eckhardt, 1970) and parsimony methods have also found widespread use in studies of molecular evolution (Fitch and Margoliash, 1967, 1970; Dayhoff and Eck, 1968; see also Fitch, 1973). Cavalli-Sforza and Edwards (1967; Edwards and Cavalli-Sforza, 1964) earlier formulated a minimum evolution method for continuous-character data.

An alternative methodology for phylogenetic inference is the compatibility method, introduced by Le Quesne (1969, 1972). He suggested that phylogenetic inference be based on finding the largest possible set of characters which could simultaneously have all states be uniquely derived on the same tree. The estimate of the phylogeny is then taken to be that tree. While Le Quesne's specific suggestions as to how this might be done have been criticized by Farris (1969), his general approach, which is based on Camin and Sokal's (1965) concept of the compatibility of two characters, has been made rigorous and extended in a series of papers by G. F. Estabrook, C. S. Johnson, Jr., and F. R. McMorris (Estabrook,

¹ This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Department of Energy, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately-owned rights.

ML – Motivação:

"Felsenstein Zone" Trees



Accuracy is proportion correct out of 1000 simulation replicates

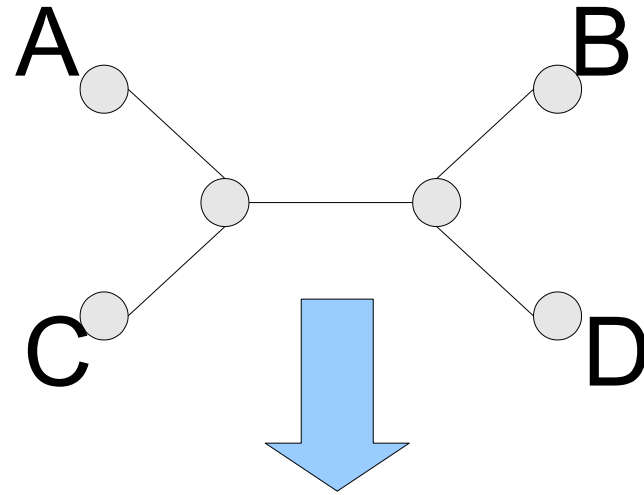
Dashed line is 1/3, the probability of getting the tree correct by pure random chance

Fig. 3, p. 531, from Swofford et al. 2001. Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. *Systematic Biology* 50: 525-539.

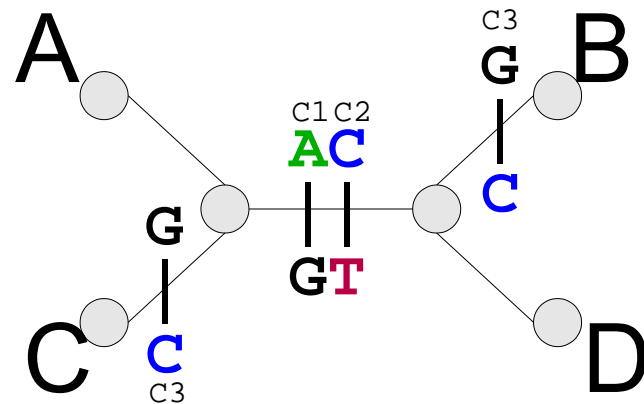
ML – Motivação:

Considere:

Taxon	c1	c2	c3
A	A	C	C
B	G	T	G
C	A	C	G
D	G	T	C



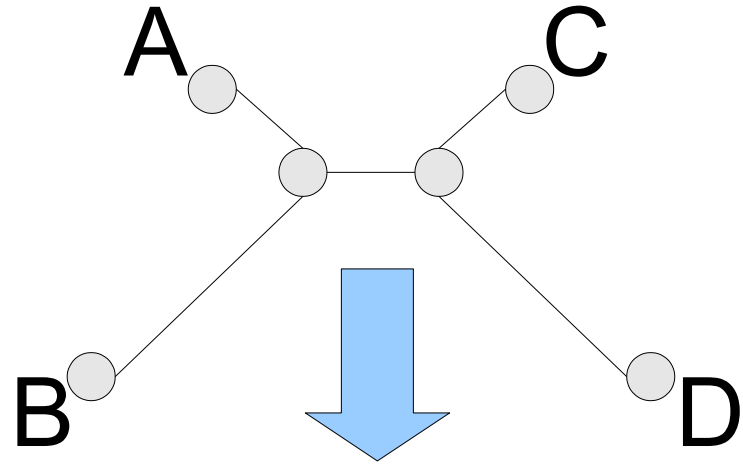
Parcimônia: 4 transformações



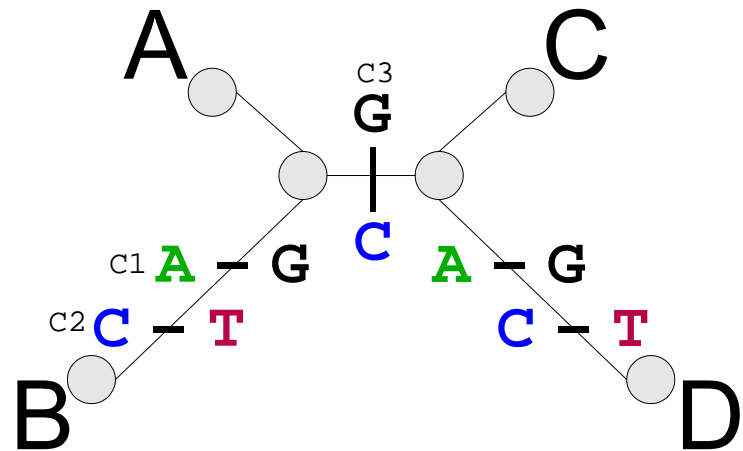
ML – Motivação:

Considere:

Taxon	c1	c2	c3
A	A	C	C
B	G	T	G
C	A	C	G
D	G	T	C



$$\text{ML: } P_{(G-C)} > P_{(A-G)} \times P_{(C-T)}$$



Probabilidades:

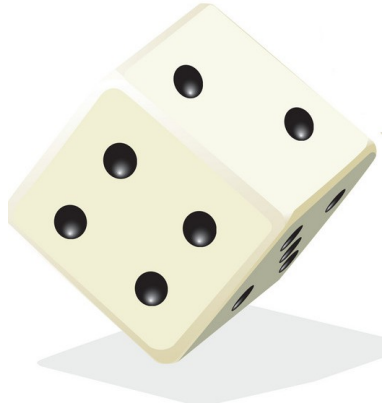
Probabilidade: proposição numérica da frequência de ocorrência de determinado evento.

$p(x)$ → se refere a probabilidade do evento x ocorrer.

X → variável aleatória é o evento – ou conjunto – cujo valor depende do resultado de processos estocásticos (aleatórios).

Espaço amostral → se refere ao conjunto de resultados possíveis de um processo aleatório.

Probabilidades:



Para dados honestos, o sorteio de uma das faces é resultado de um processo estocástico.

Em um dado com 6 faces, o conjunto de eventos possíveis configura o **espaço amostral*** (**S**):



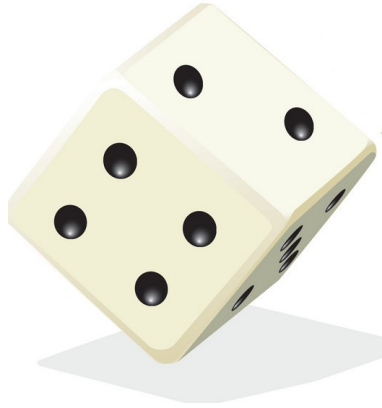
Qualquer um desses resultados é uma ou mais variáveis.

A probabilidade de um evento ($P(X=1)$) ocorrer é:

$$P_{(X=1)} = \frac{1}{S} = \frac{1}{6}$$

(*) ou espaço de resultados.

Probabilidades:



A probabilidade de um evento ($P(X=1)$) ocorrer é:

$$P_{(X=1)} = \frac{1}{S} = \frac{1}{6}$$

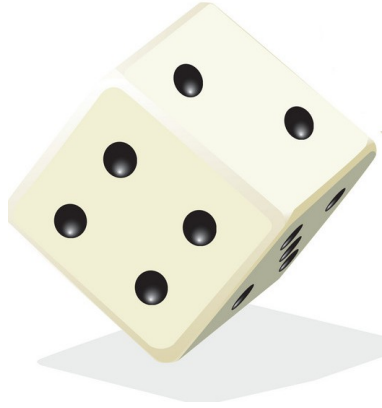
onde:

$$0 \leq P_{(x)} \leq 1$$

e

$$P_{(\bar{X}=1)} = 1 - P_{(X=1)}$$

Probabilidades:



A probabilidade da união de eventos **disjuntos** é dada pela soma das probabilidades destes eventos.

$$P_{(X=3 \text{ ou } X=5)} = P_{(X=3 \cup X=5)} = P_{(X=3)} + P_{(X=5)}$$

Exemplo: A probabilidade de jogar um dado e obter 3 ou 5 é:

$$P_{(X=3 \cup X=5)} = P_{(X=3)} + P_{(X=5)} = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

Probabilidades:



A probabilidade da **intersecção** de dois ou mais eventos é dada pela multiplicação das probabilidades destes eventos.

$$P_{(X=1 \text{ e } X=2)} = P_{(X=1 \cap X=2)} = P_{(X=1)} * P_{(X=2)}$$

Exemplo: A probabilidade de jogar dois dados e obter 1 e 2 é:

$$P_{(X=1 \cap X=2)} = P_{(X=1)} * P_{(X=2)} = \left(\frac{1}{6}\right) * \left(\frac{1}{6}\right) = \frac{1}{36}$$

Probabilidades:



A probabilidade de que o evento A ocorra, dado que o evento B já ocorreu, é chamada **probabilidade condicional**. A probabilidade do evento A, dado o evento B, é expressa da seguinte forma:

$$P_{(A|B)} = \frac{P_{(A \cap B)}}{P_{(B)}}$$

Exemplo: Suponha que você que saber a probabilidade de obter dois eventos de caras seguidos sabendo que no seu primeiro lance você já obteve uma cara?

Probabilidades:



Exemplo: Suponha que você que saber a probabilidade de obter dois eventos de caras seguidos sabendo que no seu primeiro lance você já obteve uma cara?

S = {(H,H),(H,T),(T,H),(T,T)} onde H=cara e T=coroa













$$P_{(H|T)} = \frac{P_{(H)} * P_{(T)}}{P_{(T)}} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$$

Probabilidades:



Combinando eventos de união e **intersecção**:

Qual a probabilidade de jogarmos dois dados e obter soma igual a 7?

						
	2	3	4	5	6	7
	3	4	5	6	7	8
	4	5	6	7	8	9
	5	6	7	8	9	10
	6	7	8	9	10	11
	7	8	9	10	11	12

Probabilidades:



Combinando eventos de união e **intersecção**:

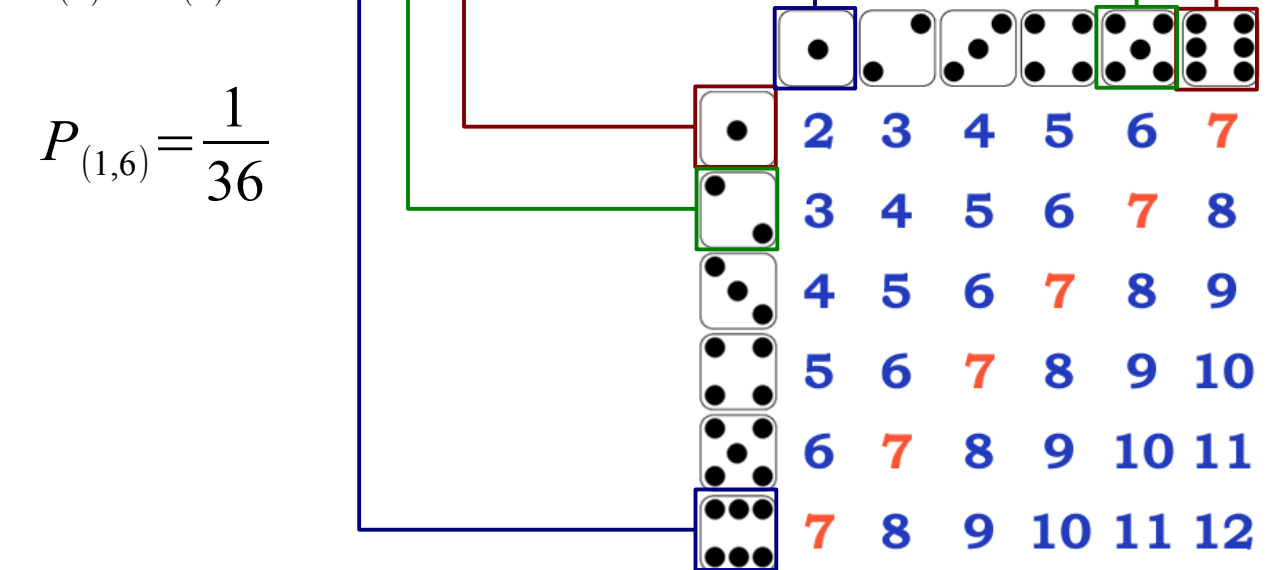
Qual a probabilidade de jogarmos dois dados e obter soma igual a 7?

$$P_{(1)} * P_{(6)}$$

$$P_{(2)} * P_{(5)}$$

$$P_{(6)} * P_{(1)}$$

$$P_{(1,6)} = \frac{1}{36}$$



Probabilidades:



Qual a probabilidade de jogarmos dois dados e obter soma igual a 7?

$$P_{(1)} * P_{(6)}$$

$$P_{(2)} * P_{(5)}$$

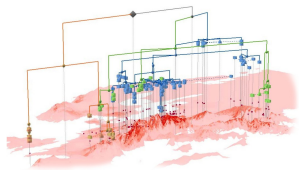
$$P_{(6)} * P_{(1)}$$

$$P_{(1,6)} = \frac{1}{36}$$

2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12
7	8	9	10	11	12	

$$P_{(\sum=7)} = P_{(1,6)} + P_{(2,5)} + P_{(3,4)} + P_{(4,3)} + P_{(5,2)} + P_{(6,1)}$$

$$P_{(\sum=7)} = \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = 1/6$$



Aula 07: Máxima Verossimilhança

Aula 7.02: Probabilidades;

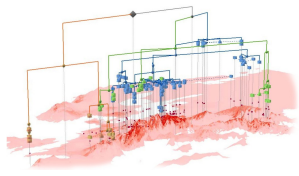
Aula 7.03: Funções, modelos e distribuições;

Aula 7.04: Máxima Verossimilhança (MLE);

Aula 7.05: MLE como critério de otimalidade;

Aula 7.06: Cálculo de probabilidade em topologias;

Aula 7.07: Considerações finais.



Aula 07: Máxima Verossimilhança

Aula 7.02: Probabilidades;

Aula 7.03: Funções, modelos e distribuições;

Aula 7.04: Máxima Verossimilhança (MLE);

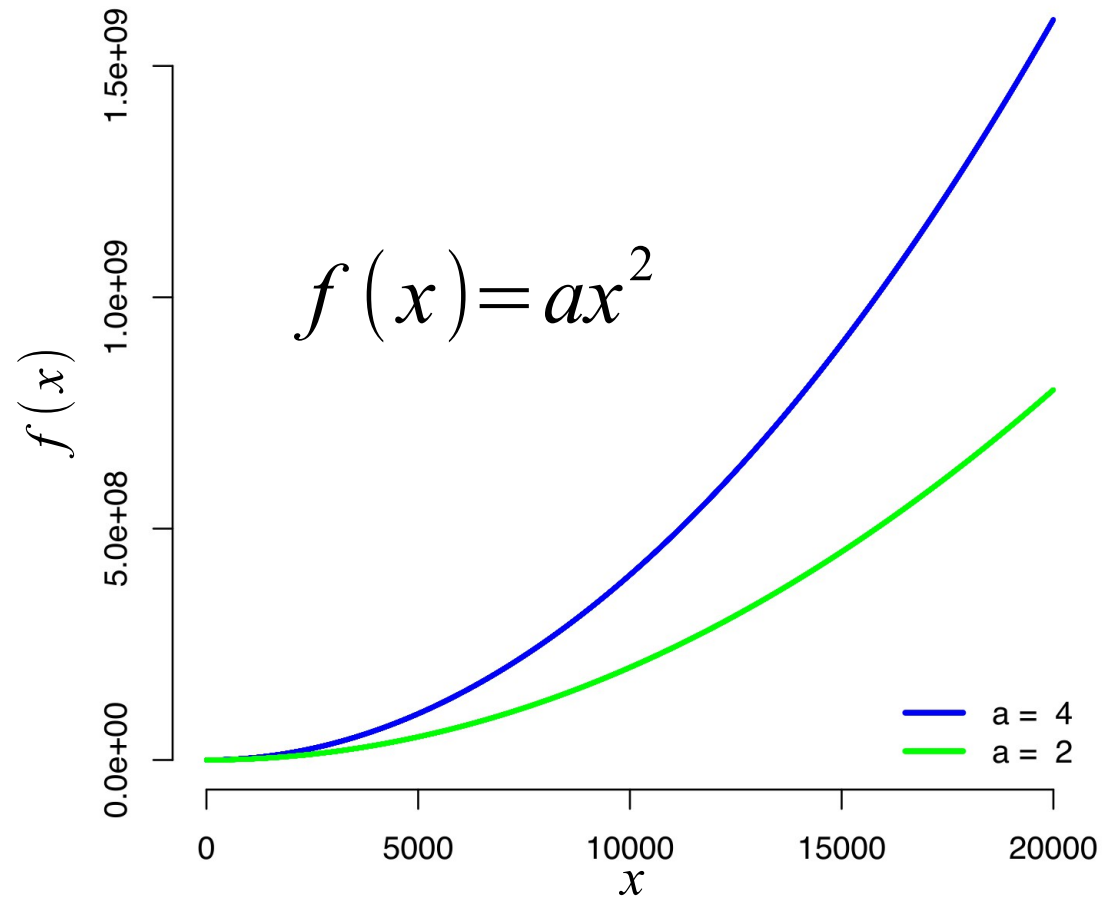
Aula 7.05: MLE como critério de otimalidade;

Aula 7.06: Cálculo de probabilidade em topologias;

Aula 7.07: Considerações finais.

