

Tutorial 14

Suporte

BIZ0433 - INFERÊNCIA FILOGENÉTICA: FILOSOFIA, MÉTODO E APLICAÇÕES.

POR DENIS JACOB MACHADO & FERNANDO P. L. MARQUES

Conteúdo

Objetivo	236
14.1 Introdução	237
14.2 Bootstrap	238
14.2.1 Considerações gerais	238
14.2.2 Bootstrap sob o critério de Parcimônia	239
14.2.3 Bootstrap sob o critério de Verossimilhança Máxima	242
14.3 Suporte de Goodman-Bremer	243
14.4 Verossimilhança diferencial (<i>Likelihood difference</i>)	246
14.5 Referências	249

Objetivo

O objetivo deste tutorial é apresentar conceitos teóricos e práticos do cálculo de índices de suporte em inferência filogenética. A parte conceitual é brevemente apresentada aqui e o leitor deve consultar Grant & Kluge [1], e referências ali citadas, para uma discussão mais aprofundada sobre os aspectos epistemológicos associados à cada métrica. O tutorial inicia-se com uma apresentação geral sobre o conceito de suporte. A seguir são apresentados 3 métricas, duas das quais são comumente utilizados em análises filogenéticas. O primeiro deles é o Bootstrap, o segundo os valores de Goodman–Bremer e, finalmente, a Verossimilhança diferencial. Os arquivos associados a este tutorial estão disponíveis no [GitHub](#). Você baixar todos os tutoriais com o seguinte comando:

```
svn checkout https://github.com/fplmarques/cladistica/trunk/tutorials/
```

“It is sometimes observed that the measure of support does not have any ‘meaning’, by which is usually meant any ‘probability interpretation’. It is indeed true that a statement of support, though derived from probabilities, does not make any assertion about the probability of a hypothesis being correct. And for a good reason: the method of support has been developed by people who explicitly deny that any such statement is generally meaningful in the context of a statistical hypothesis.” [2:33]

14.1 Introdução

Como sugere Grant & Kluge [1], a avaliação de suporte de clados é um componente importante de análises filogenéticas, mesmo que pouca atenção tenha sido dada às bases conceituais dos métodos geralmente empregados para medi-lo. Adicionalmente, o próprio conceito de “suporte” difere entre cientistas, em geral, e filogeneticistas em particular.

Em estudos filogenéticos, há uma série de métricas utilizadas para inferir índices de suporte de clados. Dentro destes índices, há aqueles considerados métricas de suporte indireto, dentro dos quais encontramos Bootstrap e Jackknife – entre outros, e métricas de suporte direto, tais como Goodman-Bremer e Verossimilhança diferencial – entre outros. Este tutorial irá apenas abordar a implementação de algumas destas métricas. Para uma revisão mais adequada sobre o assunto é recomendável a leitura de Siddall [3], Egan [4], Grant & Kluge [1], Wheeler [5] e Jacob Machado *et al.* [6]; e referências citadas por esses autores.

Existem definições explícitas sobre suporte na literatura. No entanto, filogeneticistas parecem ignorá-las. Em seu livro clássico sobre Verossimilhança, Edwards [2] considera que índices de suporte devem aferir o mérito relativo de hipóteses rivais diante dos dados observados. Para este autor, a noção de suporte pode ser concebida por uma interpretação operacional perfeitamente simples da diferença entre as verossimilhanças entre duas hipóteses com relação aos dados. Neste contexto, essa diferença expressa, à longo prazo, a frequência na qual as hipóteses rivais explicam os dados observados.

Grant & Kluge [1] oferecem uma proposta de definição objetiva para suporte em inferência filogenética. Para esses autores, métricas de suporte só são adequadas quando diretamente relacionadas com o critério de otimalidade. Para estes autores, conhecimento adquirido (b), probabilidade (p), hipótese (h) e evidência (e) são definidos como:

1. O conhecimento adquirido (b) é toda soma de conhecimento que não esteja sujeito ao teste de hipóteses ou falseamento.
2. A probabilidade (p) diz respeito à “frequência relativa” de eventos.
3. A hipótese (h) é um cenário causal ou causativo que está sujeito a teste e falseamento.
4. Evidência (e) são os dados obtidos para testar a hipótese.

Segundo Popper [7], o suporte (S) é a medida da diferença entre (*i*) probabilidade da evidência dada a hipótese e o conhecimento adquirido e (*ii*) a probabilidade da evidência dado o conhecimento adquirido:

$$S = p(e|h,b) - p(e|b) \quad (14.1)$$

Desta forma, suporte pode ser visto como uma afirmação empírica sobre a força de hipóteses $(h_1, h_2, h_3 \dots h_n)$ em relação ao corpo de evidências e dado um certo conhecimento adquirido b . Sendo que otimalidade (O) é também uma afirmação empírica sobre a força de hipóteses concorrentes em relação ao corpo de evidências dado um certo conhecimento adquirido. Seguindo essa lógica, Grant & Kluge [1] concluem que ambos, S e O , devem ser diretamente proporcionais:

$$S(h|e,b) \propto O(h|e,b) \quad (14.2)$$

Observe que otimalidade e suporte não são sinônimos (veja Tabela 14.1). Desta forma, S é diretamente proporcional, mas não igual, a O . Deste modo, para que uma medida de suporte seja objetiva, ela deve quantificar suporte como uma função do poder explanatório. Segundo esta definição de suporte, o índice Goodman-Bremer é uma medida direta do suporte, mas não os valores de Bootstrap e Jackknife.

Tabela 14.1: Suporte vs. Otimalidade

Suporte	Otimalidade
Refere-se à força <i>relativa</i> entre diferentes hipóteses	Refere-se à força <i>absoluta</i> de uma hipótese
É heurístico – facilita a descoberta	É científico
Geralmente está relacionado à clados	Diz respeito a topologias

14.2 Bootstrap

14.2.1 CONSIDERAÇÕES GERAIS

A proporção de Bootstrap é a métrica mais utilizada em estudos filogenéticos, apesar de ter sido criticado severamente quanto às violações de premissas do método [veja 3, 5; e referências citadas], inconsistência epistemológica [5], e a não-correlação desta métrica com medidas de suporte usadas na teoria da verossimilhança máxima [6]. Este método estatístico foi concebido para melhorar estimativas de parâmetros de distribuições desconhecidas. O método se vale de obter n pseudo-réplicas da amostra sobre as quais é calculado o parâmetro de interesse.

Considere o seguinte script em R:

```
1 s <- c(320.44, 303.98, 264.10, 286.90, 274.59, 332.72, 346.37, 240.83)
2 replicates <- c()      # vetor vazio que receber as médias de cada réplica
3 for (r in 1:100) {     # Loop para 100 aleatorizações, 100 pseudo-réplicas (com reposição)
4   replicate <- mean(sample(s, 8, replace = TRUE))
5   replicates <- c(replicates, replicate)
6 }
7 mean_of_replicates <- mean(replicates) # média das réplicas
8 print(mean_of_replicates) # imprime a média das médias das réplicas
```

Neste exemplo, a variável s (linha 1) recebe um vetor (lista) com 8 medidas morfométricas para qual a média estimada é 296.2412 ($\text{mean}(s)$). O método de bootstrap, se vale do uso de pseudo-réplicas ($r_1, r_2, r_3 \dots r_{100}$, loop nas linhas 3 a 6). Para cada réplica, são sorteados 8 valores de s – com repetição, ou seja uma mesma medida pode ser considerada de 0 a 8 vezes –, a média desses valores é calculada (linha 4) e a média é inserida em um vetor (replicates , linha 5). Desta forma, uma réplica pode receber os valores 303.98, 274.59, 286.90, 274.59, 240.83, 274.59, 303.98 e 346.37, cuja média seria 288.2287. Após 100 pseudo-réplicas, o script calcula a média das réplicas, ou seja a média das médias (linha 7) e finalmente imprime o resultado: 295.57¹. Em teoria, esse valor é uma estimativa mais aproximada da média real da população do qual amostra s foi obtida. É importante considerar, que o método assume (i.) que minha amostra foi obtida aleatoriamente, (ii.) que eles são independentes e (iii.) que eles são identicamente distribuídos – ou seja, foram obtidos da mesma distribuição. É interessante observar que nenhuma dessas premissas pode ser assumida para dados filogenéticos.