Funções, modelos e distribuições:

Uma **função** expressa a relação entre quantidades:

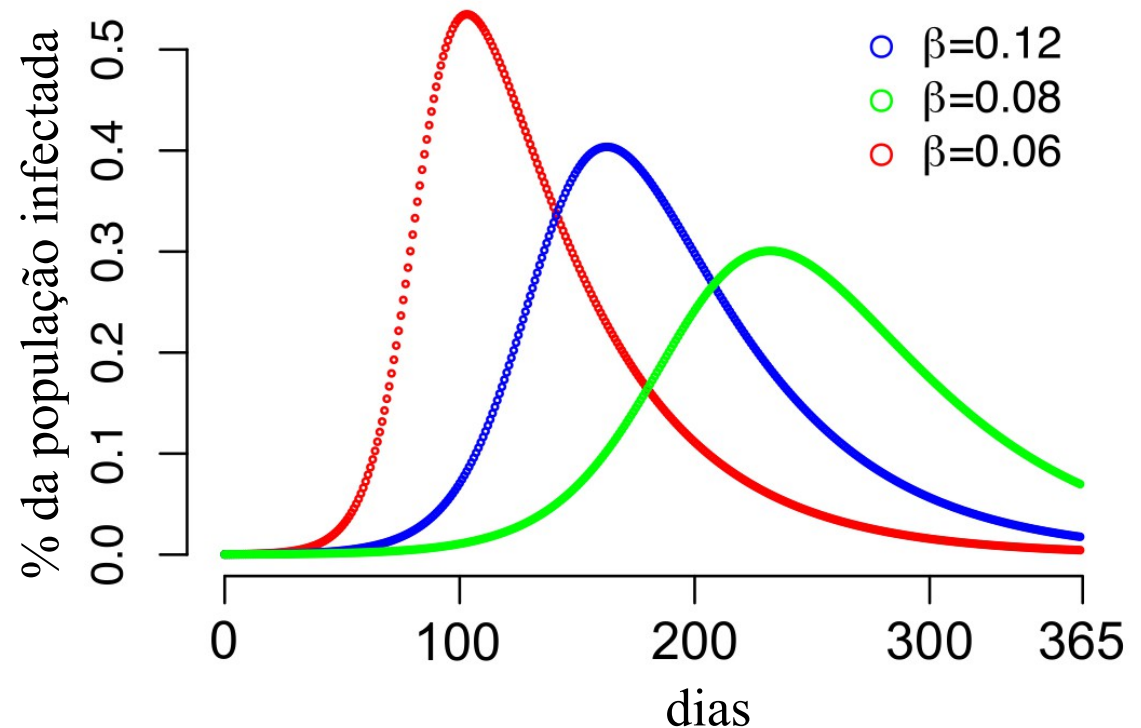


Funções, modelos e distribuições:

Uma **função** expressa a relação entre quantidades:

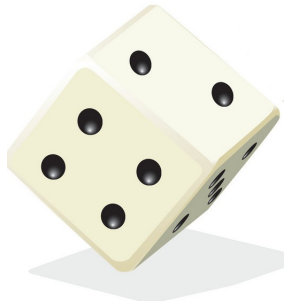
SIR model: Modela novas infecções como resultado do contato entre membros da população infectada (I) e membros suscetíveis (S), assumindo uma taxa de contato e de recuperação .

$$\frac{d}{dt} I = \beta IS - \gamma Y$$



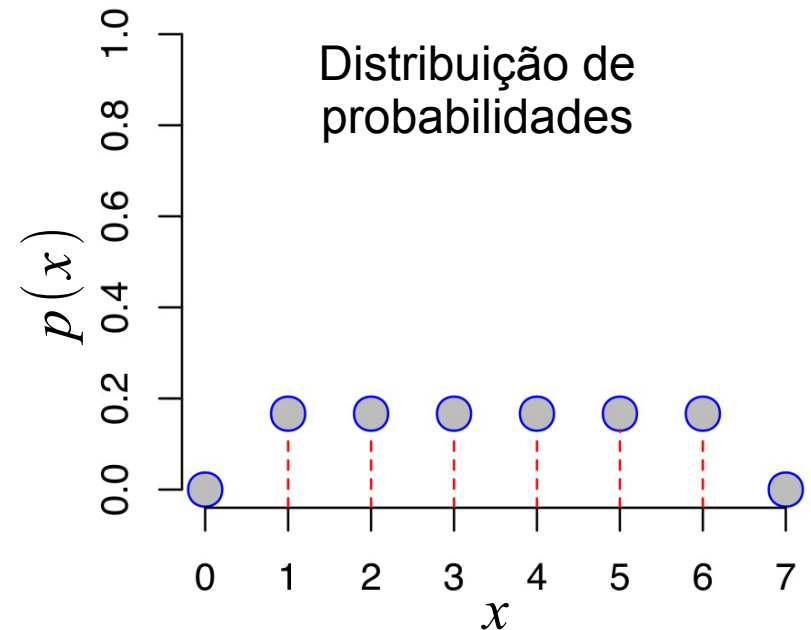
Funções, modelos e distribuições:

Quando uma função expressa a relação entre a probabilidades de eventos temos:



$$p(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{6} & 1 \leq x \leq 6 \\ 0 & x > 6 \end{cases}$$

Função probabilística



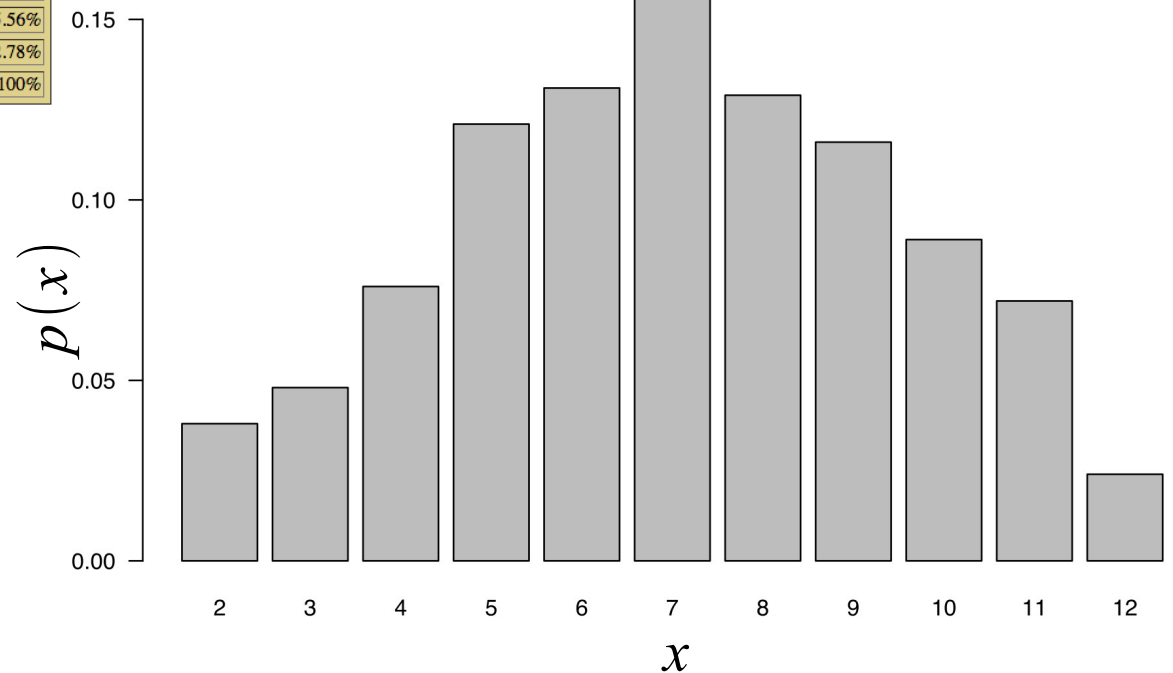
Funções, modelos e distribuições:

Quando uma função expressa a relação entre a probabilidades de eventos temos:



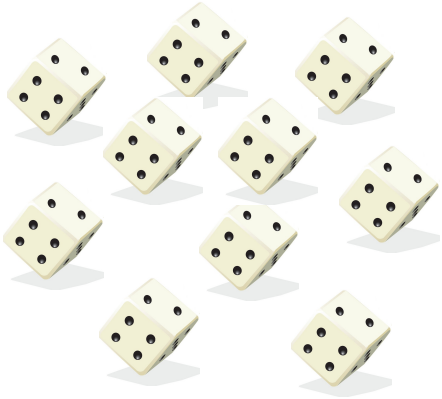
Total	Number of combinations	Probability
2	1	2.78%
3	2	5.56%
4	3	8.33%
5	4	11.11%
6	5	13.89%
7	6	16.67%
8	5	13.89%
9	4	11.11%
10	3	8.33%
11	2	5.56%
12	1	2.78%
Total	36	100%

$$\sum_x P X(x) = 1$$



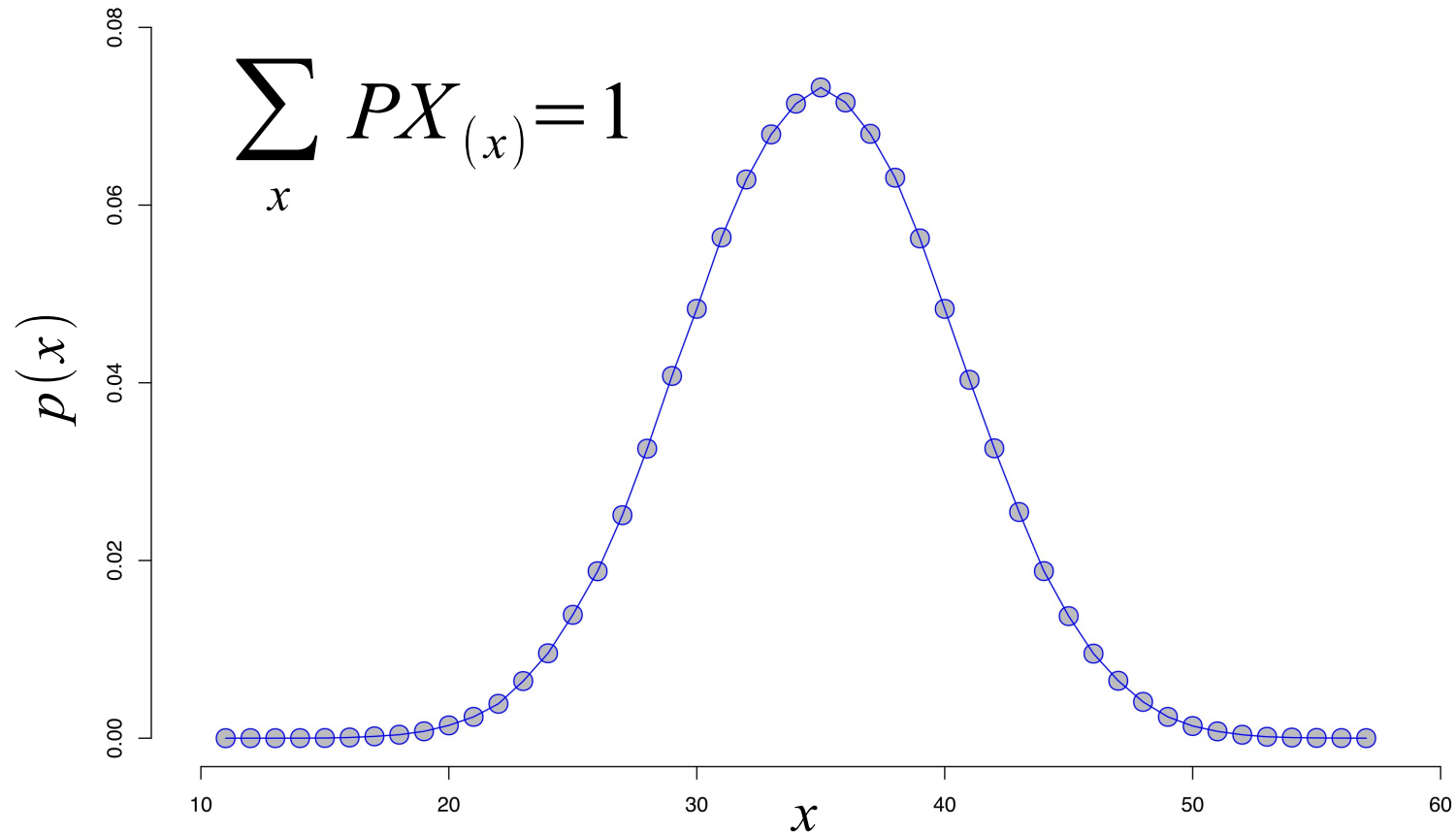
Funções, modelos e distribuições:

Quando uma função expressa a relação entre a probabilidade de eventos temos:



Simulação de 10 dados e 1.000.000 de jogadas

$$\sum_x P X_{(x)} = 1$$

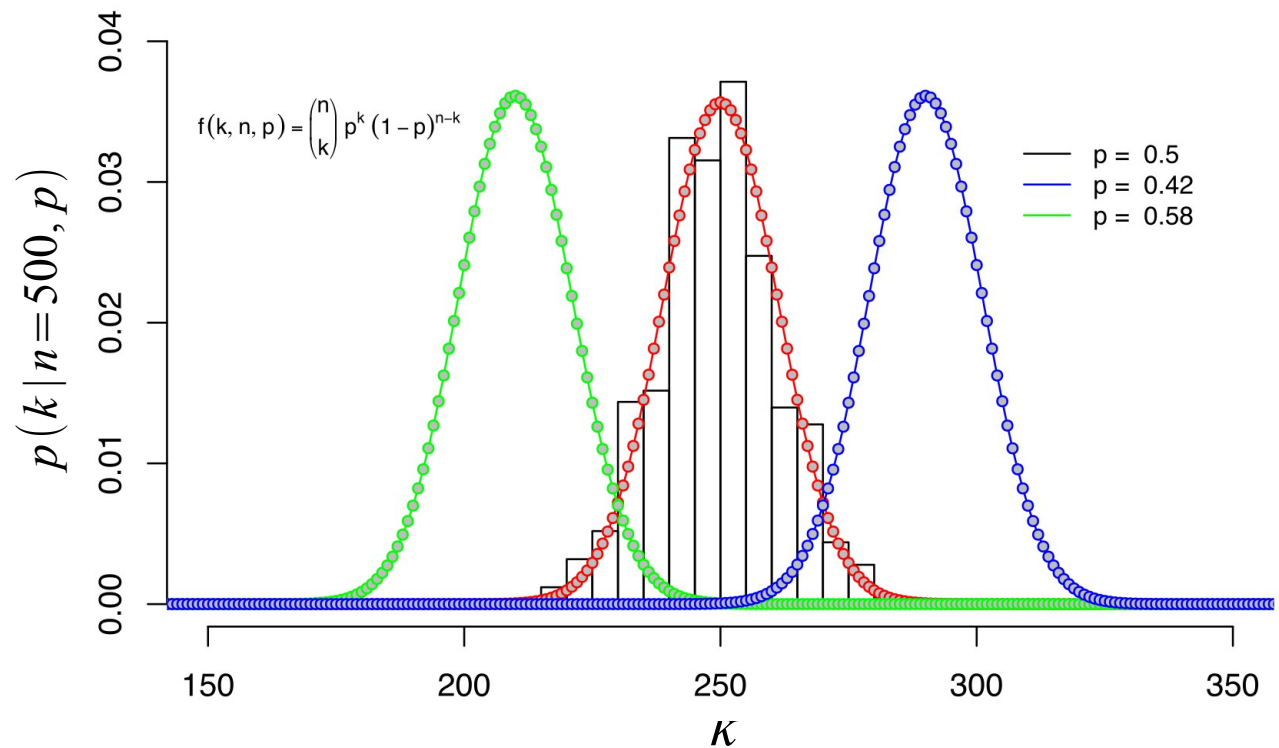


Funções, modelos e distribuições:

Funções diferentes descrevem processos e relações distintas:



Exemplo: **Distribuição binomial** é a distribuição de probabilidade discreta do número de sucessos (k) numa sequência de n tentativas.