14.2.2 BOOTSTRAP SOB O CRITÉRIO DE PARCIMÔNIA

Em sistemática filogenética, o bootstrap é aplicado da mesma maneira. As pseudo-réplicas são matrizes com o mesmo número de caracteres que a matriz original com reposição, ou seja, em uma pseudo-réplica, um ou mais caracteres podem estar representados mais que uma vez – consequentemente um ou mais caracteres podem estar presente. Para cada pseudo-réplica, faz-se uma busca e a(s) topologia(s) encontrada(s) é(são) retida(s) até que se complete o ciclo de pseudo-réplicas. Ao final, computa-se o consenso de maioria e a frequência de cada clado representa a proporção do clado em bootstrap. Desta forma, o método obedece o seguinte cronograma de execuções:

1. Determinar uma árvore ótima;
2. Re-amostrar os dados (colunas de caracteres) com reposição, gerando novas matrizes com o mesmo número de colunas que a matriz original;
3. Determinar uma árvore ótima para cada nova matriz criada usando o mesmo procedimento que o empregado em (1);
4. Repetir (2) e (3) k vezes, salvando todas as árvores assim encontradas;
5. Construir uma de consenso de maioria com todos os nós que tenham frequência ≥ 0.50

É interessante notar que para um conjunto de n , caracteres sendo que nenhum deles é homoplástico, o valor de bootstrap (V_{Bt}) para o vértice V cujo suporte reside em r caracteres é dado pela seguinte função:

$$V_{Bt} = 1 - \left(1 - \frac{r}{n}\right)^n \quad (14.3)$$

¹esse valor varia a cada execução do script.

Exercício 14.1

*=a-f) no formato xread. Adicionalmente, há um arquivo chamado `vbt.r` que pode ser editado para calcular o valor de V_{Bt} da equação acima (`$ Rscript vbt.r`). Verifique o conteúdo destes arquivos e note as diferenças que existem entre as 6 matrizes de dados. Observe também que o arquivo contém as instruções para o cálculo dos valores de bootstrap em TNT. Você deverá executar esses arquivos em TNT (e.g., `$ tnt proc bs_example_1a.tnt`) e responder às seguintes perguntas:

1. Quais são as principais diferenças que você observa nas matrizes apresentadas?

2. Os índices de Bootstrap que você obteve correspondem às expectativas teóricas da Equação 14.3?

3. Você acha que os números de caracteres necessários para prover suporte aos vértices é dependente no números de terminais?

4. Considere as duas matrizes que diferem quanto aos números de terminais mas possuem o menor números de caracteres sustentando cada clado. Você consideraria que seus dados não possuem suporte para seus resultados?

Quando proposto inicialmente por Felsenstein [8], os valores de deveriam fornecer intervalos de confiança para os vértices. De acordo com Felsenstein [8], vértices eram considerados como estatisticamente significantes quando os valores de Bootstrap eram ≥ 0.95 . Esta ideia foi subsequentemente abandonada, no entanto há pesquisadores que ainda consideram esse valor como uma medida de suporte para clados. É interessante notar que a Equação 14.3 permite a seguinte derivação:

$$r = n - n * \left(e^{\frac{\ln(1-V_{Bt})}{n}} \right) \quad (14.4)$$

Considere que o número mínimo de caracteres para resolver um diagrama binário é $t - 2$, onde t é o número de terminais. Desta forma, teríamos:

$$r = (t - 2) - (t - 2) * \left(e^{\frac{\ln(1-V_{Bt})}{t-2}} \right) \quad (14.5)$$

Essa função estima o valor de r , ou seja, número de caracteres que sustentam um vértice, para uma matriz com t terminais e determinado valor de V_{Bt} – lembrando que assume-se que não há homoplasias nos dados.

Exercício 14.2

No diretório deste tutorial há um *script* chamado `minimum_character.r` que produz um gráfico para os valores de r e t de acordo com o valor de V_{Bt} . Sua execução é feita com a seguinte linha de comando: `$ Rscript minimum_character.r`. Neste exercício você deverá fazer três execuções desse *script*. Após cada uma delas você deverá renomear o arquivo de saída `Rplots.pdf`, caso contrário a próxima execução apagará o resultado da anterior. Inicialmente, você deverá executar o *script*. Nas demais execuções, você deverá alterar os valores da variável `VBt`, linha 3, para `.99` e `.999999`. Após as execuções, avalie os gráficos resultantes e responda:

- i. Os resultados de sua análise modificam sua impressão relacionada ao exercício anterior sobre a relação entre a dependência entre número de terminais e o número de caracteres necessário para dar suporte ao nó de acordo com os valores de bootstrap? Justifique.

- ii. De acordo com suas observações e os argumentos levantados por Grant & Kluge [1], você considera essa métrica uma boa medida de suporte?

14.2.3 BOOTSTRAP SOB O CRITÉRIO DE VEROSSIMILHANÇA MÁXIMA

O cálculo de Bootstrap sob este critério de otimalidade obedece a mesma lógica de como ele é calculado sob o critério de parcimônia. Inúmeras buscas sob este critério são executadas para as pseudo-réplicas cujas topologias obtidas são utilizadas ao final para calcular um consenso de maioria. A frequência de cada clado é denominada proporção de Bootstrap.

Como as análises de Verossimilhança foram efetuadas em **GARLI** [9], iremos utilizar o mesmo programa para calcular as proporções de Bootstrap. Considere o conteúdo do arquivo `garli_bootstrap.conf` (disponível no diretório deste tutorial). Este arquivo é uma modificação do arquivo de configuração do GARLI `garli_single.conf` utilizado no Tutorial 13. O comando `datafname` foi modificado para receber o arquivo `partition1+2aln3.nex`; o comando `ofprefix` foi modificado para indicar a análise que será feita; o comando `searchreps` foi configurado para uma única busca; e finalmente, o comando `bootstrapreps` foi configurado para 100 réplicas. Adicionalmente, o modelo especificado ($TrN + I + \Gamma$) foi selecionado anteriormente pelo **jModelTest 2** (Tutorial 12).

A execução de GARLI utilizando o arquivo `garli_bootstrap.conf` gera três arquivos. Os arquivos `garli_bs.log00.log` e `garli_bs.screen.log` são os registros da execução. O arquivo `garli_bs.boot.tre` contém as topologias recuperadas para cada uma das pseudo-réplicas (100) e é este arquivo que é utilizado para calcular as proporções de Bootstrap.

A forma mais fácil de calcular estas proporções é utilizando **SumTrees** (*Phylogenetic Tree Summarization and Annotation*) que é uma ferramenta de **DendroPy**. No exemplo acima, após computada as topologias de Bootstrap, bastaria executar o seguinte comando:

```
sumtrees.py -d 0 -p -o garli_run.best_bootstrap.tre -t garli_run.best.tre garli_bs.boot.tre
```

Há várias formas de computar os valores de Bootstrap usando `sumtrees.py` e é recomendável verificar as opções disponíveis na documentação do programa². Na linha de comando acima, a opção `-d 0` define o número de decimais, `-p` define que os resultados serão expressos em porcentagem, `-o` é a opção que define o arquivo de saída (*output*) – neste caso `garli_run.best_bootstrap.tre`, `-t` é a opção que define a topologia alvo, ou seja, aquela cujos clados você deseja verificar as frequência nas topologias recuperadas durante a análise de bootstrap – neste caso `garli_run.best.tre`³, e finalmente, o arquivo no qual as topologias de Bootstrap estão, `garli_bs.boot.tre`.

A etapa final deste processo é observar os valores de Bootstrap no programa FigTree⁴. Este programa já foi utilizado no Tutorial 2 e é utilizado para visualização e edição de topologias. Para verificar os valores de Bootstrap, basta seguir as seguintes etapas:

1. Abra o arquivo `garli_run.best_bootstrap.tre` em FigTree.

²<https://pythonhosted.org/DendroPy/programs/sumtrees.html>

³Esta topologia foi recuperada durante o Tutorial 12

⁴ O Tutorial 2 possui um breve vídeo de como o programa funciona.