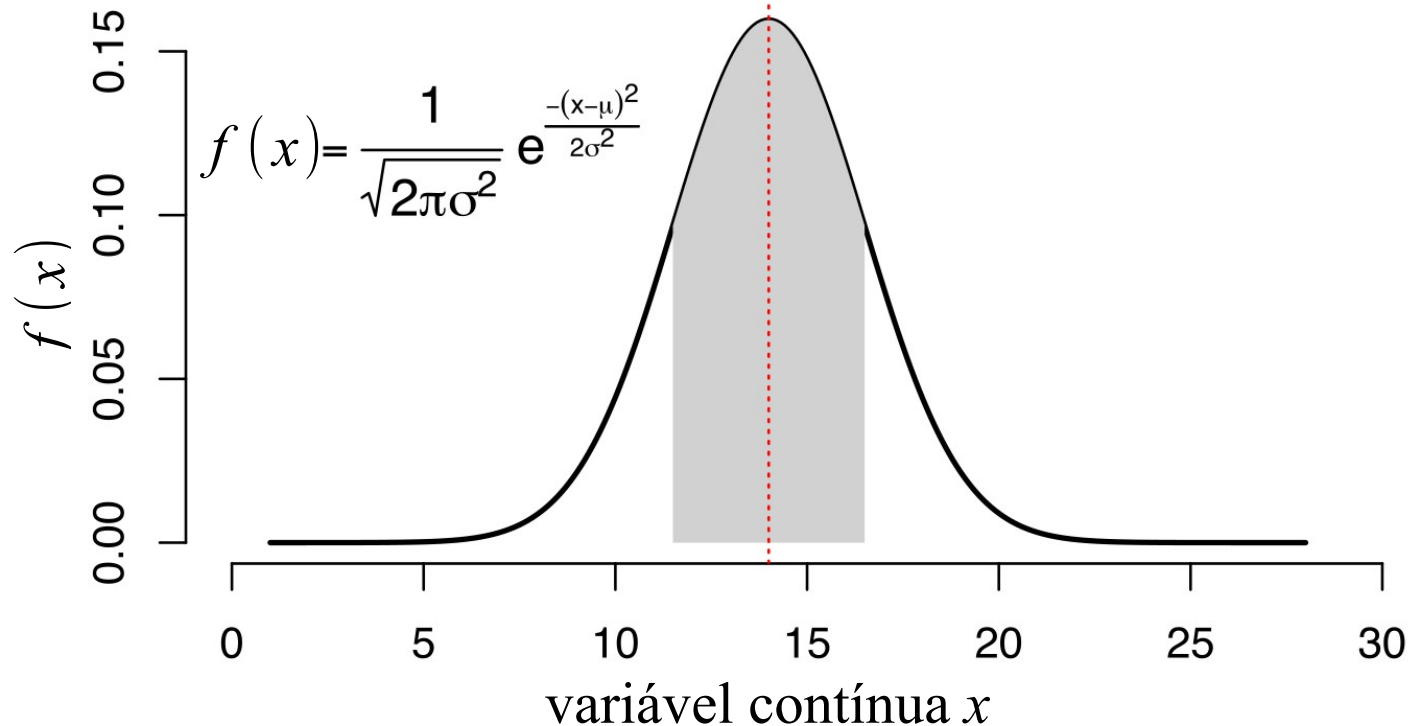


Funções, modelos e distribuições:

Funções diferentes descrevem processos e relações distintas:

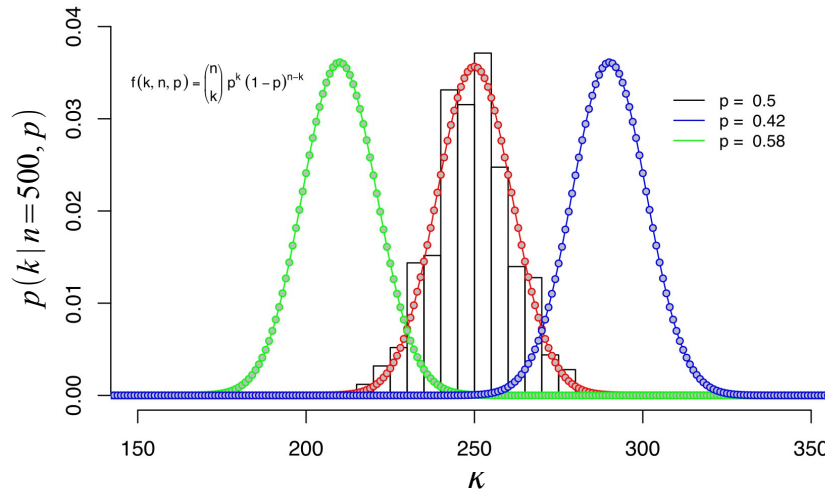


Exemplo: **Distribuição normal** descreve o comportamento de variáveis contínuas com variância simétrica em torno da média.



Funções, modelos e distribuições:

Função Massa de Probabilidade

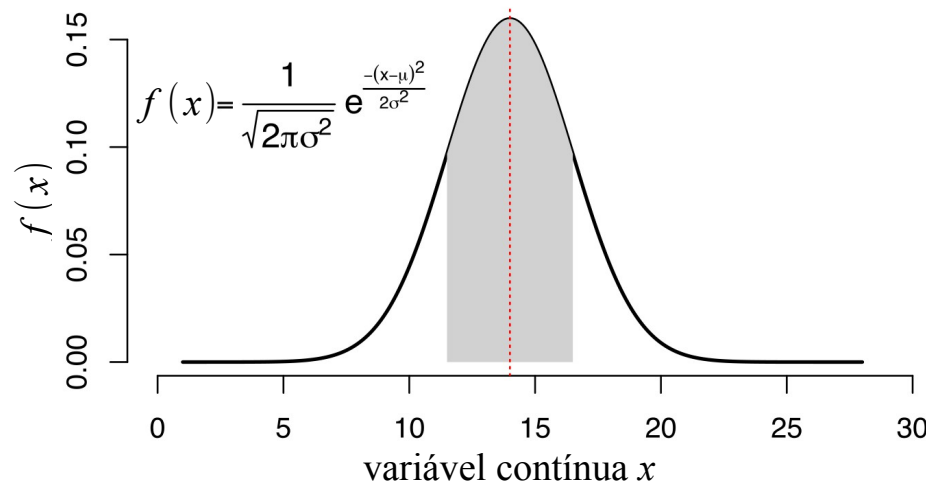


$$P(x \leq X \leq x+a) = \sum_{x: x \leq x \leq x+a} PX(x)$$

$$PX(x) \geq 0$$

$$\sum_x PX(x) = 1$$

Função Densidade de Probabilidade



$$P(x \leq X \leq x+a) = \int_x^{x+a} fX(x) dx$$

$$fX(x) \geq 0$$

$$\int_{-\infty}^{\infty} fX(x) dx = 1$$

Estimativas por Máxima Verossimilhança:

Método utilizado para estimar valores para parâmetros () de modelos na medida em que os valores encontrados visam maximizar a verossimilhança de que o processo descrito pelo modelo produziu os dados observados.

Função → Modelo

$$f(k, n, p) = \binom{n}{k} p^k (1 - p)^{n-k}$$

$$\theta = \begin{cases} k = \text{número de sucessos} \\ n = \text{número de eventos} \\ p = \text{probabilidade de } k \end{cases}$$

Estimativas por Máxima Verossimilhança:

Um **modelo** estatístico é a formalização matemática da relação entre variáveis que correspondem a observações potenciais que inclui a descrição das incertezas sobre estas observações devido à variabilidade natural, erros ou informação incompleta.

Parâmetros estatísticos podem ser vistos como as características numéricas de um modelo.

Função \rightarrow Modelo

$$f(k, n, p) = \binom{n}{k} p^k (1 - p)^{n-k}$$

$$\theta = \begin{cases} k = \text{número de sucessos} \\ n = \text{número de eventos} \\ p = \text{probabilidade de } k \end{cases}$$

Estimativas por Máxima Verossimilhança:



Sir Ronald Aylmer Fisher
(17 February 1890 – 29 July
1962)

$$L(\theta | x) \propto P(x | \theta)$$

A verossimilhança (L) de um conjunto de parâmetros (θ), dado alguma observação (x) é proporcional a probabilidade daquela observação ter ocorrido dados os valores daqueles parâmetros.

Estimativas por Máxima Verossimilhança:

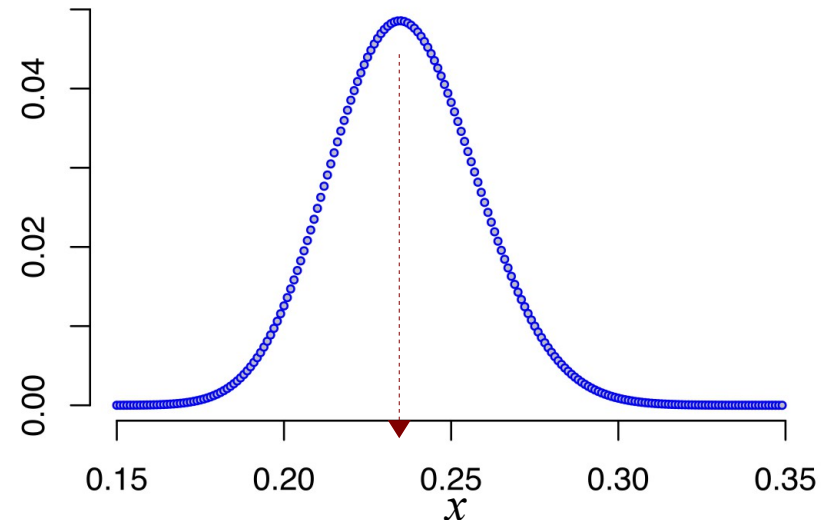


Sir Ronald Aylmer Fisher
(17/02/1890 – 29/07/1962)

$$L(\theta | x) \propto P(x | \theta)$$

Fisher (1912) empregou a função de verossimilhança $f(\theta|x)$ com a ideia de que o(s) valor(es) de θ que maximizam a probabilidade dos dados observados (x) seria um bom estimador de θ .

$$f(\theta|x) = L(\theta|x)$$



Estimativas por Máxima Verossimilhança:



Suponha que eu tenha jogado uma moeda 375 vezes e tenha obtido 88 caras. Como eu poderia estimar a probabilidade de obter caras ($P(\text{ca})$) com esta moeda usando verossimilhança.

1. Qual é a função (modelo) de probabilidades que descreve mais adequadamente o processo responsável pelo resultado?
2. Qual é o valor de $P(\text{ca})$ que maximiza essa função?

Estimativas por Máxima Verossimilhança:



Suponha que eu tenha jogado uma moeda 375 vezes e tenha obtido 88 caras. Como eu poderia estimar a probabilidade de obter caras ($P(\text{ca})$) com esta moeda usando verossimilhança.

1. Qual é a função (modelo) de probabilidades que descreve mais adequadamente o processo responsável pelo resultado?

A **distribuição binomial** é a distribuição de probabilidade discreta do número de sucessos numa sequência de tentativas tais que:

i. Cada tentativa tem exclusivamente como resultado duas possibilidades, sucesso ou fracasso;

ii. Cada tentativa é independente das demais;

A probabilidade de sucesso a cada tentativa permanece constante independente das demais;

A variável de interesse, ou pretendida, é o número de sucessos nas tentativas.

Estimativas por Máxima Verossimilhança:

Função da distribuição binomial:

$$f(k, n, p) = \binom{n}{k} p^k (1-p)^{n-k}$$

onde k é no número de eventos de sucesso (i.e., pergunta de aluna), n é o número de perguntas feitas, e p é a probabilidade deste evento ocorrer, neste caso, ele é o parâmetro (θ) que queremos estimar.



Estimativas por Máxima Verossimilhança:

Suponha que eu tenha jogado uma moeda 375 vezes e tenha obtido 88 caras. Como eu poderia estimar a probabilidade de obter caras ($P(\text{ca})$) com esta moeda usando verossimilhança.



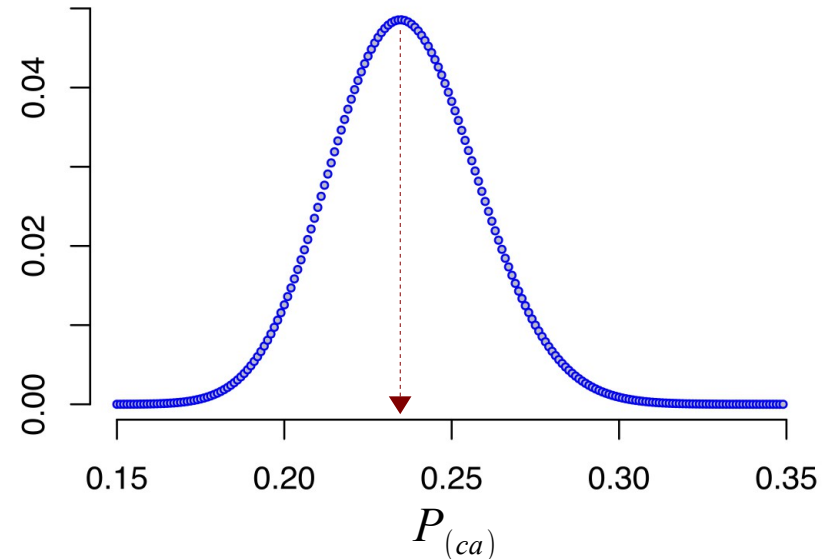
2. Qual é o valor de $P(\text{ca})$ que maximiza essa função?

$$f(k, n, p) = \binom{n}{k} p^k (1-p)^{n-k}$$

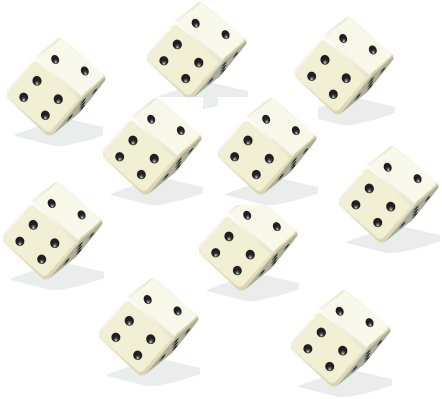
$$f(P_{(ca)}, n, k) = L(P_{(ca)} | n, k)$$

$$\max L(P_{(ca)} | 375, 88) = 0.04856$$

$$\text{para } P_{(ca)} = 0.235$$



Estimativas por Máxima Verossimilhança:



Testando hipóteses:

Suponha que você jogue 10 dados simultaneamente e todos eles repousam com a face 3 para cima. Você quer testar duas hipóteses:

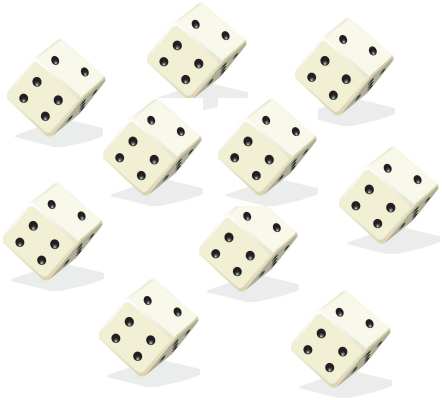
$H_1 =$ dados honestos

$H_2 =$ dados viciados

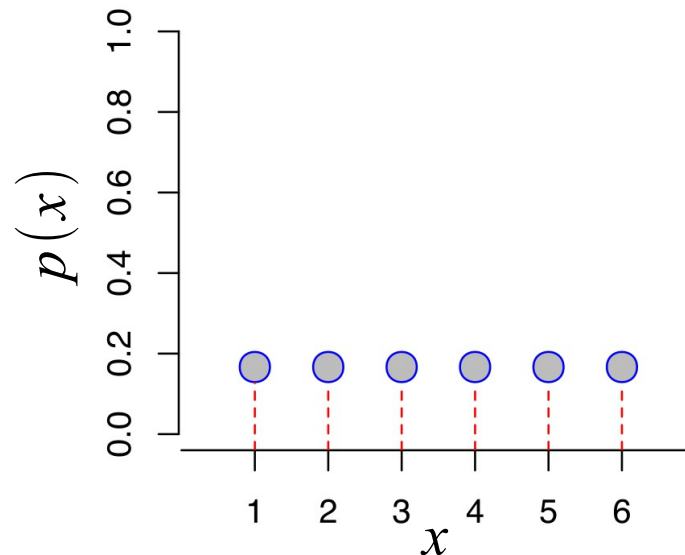
Estimativas por Máxima Verossimilhança:

Testando hipóteses:

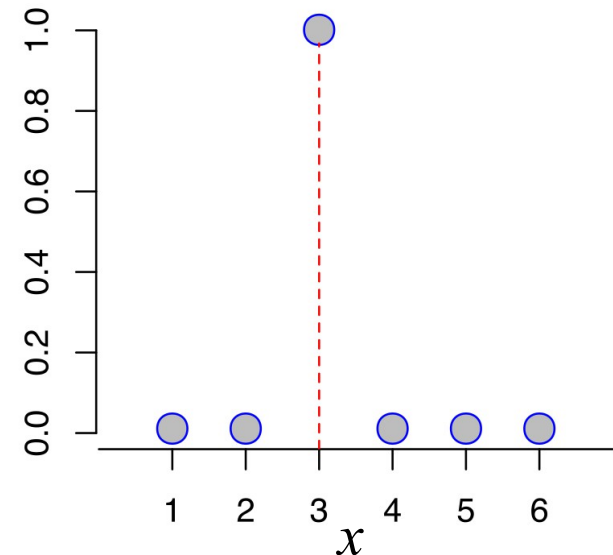
Suponha que você jogue 10 dados simultaneamente e todos eles repousam com a face 3 para cima. Você quer testar duas hipóteses:



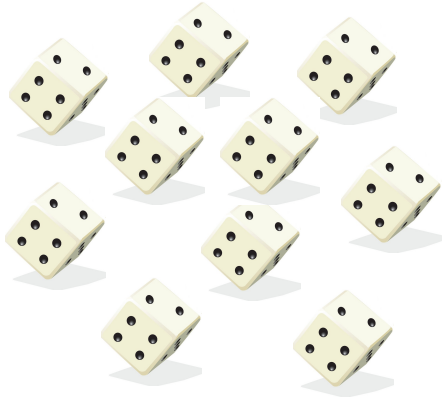
$H_1 =$ dados honestos



$H_2 =$ dados viciados

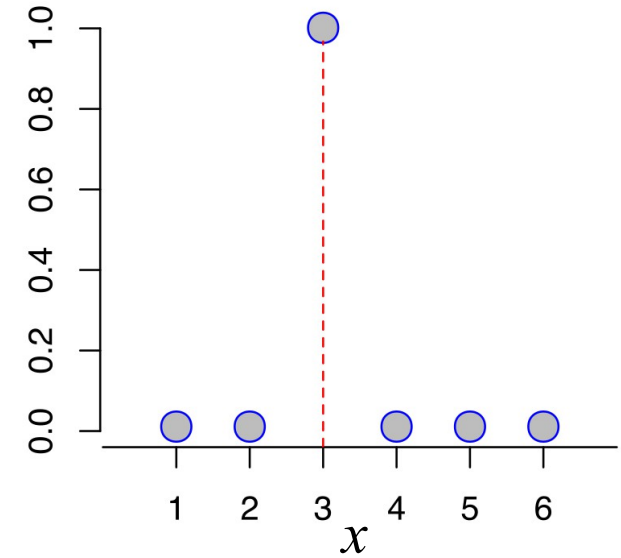
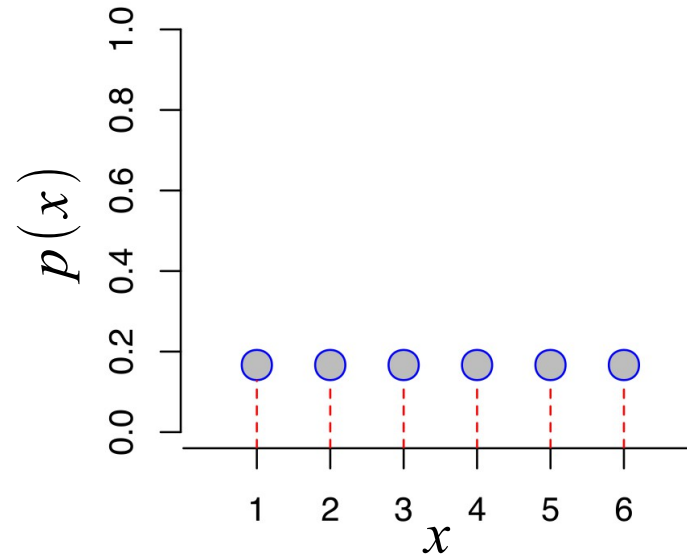


Estimativas por Máxima Verossimilhança:



$H_1 = \text{dados honestos}$

$H_2 = \text{dados viciados}$



$$P_{(10 \times 3 | H_1)} = \left(\frac{1}{6}\right)^{10} = 1.653817e-08$$

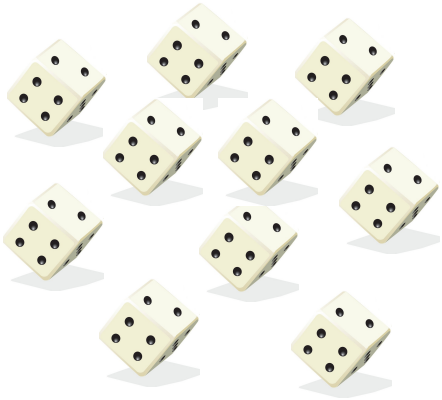
$$P_{(10 \times 3 | H_2)} = 1^{10} = 1.0$$

$$L_{(H_1 | 10 \times 3)} = 1.653817e-08$$

$$L_{(H_2 | 10 \times 3)} = 1.0$$

$$L_{(H_2 | 10 \times 3)} > L_{(H_1 | 10 \times 3)}$$

Estimativas por Máxima Verossimilhança:



Testando hipóteses:

Suponha que você jogue 10 dados simultaneamente e todos eles repousam com a face 3 para cima. Você quer testar duas hipóteses:

$H_1 =$ dados honestos

$H_2 =$ dados viciados

$$L_{(H_2|10 \times 3)} > L_{(H_1|10 \times 3)}$$

Verossimilhança: $L(\theta | x) \propto P(x | \theta)$

Probabilidades: $P(\theta | x) \neq P(x | \theta)$

Máxima Verossimilhança vs. Probabilidades



Considere:

x = Professor ouve barulho na goiabeira

θ = *Jesus subiu na goiabeira*

$$L(\theta | x) \propto P(x | \theta)$$

$L(\text{Jesus} | \text{barulho}) \propto P(\text{barulho} | \text{Jesus}) = \text{Alta}$

$P(\text{Jesus} | \text{barulho}) = \text{Baixa}$



O MDH adverte: a Ministra de Estado da Mulher, da Família e dos Direitos Humanos discorda!

Máxima Verossimilhança e Inferência filogenética

"Felsenstein Zone" Trees

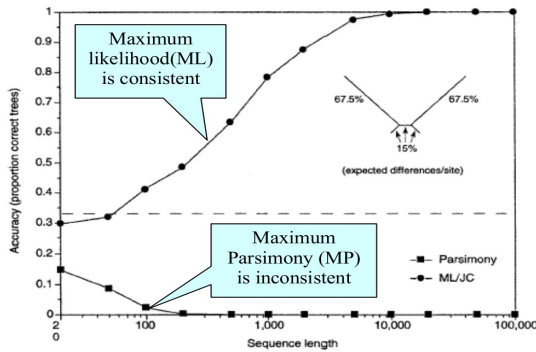


Fig. 3, p. 531, from Swofford et al. 2001. Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. *Systematic Biology* 50: 525-539.

		↓ ↓ ↓
Dados	sp.X	CTGGCTACGT
	sp.A	TGGAGTAAGT
	sp.B	CCTAGCAAGT
	sp.C	CCTGATTGCA

Codificação

Parcimônia:

EVIDÊNCIAS: transformação de estados de caracteres.

- 1a. posição: C ↔ T
- 3a. posição: G ↔ T
- 10a. posição: T ↔ A

Seleção

Menor distância
Patrística

Probabilística (ML):

EVIDÊNCIAS: transformação de estados de caracteres.

- 1a. Posição: C ↔ T
- 3a. posição: G ↔ T
- 10a. posição: T ↔ A

$$\max L(\text{dados} | Q, T, v)$$

Onde:

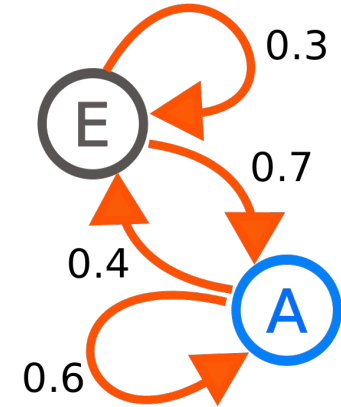
Q é a modelo de substituição,
T é a topologia e
v são os comprimentos de ramos.

Máxima Verossimilhança e Inferência filogenética

Modelos de substituição (e.g., nucleotídeos):



Uma cadeia de Markov é um modelo estocástico que descreve uma sequência de eventos possíveis para os quais a probabilidade de cada evento depende somente do estado obtido no evento anterior.



$$Q = P_{(transf.)} = \begin{pmatrix} P_{(AA|v)} & P_{(CA|v)} & P_{(GA|v)} & P_{(TA|v)} \\ P_{(AC|v)} & P_{(CC|v)} & P_{(GC|v)} & P_{(TC|v)} \\ P_{(AG|v)} & P_{(CG|v)} & P_{(GG|v)} & P_{(TG|v)} \\ P_{(AT|v)} & P_{(CT|v)} & P_{(GT|v)} & P_{(TT|v)} \end{pmatrix}$$

Máxima Verossimilhança e Inferência filogenética

Modelos de substituição (e.g., nucleotídeos):

Model	Exchangeability parameters	Base frequency parameters	Reference
JC69 (or JC)	$a = b = c = d = e = f$	$\pi_A = \pi_C = \pi_G = \pi_T = 0.25$	Jukes and Cantor (1969) ^[13]
F81	$a = b = c = d = e = f$	all π_i values free	Felsenstein (1981) ^[14]
K2P (or K80)	$a = c = d = f$ (transversions), $b = e$ (transitions)	$\pi_A = \pi_C = \pi_G = \pi_T = 0.25$	Kimura (1980) ^[15]
HKY85	$a = c = d = f$ (transversions), $b = e$ (transitions)	all π_i values free	Hasegawa et al. (1985) ^[16]
K3ST (or K81)	$a = f$ (γ transversions), $c = d$ (β transversions), $b = e$ (transitions)	$\pi_A = \pi_C = \pi_G = \pi_T = 0.25$	Kimura (1981) ^[17]
TN93	$a = c = d = f$ (transversions), b ($A \leftrightarrow G$ transitions), e ($C \leftrightarrow T$ transitions)	all π_i values free	Tamura and Nei (1993) ^[18]
SYM	all exchangeability parameters free	$\pi_A = \pi_C = \pi_G = \pi_T = 0.25$	Zharkikh (1994) ^[19]
GTR (or REV ^[9])	all exchangeability parameters free	all π_i values free	Tavaré (1986) ^[7]

$$Q = P_{(transf.)} = \begin{pmatrix} P_{(AA|v)} & P_{(CA|v)} & P_{(GA|v)} & P_{(TA|v)} \\ P_{(AC|v)} & P_{(CC|v)} & P_{(GC|v)} & P_{(TC|v)} \\ P_{(AG|v)} & P_{(CG|v)} & P_{(GG|v)} & P_{(TG|v)} \\ P_{(AT|v)} & P_{(CT|v)} & P_{(GT|v)} & P_{(TT|v)} \end{pmatrix}$$

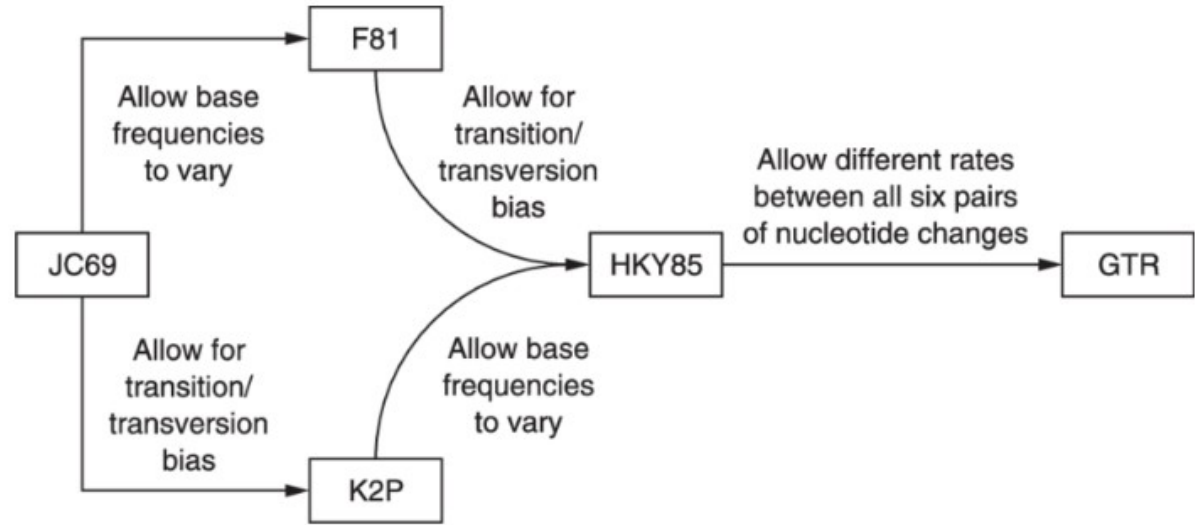
Máxima Verossimilhança e Inferência filogenética

Modelos de substituição (e.g., nucleotídeos):

Jukes-Cantor Substitution Probabilities

$\mu t = 0.25$

	A	C	G	T
A	0.5259	0.1580	0.1580	0.1580
C	0.1580	0.5259	0.1580	0.1580
G	0.1580	0.1580	0.5259	0.1580
T	0.1580	0.1580	0.1580	0.5259



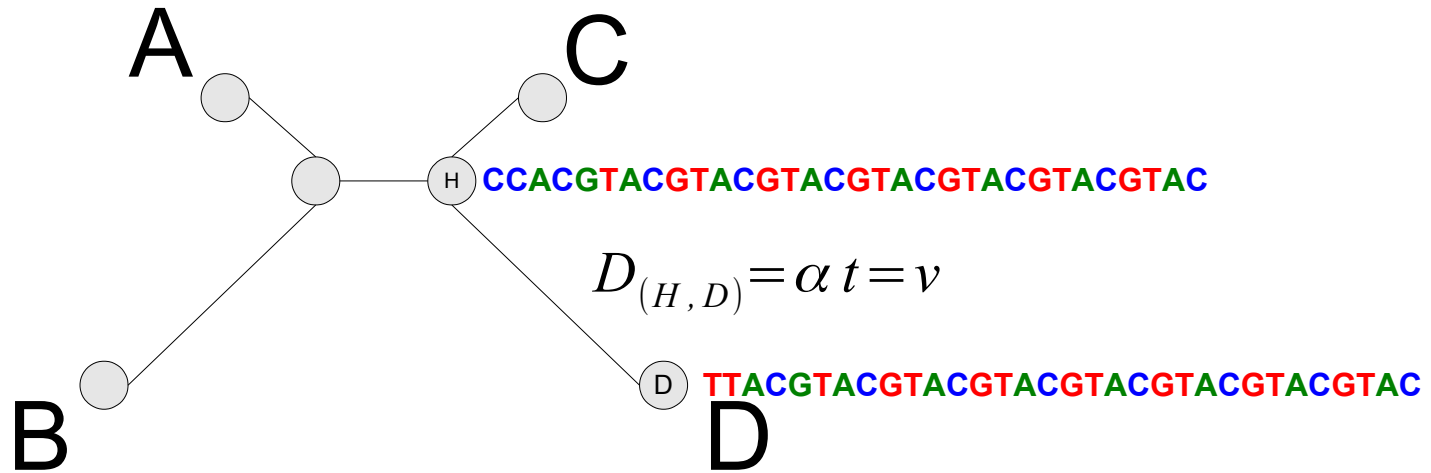
Kimura Two-Parameter Substitution Probabilities