2. O programa exibirá uma janela solicitando que você defina à que se referem os valores das topologias. Digite “bootstrap” e pressione OK.
3. Enraize a topologia no táxon “Taxon1”.
4. No canto inferior esquerdo na janela de Figtree, selecione “Branch Lables”.
5. Abra as opções de “Branch Lables” – pressionando o triângulo à esquerda desta opção – e em “Display”, “selecione bootstrap”.

Exercício 14.3

Neste exercício, você deverá visualizar o arquivo `garli_bs.boot.tre` em FigTree e anotar os valores de Bootstrap na Figura 14.1.

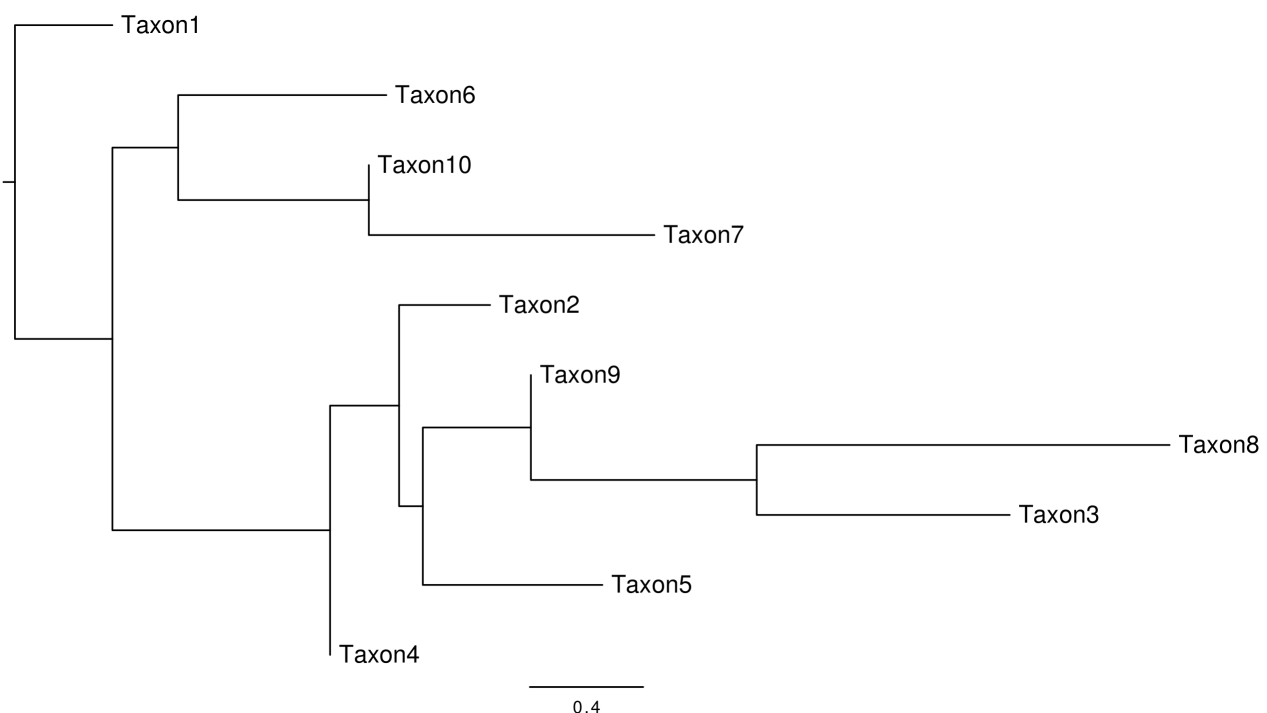


Figura 14.1: Topologia selecionada sob o critério de Verossimilhança máxima dos dados em `partition1+2aln3.nex` com lnL -1369.5791.

14.3 Suporte de Goodman-Bremer

Ao contrário dos valores de Bootstrap, o suporte de Goodman-Bremer (GB) [10, 11] é uma medida empírica de suporte dado pela evidência a um determinado clado. Portanto, é considerado um valor de suporte direto. O suporte de Goodman-Bremer é a diferença de comprimento (*i.e.*, custo, número de passos) entre o cladograma mais parcimonioso (L) e o cladograma mais parcimonioso que não apresenta um determinado clado de interesse (L'):

$$GB = L' - L \quad (14.6)$$

Em outras palavras, o valor de Goodman-Bremer representa o número adicional de passos em uma topologia sub-ótima necessário para colapsar determinado nó. Uma outra maneira de pensar sobre essa métrica é que o seu valor indica a quantidade relativa de contra-evidências necessária para que você não tenha suporte nenhum sobre determinado grupo. Vale considerar que o GB é dependente do enraizamento da topologia e que grupos que estariam ausentes no consenso estrito, e portanto não são suportados pelas evidências, apresentam $GB = 0$.