$\mu t = 0.25 \quad \kappa = 2.0$

	A	C	G	T
A	0.4535	0.1580	0.2304	0.1580
C	0.1580	0.4535	0.1580	0.2304
G	0.2304	0.1580	0.4535	0.1580
T	0.1580	0.2304	0.1580	0.4535

Máxima Verossimilhança e Inferência filogenética

Estimativas de comprimento de ramo:



Diferença entre sequências:

CCACGTACGTACGTACGTACGTACGTACGTAC
**
TTACGTACGTACGTACGTACGTACGTACGTAC

Máxima Verossimilhança e Inferência filogenética

Estimativas de comprimento de ramo:

Diferença entre sequências:

CCACGTACGTACGTACGTACGTACGTACGTAC
**
TTACGTACGTACGTACGTACGTACGTACGTAC

probabilidade de não
mudança de estado*

$$P(ii) = \left(\frac{1}{4}\right) * \left(\frac{1}{4} + \frac{3}{4} e^{-4v}\right)$$

probabilidade de
mudança de estado*

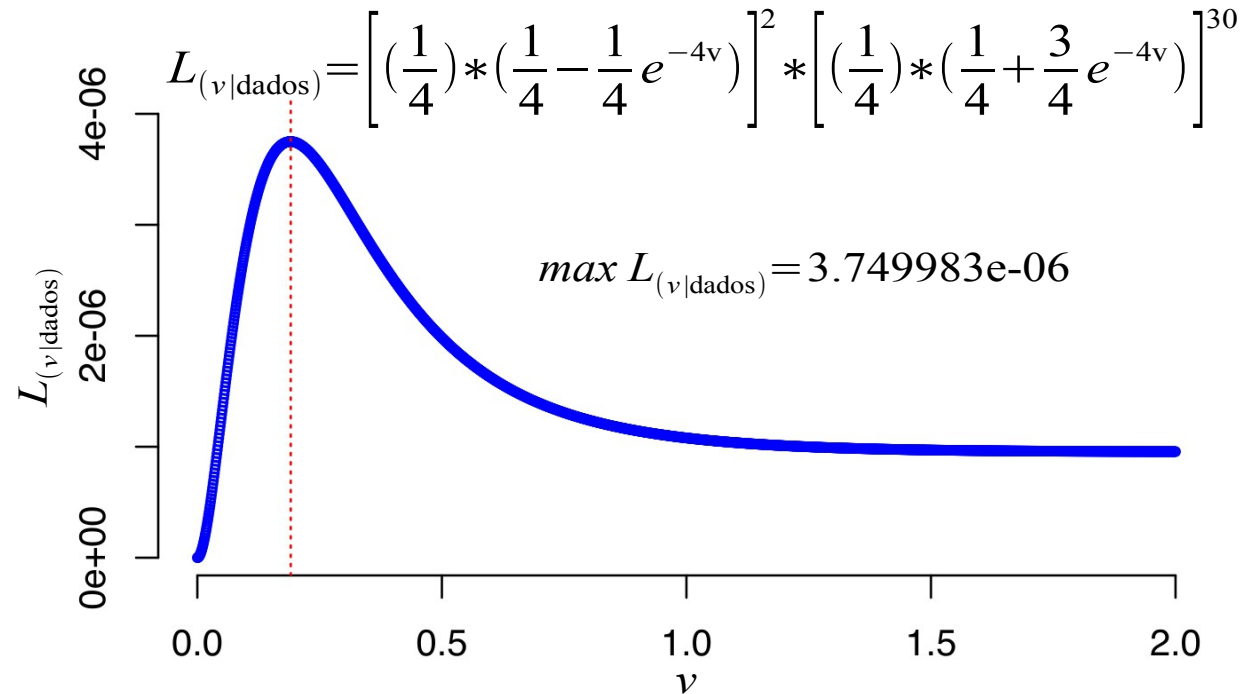
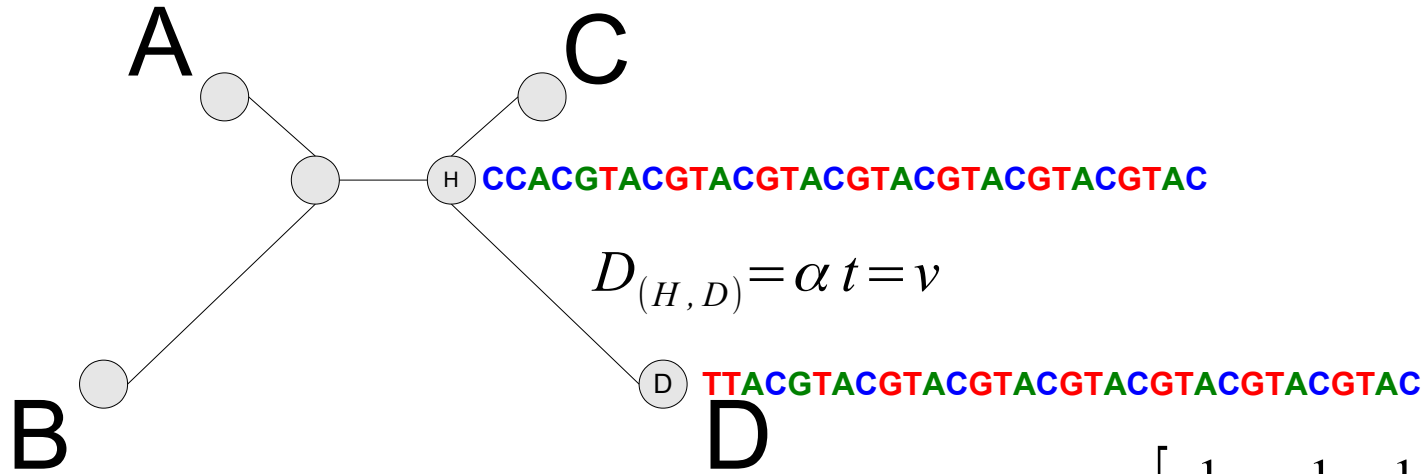
$$P(ij) = \left(\frac{1}{4}\right) * \left(\frac{1}{4} - \frac{1}{4} e^{-4v}\right)$$

$$L_{(v|\text{dados})} = \left[\left(\frac{1}{4}\right) * \left(\frac{1}{4} - \frac{1}{4} e^{-4v}\right) \right]^2 * \left[\left(\frac{1}{4}\right) * \left(\frac{1}{4} + \frac{3}{4} e^{-4v}\right) \right]^{30}$$

* baseado no modelo de JC69 (distribuição de Poisson) e frequências iguais de pares de base.

Máxima Verossimilhança e Inferência filogenética

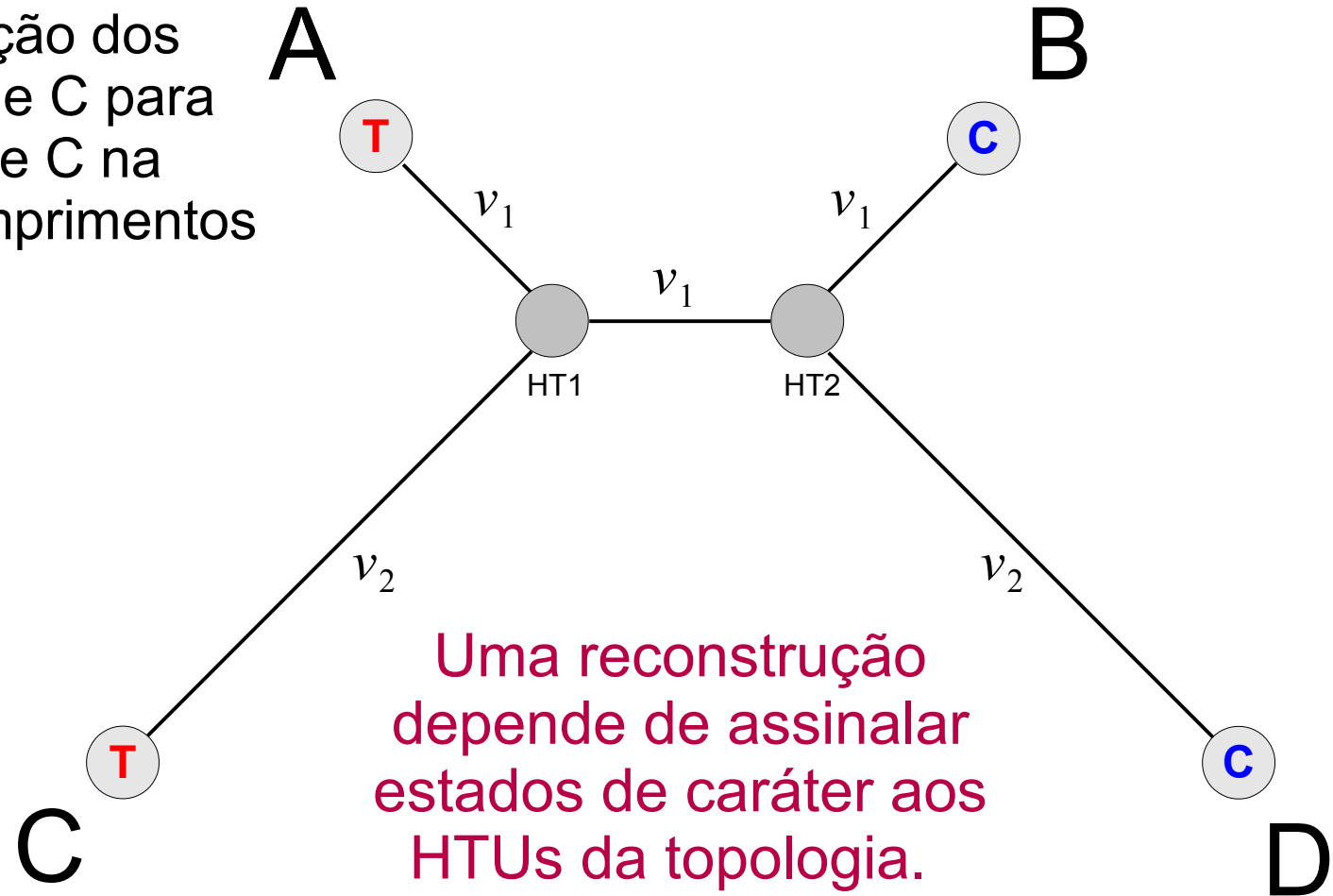
Estimativas de comprimento de ramo:

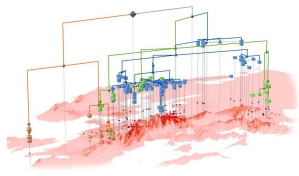


Máxima Verossimilhança e Inferência filogenética

Verossimilhança de reconstruções:

Considere a distribuição dos estados de caráter T e C para os terminais A, B, C, e C na topologia T_1 com comprimentos de ramos v_1 e v_2 .





Aula 07: Máxima Verossimilhança

Aula 7.02: Probabilidades;

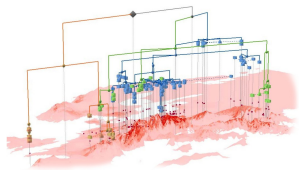
Aula 7.03: Funções, modelos e distribuições;

Aula 7.04: Máxima Verossimilhança (MLE);

Aula 7.05: MLE como critério de otimalidade;

Aula 7.06: Cálculo de probabilidade em topologias;

Aula 7.07: Considerações finais.



Aula 07: Máxima Verossimilhança

Aula 7.02: Probabilidades;

Aula 7.03: Funções, modelos e distribuições;

Aula 7.04: Máxima Verossimilhança (MLE);

Aula 7.05: MLE como critério de otimalidade;

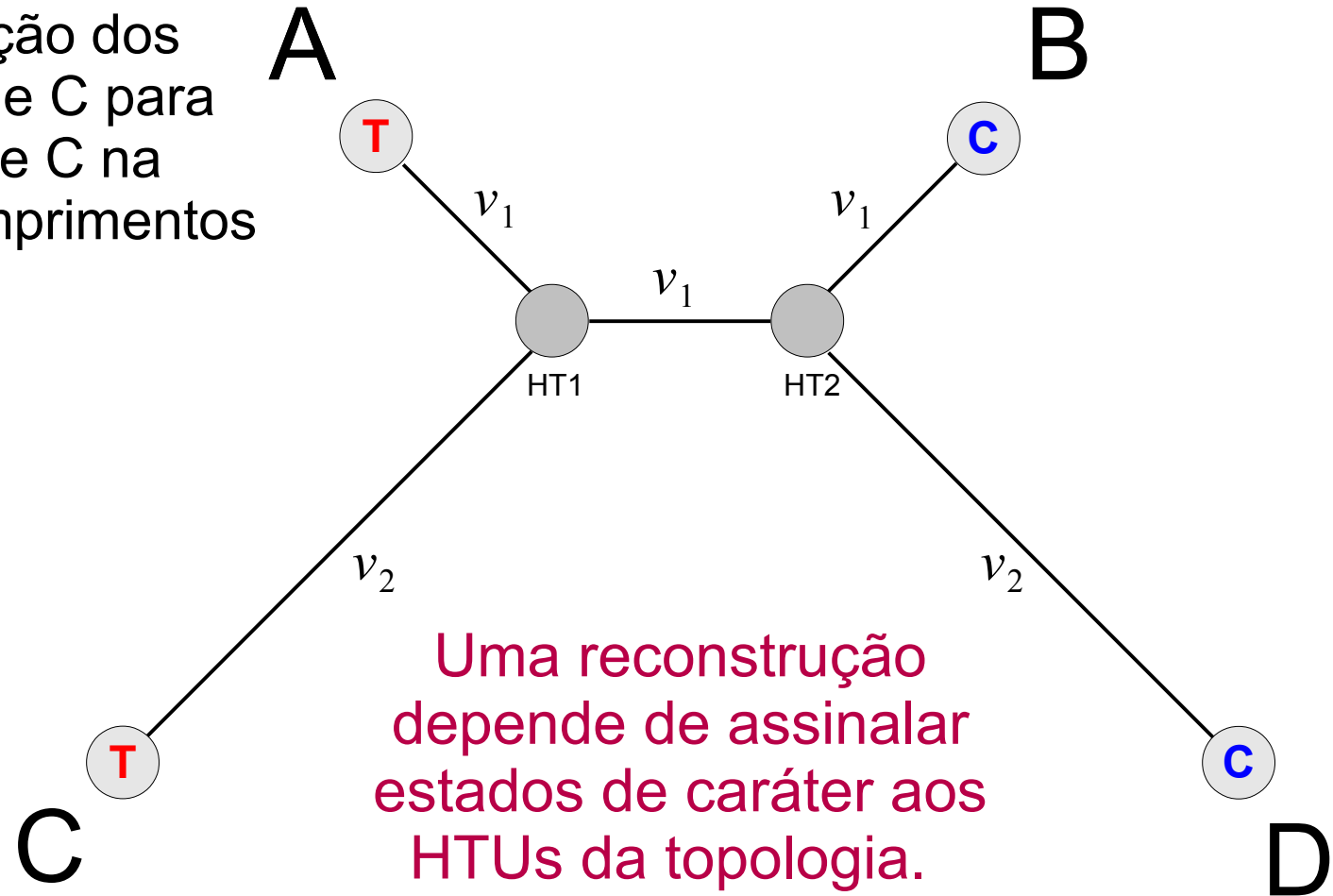
Aula 7.06: Cálculo de probabilidade em topologias;

Aula 7.07: Considerações finais.