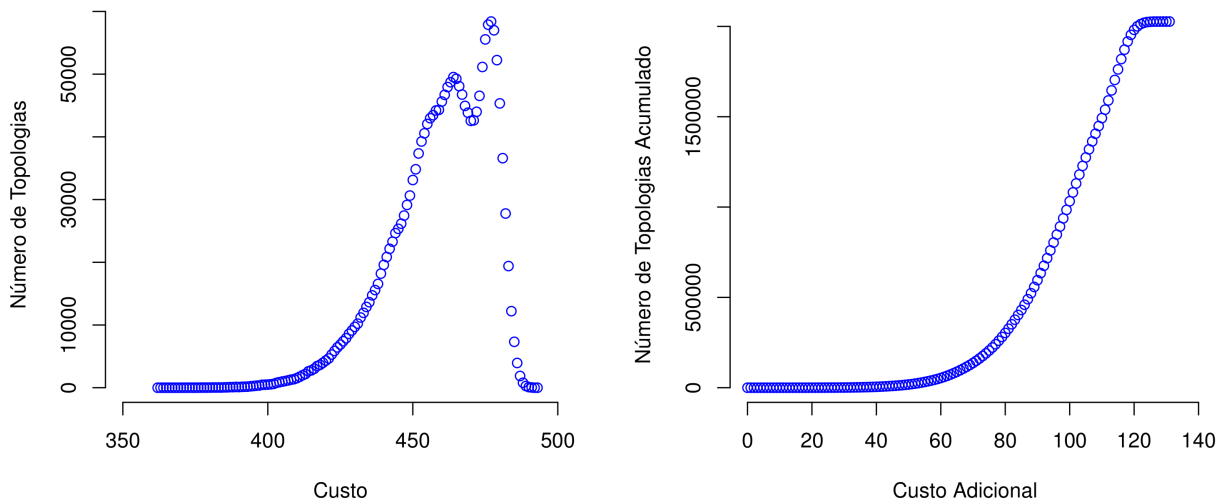


Figura 14.2: Distribuição de topologias de acordo com seu custo (esquerda) e acúmulo de topologias de acordo com número de passos adicionais com relação à topologia mais curta para os dados do arquivo `total_evidence.tnt` (direita).

Há algumas propriedades sobre esse índice que você deve considerar. Está claro que para computá-lo é necessário avaliar topologias sub-ótimas e isso pode ser um problema se você está diante de um espaço topológico complexo. Considere por exemplo o arquivo de dados `total_evidence.tnt`. Neste arquivo estão concatenados os dados de `partition1.fas`, `partition2.fas` e `partition3.tnt`. A enumeração destes dados, que contém 10 terminais – portanto 2.027.025 topologias possíveis, resulta em custos que variam de 362 a 493 (Figura 14.2). Suponha que para determinado nó, o índice de GB seja 120, você teria que ter avaliado praticamente todo o espaço de topologias. Felizmente, para esses dados, o clado com maior índice de GB é 14. Desta forma, para esses dados o cômputo desses índices parece ser trivial – o que certamente não é o caso de muitos dados reais.

Há várias estratégias adotadas para calcular o índice de BG. Em sua forma mais primitiva, você poderia compilar todas as topologias com n passos de distância da topologia mais curta que você obteve para seus dados e posteriormente, calcular passo a passo esse índice filtrando as topologias $c + n$, no qual c é o custo mínimo e n é o número de passos adicionais, calculando o consenso, e verificando quais nós são colapsados. Fazer isso manualmente é laborioso e demanda tempo computacional e analítico.