Máxima Verossimilhança e Inferência filogenética

Verossimilhança de reconstruções:

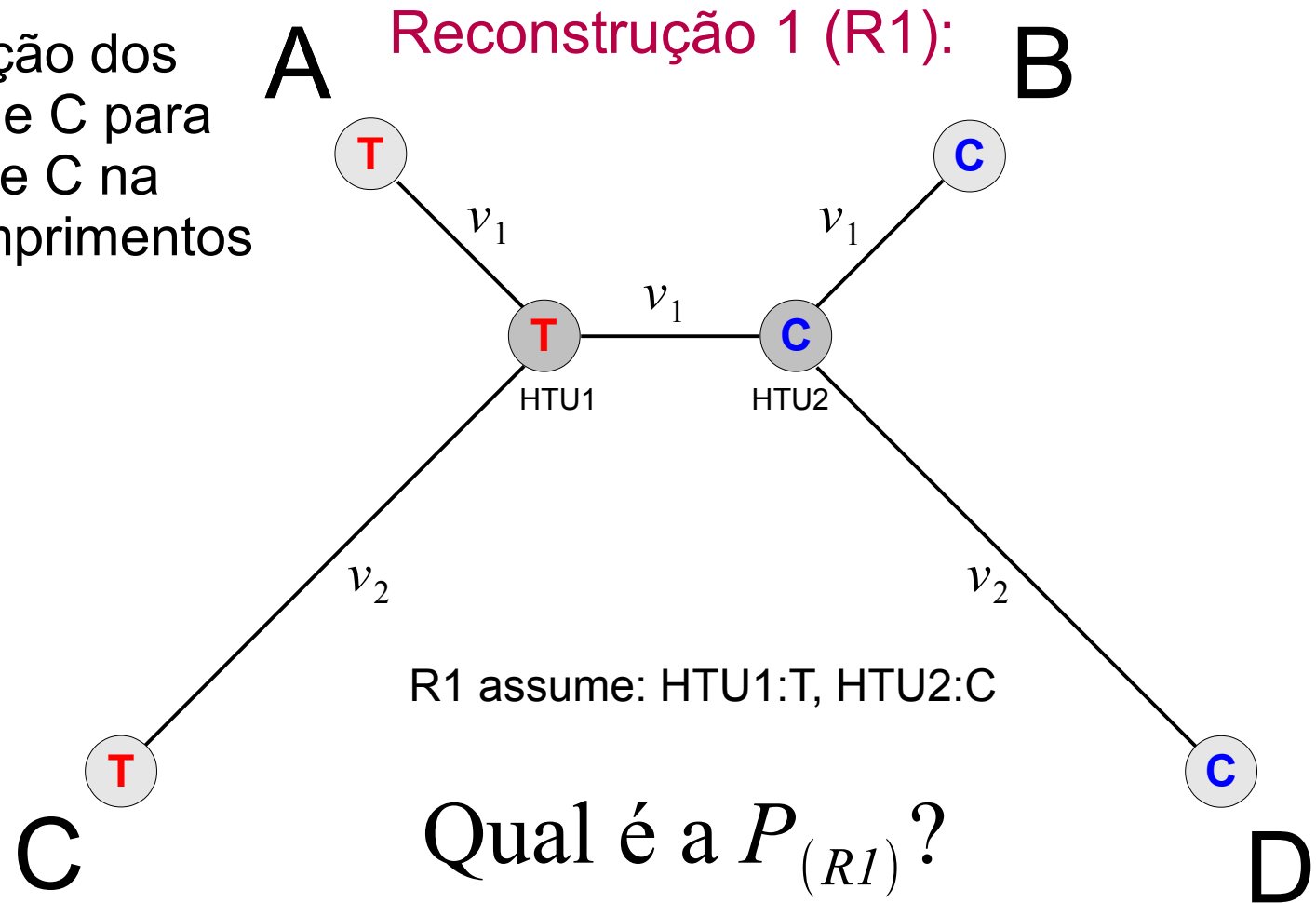
Considere a distribuição dos estados de caráter T e C para os terminais A, B, C, e C na topologia T_1 com comprimentos de ramos v_1 e v_2 .



Máxima Verossimilhança e Inferência filogenética

Verossimilhança de reconstruções:

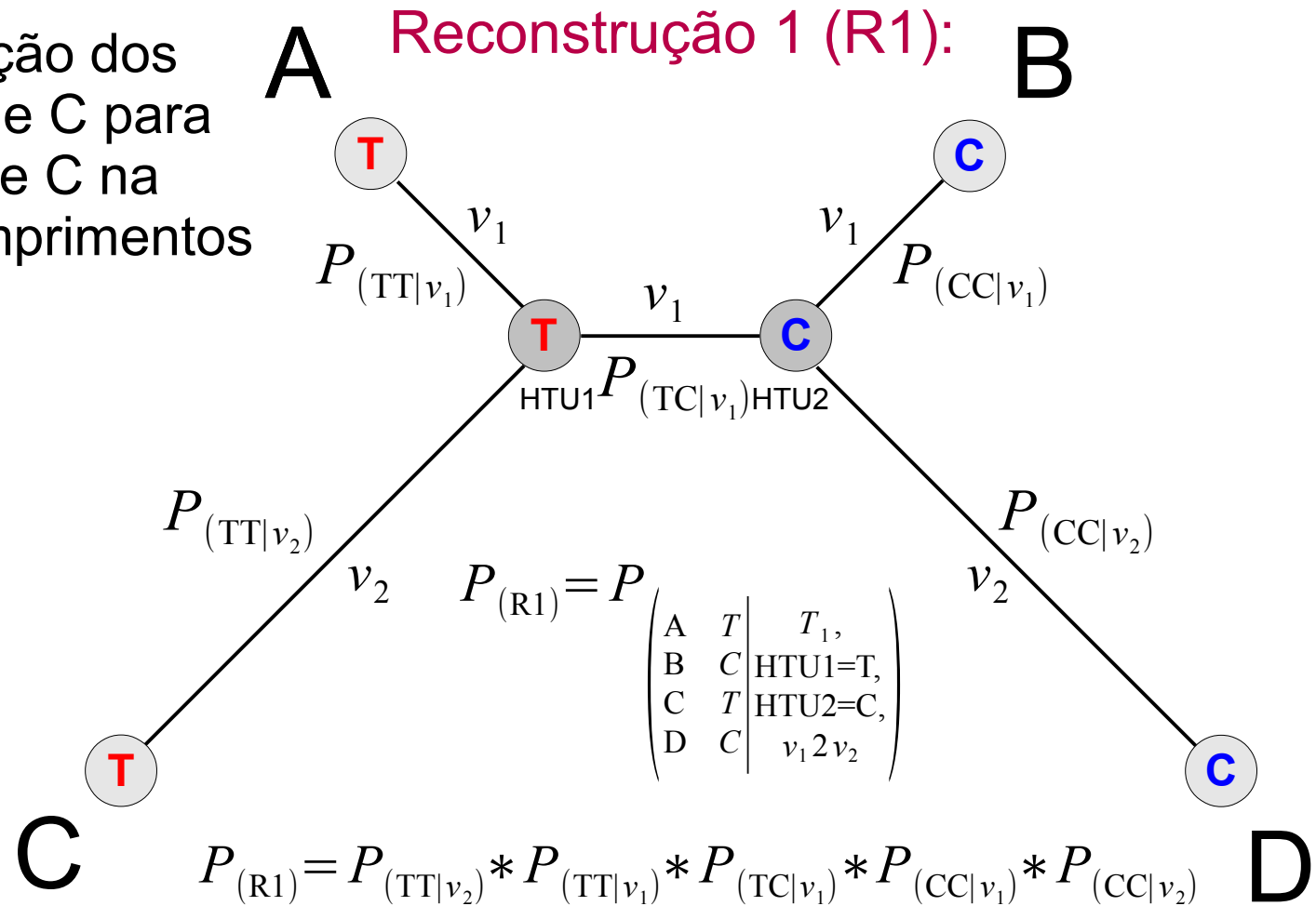
Considere a distribuição dos estados de caráter T e C para os terminais A, B, C, e C na topologia T_1 com comprimentos de ramos v_1 e v_2 .



Máxima Verossimilhança e Inferência filogenética

Verossimilhança de reconstruções:

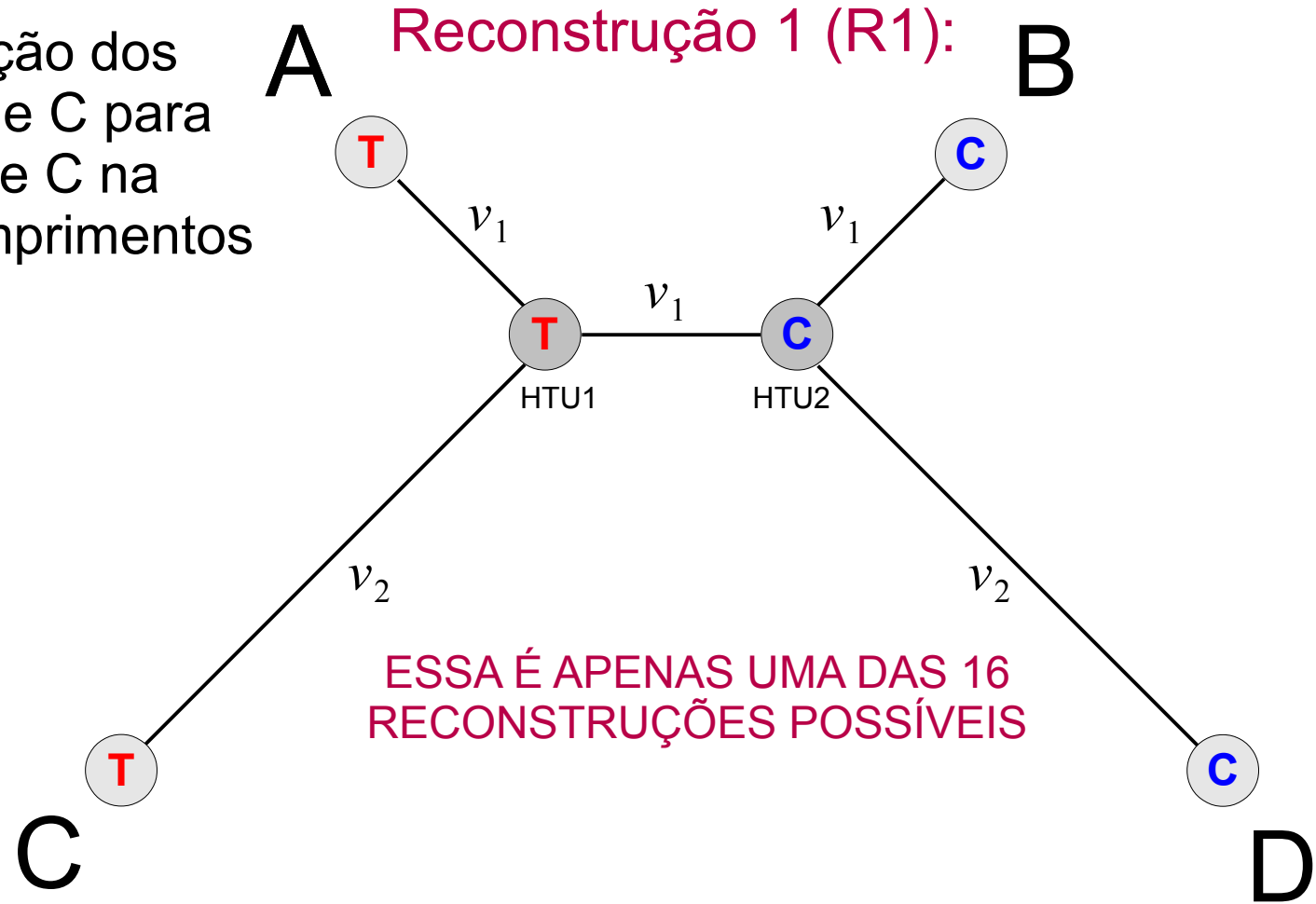
Considere a distribuição dos estados de caráter T e C para os terminais A, B, C, e C na topologia T_1 com comprimentos de ramos v_1 e v_2 .



Máxima Verossimilhança e Inferência filogenética

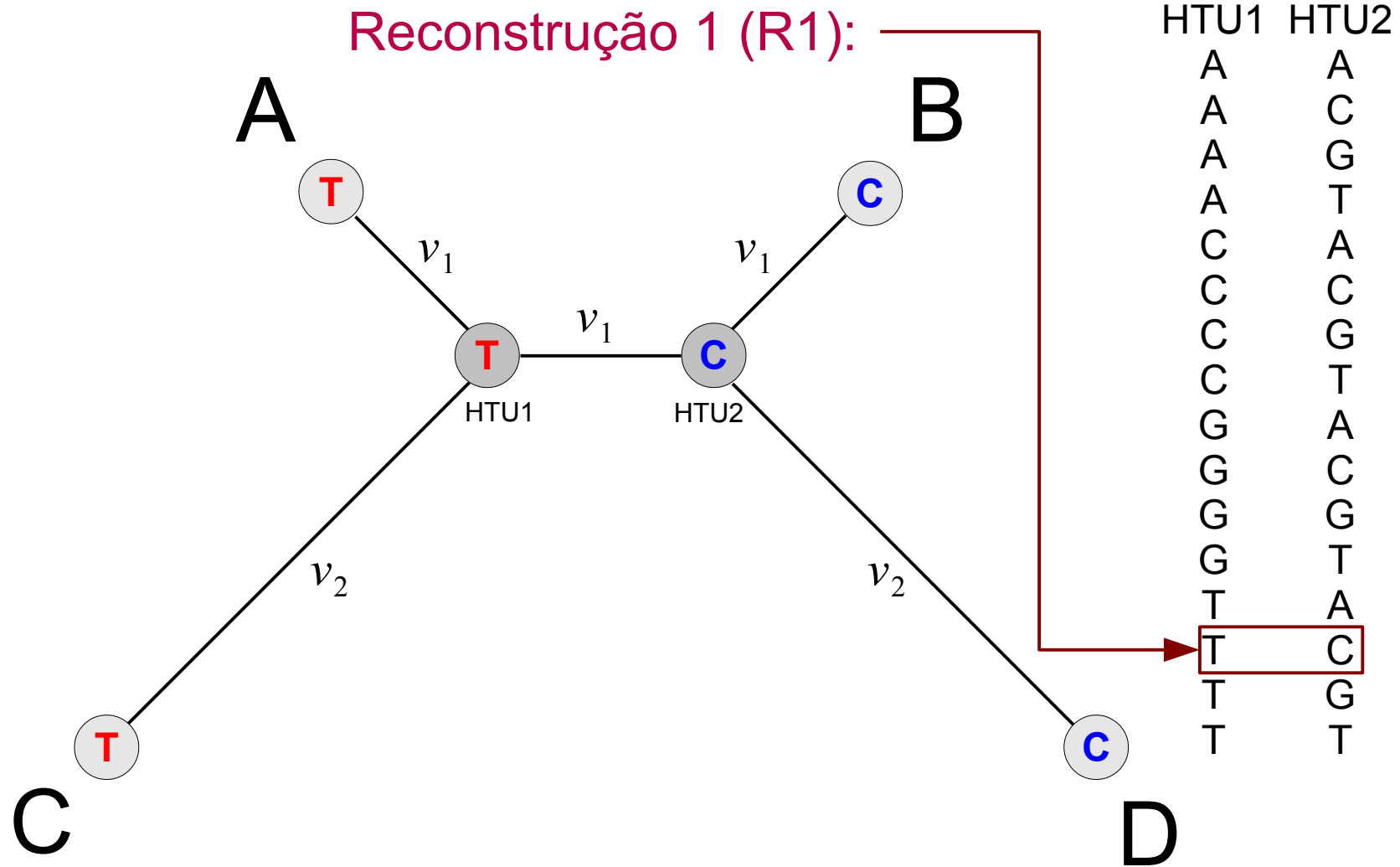
Verossimilhança de reconstruções:

Considere a distribuição dos estados de caráter T e C para os terminais A, B, C, e C na topologia T_1 com comprimentos de ramos v_1 e v_2 .