A melhor forma de calcular o índice de GB é utilizando um macro disponível em TNT chamado

BREMER.RUN. Este macro utiliza buscas com os algoritmos de novas tecnologias para compilar de topologias sub-ótimas, que posteriormente são utilizadas para obter os índices. Para executar esse macro é necessário que você possua uma topologia na memória. Caso seus dados resultem em múltiplas topologias, o ideal seria você calcular o consenso estrito desta topologia, mantê-lo na memória e descartar as topologias fundamentais (veja Seção 6.2 do Tutorial 6) e, finalmente, executar o macro do TNT.

O POY também permite o cálculo deste índice utilizando algumas de suas estratégias internas. Porém isso é feito assumindo homologia estática dos dados, uma vez que o uso de homologia dinâmica para esse propósito se mostrou computacionalmente inviável – além de desnecessário. Sendo assim, considero mais efetivo submeter o alinhamento implícito das análises em POY para o cálculo de BG. Considere o exemplo abaixo:

```
read (" partition1 . fas ", " partition2 . fas ", " partition3 . tnt ")
set ( root : " Taxon1 " )
search ( max_time : 0 : 0 : 5 )
select ( )
set ( iterative : exact )
swap ( tbr )
fuse ( )
select ( best : 1 )
transform ( all , ( static_approx ) )
report ( " total_evidence . tnt " , phastwinclad )
exit ( )
```

Neste *script* (denominado `poy_run.poy` e disponível no diretório deste tutorial), os dados são lidos e analisados utilizando os algoritmos de busca do comando `search` por 5 minutos, a melhor topologia é selecionada e rediagnosticada via *Iterative Pass* com refinamento, uma única topologia mais curta é selecionada, todos os dados são transformados em homologia estática e impresso no arquivo `total_evidence.tnt` em formato TNT. Essa lógica analítica é uma boa estratégia para dados sujeitos a homologia dinâmica para os quais você deseja calcular índices que dependam de homologia estática – e consequentemente valores de Bootstrap (caso você ainda considere esse último útil para alguma coisa). Para obter os valores de BG para esse arquivo, então, basta: (i.) analisá-lo em TNT e obter uma topologia e (ii.) executar o macro com o comando `run BREMER.RUN`.

Exercício 14.4

Neste exercício você deverá obter os índices de bootstrap e GB para o arquivo `total_evidence.tnt`. Use o espaço abaixo para desenhar a topologia e anotar os índices obtidos e, posteriormente, responda as seguintes perguntas:

i. Qual é o maior índice de GB que um determinado vértice (nó) pode possuir?

ii. Existe alguma correlação entre os índices de bootstrap e GB?

14.4 Verossimilhança diferencial (*Likelihood difference*)

A Verossimilhança diferencial é equivalente ao índice de GB dentro do contexto de análises sob o critério de Verossimilhança. Embora pouco empregado [veja referências em 1, 6], esse índice possui todas as propriedades desejáveis de um índice de suporte [1, 2]. O índice é dado pela seguinte formulação:

$$P(e|h,b) - P(e|h',b) \quad (14.7)$$

no qual h é a hipótese selecionada e h' é a hipótese alternativa. Como GB, este índice está diretamente relacionado ao critério de otimalidade, e portando, deve ser considerado um índice de suporte direto. Dentro do contexto de análises filogenéticas, a diferença de Verossimilhança é aplicada em relação à topologia ótima e aquela considerada ótima na ausência do clado em questão. Atualmente não há programas que possibilitam o cômputo imediato deste índice. Sua obtenção, portanto, dependerá de uma série de análises sequenciais que iremos demonstrar a seguir.

Suponha que seus dados estejam em `partition1+2a1n3.nex` e que após selecionar o melhor modelo para sua análise, você configurou o arquivo `garli_single.conf`, ambos disponíveis

no diretório deste tutorial. O resultado desta análise (arquivos `garli_run.*.*`) resultou em uma topologia com $\ln L$ -1369.5791. Figura 14.1 do Exercício 14.3 é resultado desta análise.

Para calcular a diferença de Verossimilhança para o clado (Taxon8+Taxon3) é necessário obter a topologia ótima na qual esse clado não existe. Isso é feito utilizando o que chamamos de buscas sob restrição reversa (*reverse/negative constraint*). A primeira etapa seria definir a topologia que contém esse clado. Para esse caso em particular, seria:

```
(Taxon1, Taxon2, Taxon4, Taxon5, Taxon6, Taxon7, Taxon9, Taxon10, (Taxon8, Taxon3))
```

Esta topologia deve estar contida em um arquivo texto com um sinal – (negativo), que indica ao GARLI que ela deve ser utilizada como restrição reversa. Chamaremos esse arquivo de `constraint_8+3.tre`. O próximo passo é editar o arquivo de configuração de GARLI. Para isso, uma cópia do arquivo `garli_single.conf` foi feita e denominada `garli_constrain.conf` e os comandos `ofprefix` e `constraintfile` foram modificados para o cálculo (veja conteúdo desses arquivos).

O resultado desta análise gerou uma topologia com $\ln L$ -1369.8372, a qual é considerada a topologia ótima sob este critério na qual o clado (Taxon8+Taxon3) não existe. Para calcular a Verossimilhança diferencial para este clado, basta subtrair o $\ln L$ da hipótese sob restrição reversa daquela obtida na primeira análise onde este clado foi recuperado. Assim

$$LR_{T8+T3} = \ln L_{comT8+T3} - \ln L_{semT8+T3} \quad (14.8)$$

Portanto,

$$LR_{T8+T3} = 0.2581. \quad (14.9)$$

Exercício 14.5

Neste exercício você deverá obter os índices das Razões de Verossimilhança para os cinco clados definidos abaixo, provenientes da Figura 14.1, utilizando as etapas descritas acima. Anote os valores obtidos na Figura 14.1 e responda:

Clados de interesse:

```
(Taxon1, Taxon2, Taxon4, Taxon5, Taxon6, Taxon7, Taxon10, (Taxon9, Taxon8, Taxon3))  
(Taxon1, Taxon2, Taxon4, Taxon6, Taxon7, Taxon10, (Taxon5, Taxon9, Taxon8, Taxon3))  
(Taxon1, Taxon4, Taxon6, Taxon7, Taxon10, (Taxon2, Taxon5, Taxon9, Taxon8, Taxon3))  
(Taxon1, Taxon6, Taxon7, Taxon10, (Taxon4, Taxon2, Taxon5, Taxon9, Taxon8, Taxon3))
```

- i.* Existe alguma correlação observável entre os índices de Bootstrap obtidos no Exercício 14.3 e Verossimilhança diferencial?

14.5 Referências

1. Grant, T & Kluge, A. 2008. Clade support measures and their adequacy. *Cladistics* **24**(6): 1051–1064.
2. Edwards, A. W. F. 1992. Likelihood. Baltimore, Maryland: The John Hopkins University Press, 1992. 275 pp.
3. Siddall, M. E. em *Methods and Tools in Biosciences and Medicine* eds. DeSalle, R.; Wheeler, W. C. & Giribet, G., 80–101. Basel: Birkhäuser Verlag, 2001.
4. Egan, M. G. 2006. Support versus Corroboration. *Journal of Biomedical Informatics* **39**: 72–85.
5. Wheeler, W. C. 2012. Systematics: a course of lectures. Malaysia: Wiley-Blackwell, 2012. 426.
6. Jacob Machado, D.; Portella de Luna Marques, F.; Jiménez-Ferbans, L. & Grant, T. 2022. An empirical test of the relationship between the bootstrap and likelihood ratio support in maximum likelihood phylogenetic analysis. *Cladistics* **38**(3): 392–401.
7. Popper, K. 1959. The Logic of Scientific Discovery. London: Routledge, 1959, 513. ISBN: 041507892X.
8. Felsenstein, J. 1985. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution* **39**(4): 783–791.
9. Zwickl, D. J. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Tese de doutorado (The University of Texas, Austin, 2006).
10. Goodman, M; Olson, C.; Beeher, J. & Czelusniak, J. 1982. New perspectives in the molecular biological analysis of mammalian phylogeny. *Acta Zoologica Fennica* **169**: 19–35.
11. Bremer, K. 1988. The Limits of Amino Acid Sequence Data in Angiosperm Phylogenetic Reconstruction. *Evolution* **42**(4): 795–803.