Máxima Verossimilhança e Inferência filogenética

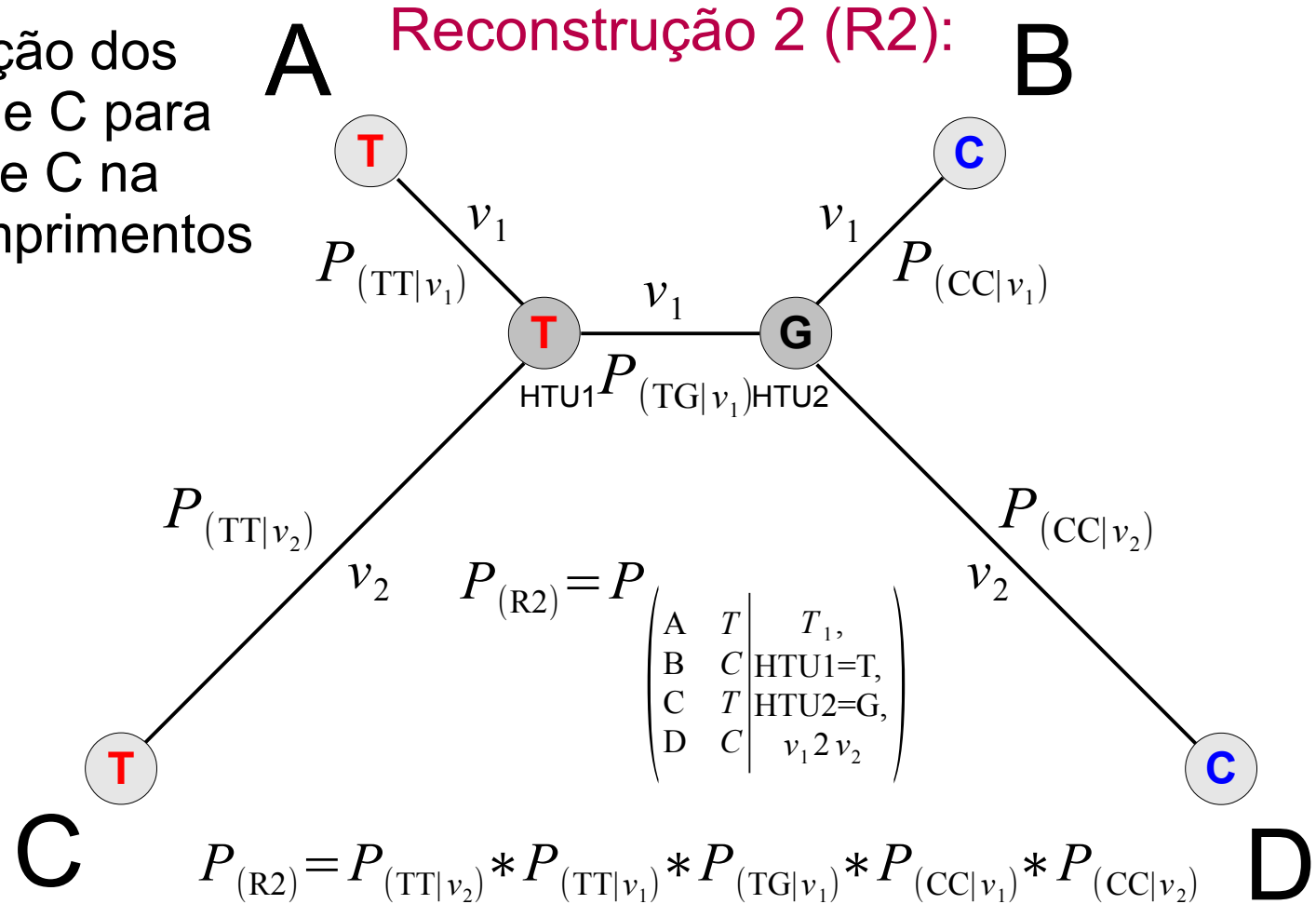
Verossimilhança de reconstruções:



Máxima Verossimilhança e Inferência filogenética

Verossimilhança de reconstruções:

Considere a distribuição dos estados de caráter T e C para os terminais A, B, C, e C na topologia T_1 com comprimentos de ramos v_1 e v_2 .



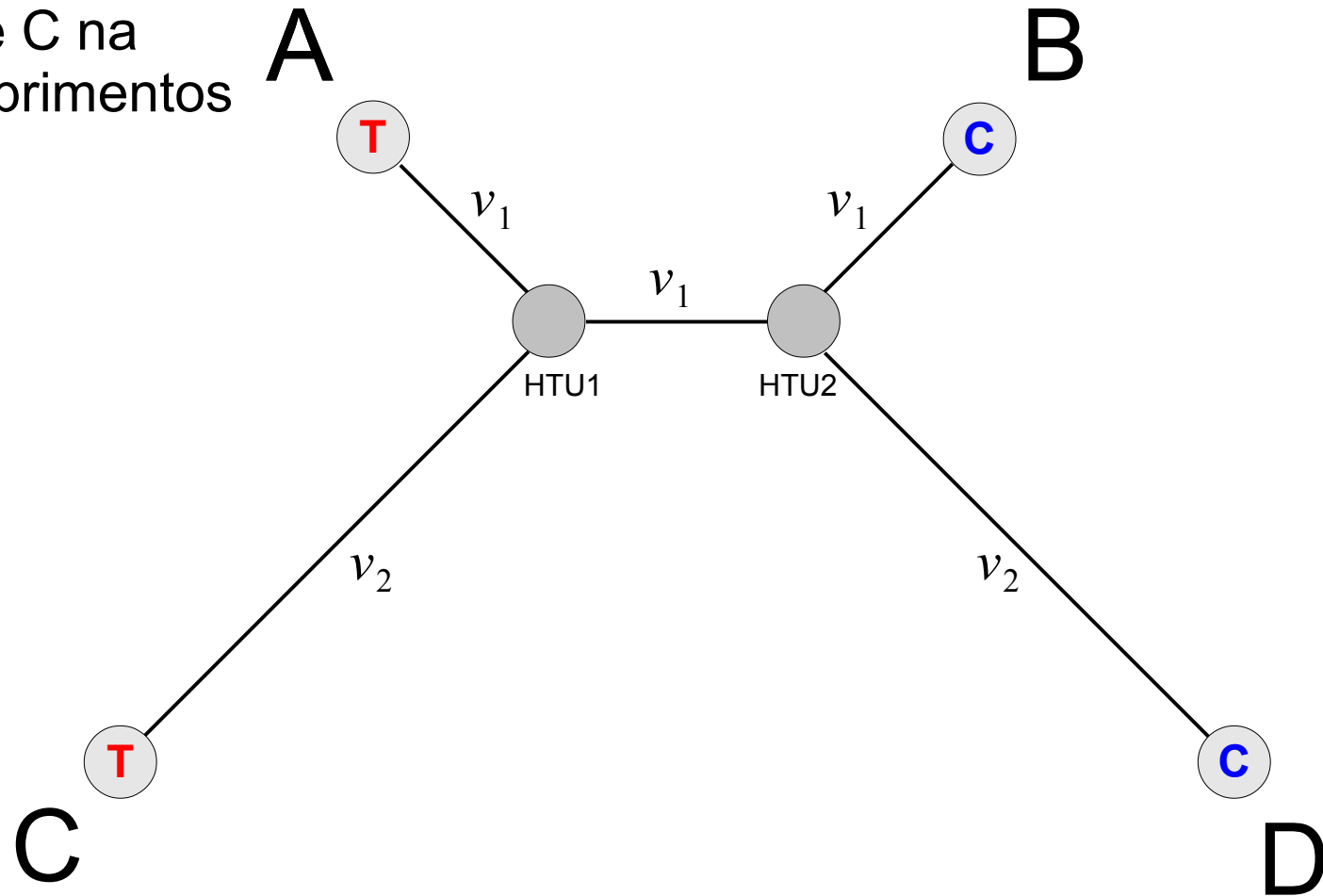
Máxima Verossimilhança e Inferência filogenética

Verossimilhança de topologia:

Considere a distribuição dos estados de caráter T e C para os terminais A, B, C, e C na topologia T_1 com comprimentos de ramos v_1 e v_2 .

Mas afinal...

Qual é a probabilidade da T_1 ?



Máxima Verossimilhança e Inferência filogenética

Verossimilhança de topologia:

Todas essas reconstruções são possíveis, mas somente uma delas ocorreu. Como não sabemos qual, integramos as probabilidades!

$$P_{(R1)} = P_{(TT|v_2)} * P_{(TT|v_1)} * P_{(TC|v_1)} * P_{(CC|v_1)} * P_{(CC|v_2)}$$

$$P_{(R2)} = P_{(TT|v_2)} * P_{(TT|v_1)} * P_{(TG|v_1)} * P_{(GC|v_1)} * P_{(gC|v_2)}$$

$$P_{(R3)} = P_{(TT|v_2)} * P_{(TT|v_1)} * P_{(TT|v_1)} * P_{(TC|v_1)} * P_{(TC|v_2)}$$

$$P_{(R4)} = P_{(TT|v_2)} * P_{(TT|v_1)} * P_{(TA|v_1)} * P_{(AC|v_1)} * P_{(AC|v_2)}$$

$$P_{(R5)} = P_{(TA|v_2)} * P_{(TA|v_1)} * P_{(AA|v_1)} * P_{(AC|v_1)} * P_{(AC|v_2)}$$

$$P_{(R6)} = P_{(TA|v_2)} * P_{(TA|v_1)} * P_{(AC|v_1)} * P_{(CC|v_1)} * P_{(CC|v_2)}$$

$$P_{(R7)} = P_{(TA|v_2)} * P_{(TA|v_1)} * P_{(AG|v_1)} * P_{(GC|v_1)} * P_{(GC|v_2)}$$

$$P_{(R8)} = P_{(TA|v_2)} * P_{(TA|v_1)} * P_{(AT|v_1)} * P_{(TC|v_1)} * P_{(TC|v_2)}$$

$$P_{(R9)} = P_{(TC|v_2)} * P_{(TC|v_1)} * P_{(CA|v_1)} * P_{(AC|v_1)} * P_{(AC|v_2)}$$

$$P_{(R10)} = P_{(TC|v_2)} * P_{(TC|v_1)} * P_{(CA|v_1)} * P_{(AC|v_1)} * P_{(AC|v_2)}$$

$$P_{(R11)} = P_{(TC|v_2)} * P_{(TC|v_1)} * P_{(CG|v_1)} * P_{(GC|v_1)} * P_{(GC|v_2)}$$

$$P_{(R12)} = P_{(TC|v_2)} * P_{(TC|v_1)} * P_{(CT|v_1)} * P_{(TC|v_1)} * P_{(TC|v_2)}$$

$$P_{(R13)} = P_{(TG|v_2)} * P_{(TG|v_1)} * P_{(GA|v_1)} * P_{(AC|v_1)} * P_{(AC|v_2)}$$

$$P_{(R14)} = P_{(TG|v_2)} * P_{(TG|v_1)} * P_{(GC|v_1)} * P_{(CC|v_1)} * P_{(CC|v_2)}$$

$$P_{(R15)} = P_{(TG|v_2)} * P_{(TG|v_1)} * P_{(GG|v_1)} * P_{(GC|v_1)} * P_{(GC|v_2)}$$

$$P_{(R15)} = P_{(TG|v_2)} * P_{(TG|v_1)} * P_{(GT|v_1)} * P_{(TC|v_1)} * P_{(TC|v_2)}$$

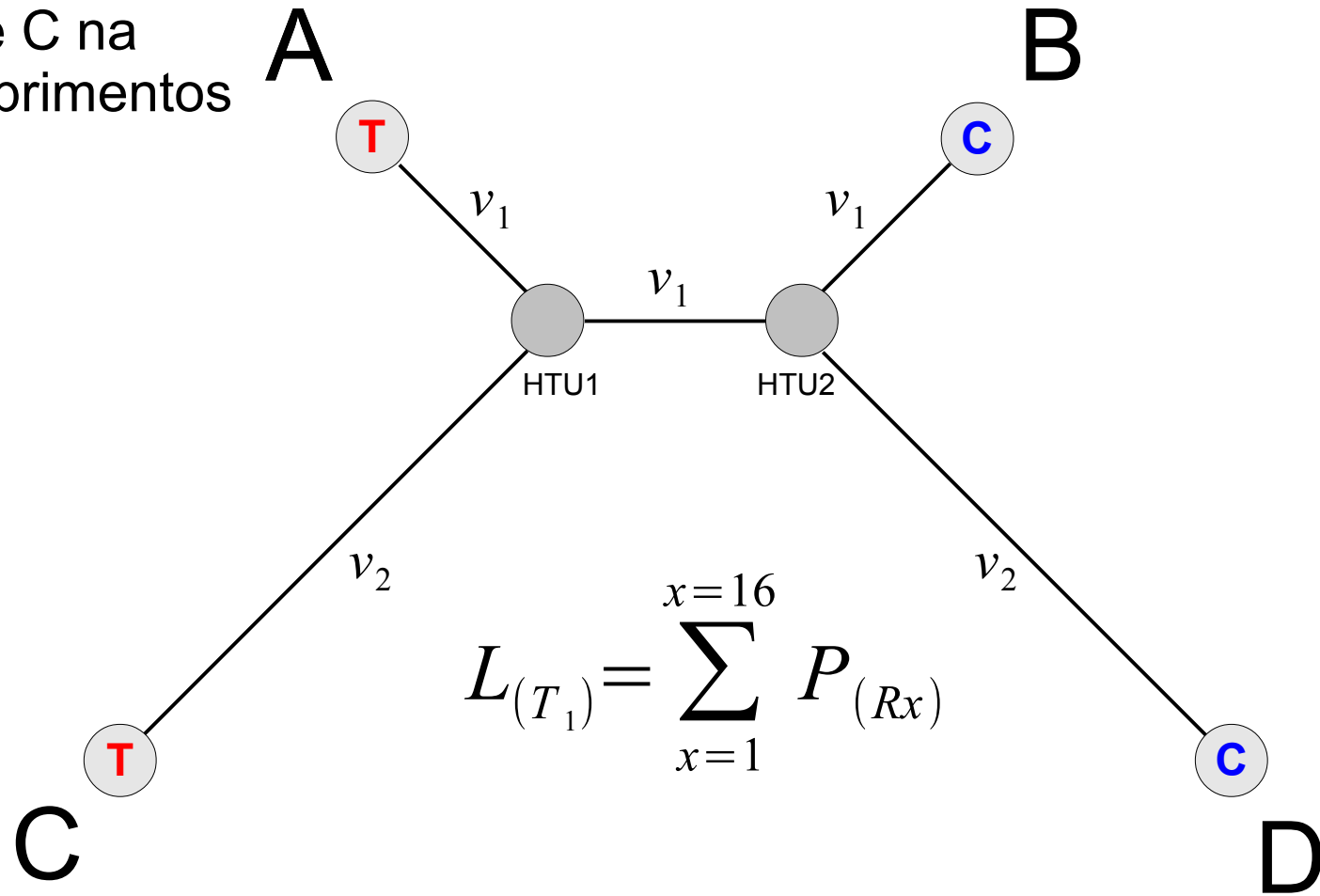
Máxima Verossimilhança e Inferência filogenética

Verossimilhança de topologia:

Considere a distribuição dos estados de caráter T e C para os terminais A, B, C, e C na topologia T_1 com comprimentos de ramos v_1 e v_2 .

Mas afinal...

Qual é a Verossimilhança da T_1 ?



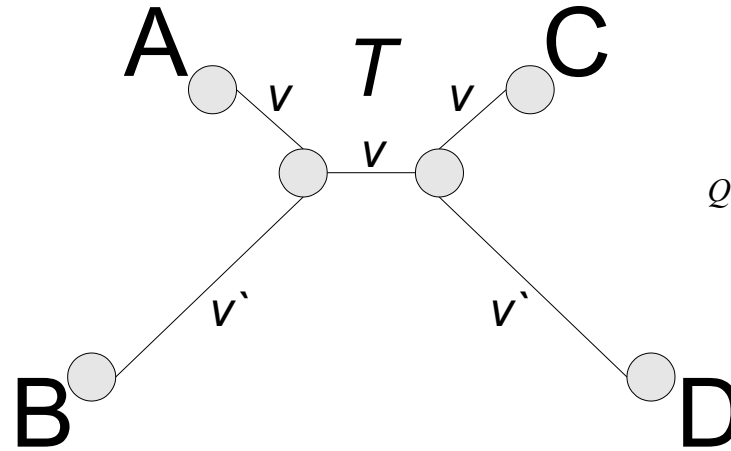
$$L_{(T_1)} = \sum_{x=1}^{x=16} P_{(Rx)}$$

Máxima Verossimilhança e Inferência filogenética

Verossimilhança de topologia:

Considere:

	Taxon	C1	C2	C3
D = {	A	A	C	C
	B	G	T	G
	C	A	C	G
	D	G	T	C



$$Q = \begin{pmatrix} P_{(AA|v)} & P_{(CA|v)} & P_{(GA|v)} & P_{(TA|v)} \\ P_{(AC|v)} & P_{(CC|v)} & P_{(GC|v)} & P_{(TC|v)} \\ P_{(AG|v)} & P_{(CG|v)} & P_{(GG|v)} & P_{(TG|v)} \\ P_{(AT|v)} & P_{(CT|v)} & P_{(GT|v)} & P_{(TT|v)} \end{pmatrix}$$

Se a matriz de dados possui 3 caracteres, como calcular:

$$L(T, v, Q | D)$$

Máxima Verossimilhança e Inferência filogenética

Verossimilhança de topologia:

Considere a distribuição dos estados de caráter T e C para os terminais A, B, C, e C na topologia T_1 com comprimentos de ramos v_1 e v_2 .

Mas afinal...

Qual é a probabilidade da T_1 ?

$$L_{(T_1)} = \sum_{x=1}^{x=16} P_{(Rx)}$$

$$L_{(T_1)} \propto P_{(\text{dados}|Q, T_1, v)}$$

$$L_{(Q, T_1, v|\text{dados})} \propto P_{(\text{dados}|Q, T_1, v)}$$

Máxima Verossimilhança e Inferência filogenética

Verossimilhança de topologia:

Considere:

Taxon	c1	c2	c3
A	A	C	C
B	G	T	G
C	A	C	G
D	G	T	C

Como calcular $L_{(T, \nu, Q|D)}$:

1. Compute $L_{(T, \nu, Q|C_{(1a3)})}$.
2. Multiplique as $L_{(T, \nu, Q|C_n)}$:

$$L_{(T, \nu, Q|D)} = \prod_{n=1}^{n=3} L_{(T, C_n)}$$

Função objetiva de ML: $\max L_{(T, \nu, Q|D)}$

Máxima Verossimilhança e Inferência filogenética

Verossimilhança de topologia:

Considere:

Taxon	c1	c2	c3
A	A	C	C
B	G	T	G
C	A	C	G
D	G	T	C

Como calcular $L(T, \nu, Q | D)$:

1. Compute $L_{(T, \nu, Q | C_{(1a3)})}$:

$$L_{(T|C1)} = \sum_{x=1}^{x=16} P_{(RxC_1)}$$

$$L_{(T|C2)} = \sum_{x=1}^{x=16} P_{(RxC_2)}$$

$$L_{(T|C3)} = \sum_{x=1}^{x=16} P_{(RxC_3)}$$

Máxima Verossimilhança e Inferência filogenética

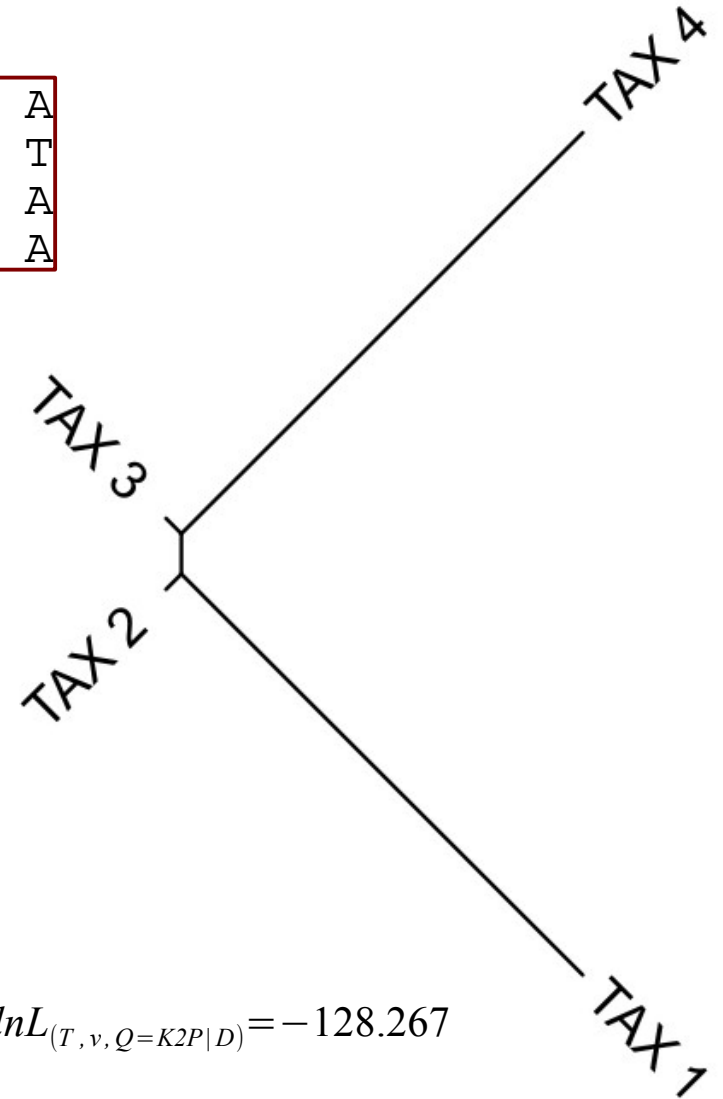
Considere:

TAX_1	A	TA	TTTTTTTTTTTT	AAAAA	AAAAAAAAAAAA	A	A
TAX_2	A	CG	AAAAAAAAAAAA	AAAAA	AAAAAAAAAAAA	A	T
TAX_3	C	CG	AAAAAAAAAAAA	AAAAA	AAAAAAAAAAAA	T	A
TAX_4	C	TA	AAAAAAAAAAAA	AAAAA	TTTTTTTTTTTT	A	A

Não informativos (P)

$((1,4),(2,3))$

$((1,2),(3,4))$



$$\ln L_{(T, v, Q=K2P|D)} = -128.267$$

Máxima Verossimilhança e Inferência filogenética

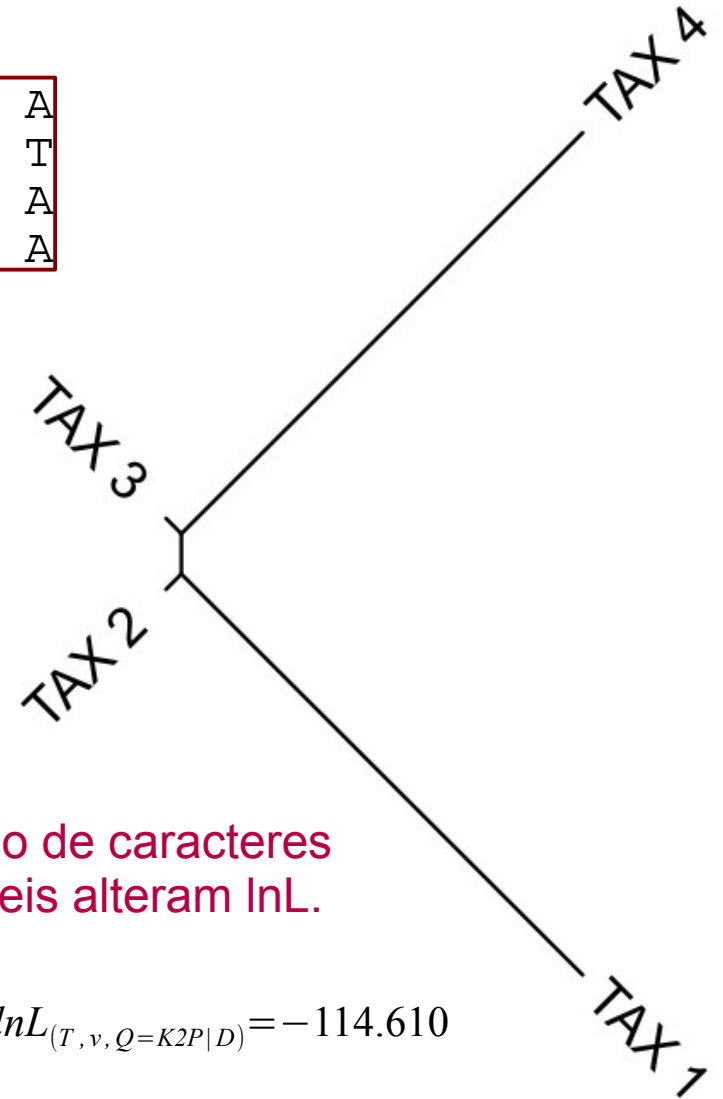
Considere:

TAX_1	A	TA	TTTTTTTTTTTT	AAAAAAAAAAAA	A	A
TAX_2	A	CG	AAAAAAAAAAAA	AAAAAAAAAAAA	A	T
TAX_3	C	CG	AAAAAAAAAAAA	AAAAAAAAAAAA	T	A
TAX_4	C	TA	AAAAAAAAAAAA	TTTTTTTTTTTT	A	A

Não informativos (P)

$((1,4),(2,3))$

$((1,2),(3,4))$



Remoção de caracteres invariáveis alteram $\ln L$.

$$\ln L_{(T, v, Q=K2P|D)} = -114.610$$

Máxima Verossimilhança e Inferência filogenética

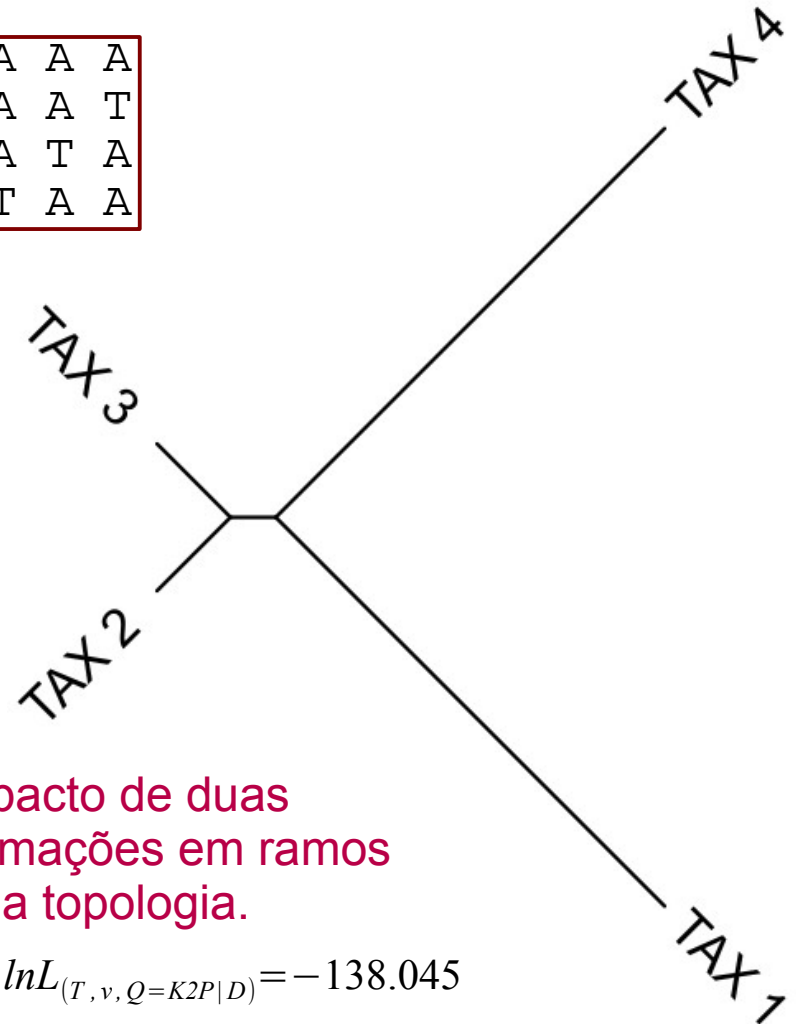
Considere:

TAX_1	A	TA	TTTTTTTTTTTT	AAAA	AAAAAAAAAAAA	A	A
TAX_2	A	CG	AAAAAAAAAAAA	TTAA	AAAAAAAAAAAA	A	T
TAX_3	C	CG	AAAAAAAAAAAA	AATT	AAAAAAAAAAAA	T	A
TAX_4	C	TA	AAAAAAAAAAAA	AAAA	TTTTTTTTTTTT	A	A

Não informativos (P)

$((1,4),(2,3))$

$((1,2),(3,4))$



Máxima Verossimilhança e Inferência filogenética

Considerações filosóficas e epistemológicas:

Qual é o melhor método de inferência?

Qual é a melhor árvore (topologia)?

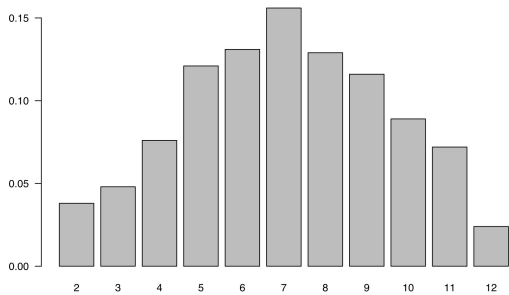
Não há uma única resposta para a pergunta.

Para ambas necessário adotar um critério de avaliação objetivo!

MÉTODOS DE DISTÂNCIA
VS.
TRANSFORMAÇÃO DE CARACTERES

Máxima Verossimilhança e Inferência filogenética

Considerações filosóficas e epistemológicas:



Sistemática, assim como muitas outras disciplinas em Biologia, é uma ciência histórica na qual os eventos que buscamos explicar são únicos no tempo e no espaço.

"If each character observation is a unique object, it cannot be a random sample drawn from a parameterized distribution expressed as a model of change (e.g., GTR+I+G)."

(Wheeler 2012:272)

Máxima Verossimilhança e Inferência filogenética

Considerações filosóficas e epistemológicas:



Probabilidades: duas interpretações

i. Grau de crença:

Confiança relativa em determinado resultado, considere:

"Esta moeda tem 80% de chance de ser honesta."

Máxima Verossimilhança e Inferência filogenética

Considerações filosóficas e epistemológicas:



Probabilidades: duas interpretações

ii. Uma proposição verdadeira sobre o universo natural semelhante à eventos repetidos:

Implica em uma proposição sobre o que é ontologicamente real, considere:

"Esta moeda tem 80% de chance de dar coroa em uma jogada."

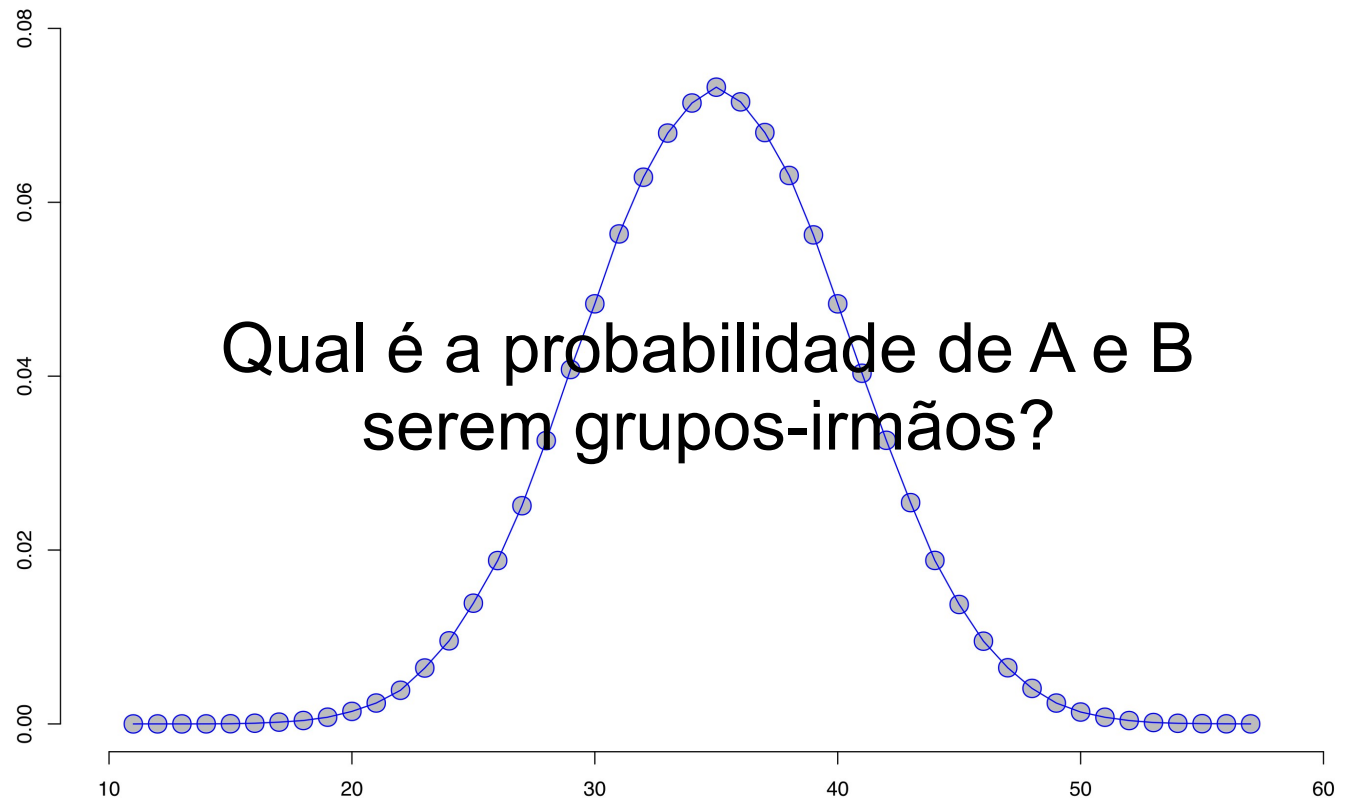


Oferece uma propriedade da moeda e uma proposição relacionada à sua realidade física.

Máxima Verossimilhança e Inferência filogenética

Considerações filosóficas e epistemológicas:

Ciências históricas não permitem facilmente a interpretação frequencista (i.e., ii)



Escolha de critérios de otimização

Critérios de otimização são premissas analíticas utilizadas para interpretar eventos historicamente únicos. Como tal, eles não podem ser testados empiricamente por acurácia. Por outro lado, congruência pode ser um caminho a ser seguido. No entanto, em seu cerne, critérios de otimização são pressupostos que requer suporte e justificativa quando empregados. Pressupostos diferentes levam a resultados diferentes. Somente justificando nossas premissas nós podemos defender nossas conclusões.

(Wheeler 2012:272)