Teste de modelos (AIC)

Caetano - 2021

Definindo – Free parameters

- Parâmetros livres são aqueles estimados usando uma busca heurística ou a maximização direta da verossimilhança do modelo.
- Modelos com mais parâmetros livres são considerados mais complexos.
- A complexidade do modelo, via de regra, aumenta o valor de verossimilhança.

Tutorial

Definindo – Nested models

- Modelos aninhados são aqueles que podem ser descritos como um caso especial de um modelo mais geral.
- Os modelos de diversificação que criamos são todos aninhados, pois são casos especiais do modelo livre.

Definindo – Likelihood ratio test

- No caso de modelos aninhados, podemos testar qual é o melhor modelo usando a razão das verossimilhanças.
- A razão das verossimilhanças testa se temos suporte para aceitar o modelo mais complexo ao invés do modelo mais simples (nulo).

Definindo – Likelihood ratio test

- No caso de modelos aninhados, podemos testar qual é o melhor modelo usando a razão das verossimilhanças.
- A razão das verossimilhanças testa se temos suporte para aceitar o modelo mais complexo ao invés do modelo mais simples (nulo).

free_fit ← two free parameters

equal_rates_fit ← one free parameter

Likelihood ratio test

$LR = -2 * (\log(lik_{nested}) - \log(lik_{free}))$

LR = 28.93; p valor < 0.01

Definindo – Likelihood ratio test

- No caso de modelos aninhados, podemos testar qual é o melhor modelo usando a razão das verossimilhanças.
- A razão das verossimilhanças testa se temos suporte para aceitar o modelo mais complexo ao invés do modelo mais simples (nulo).

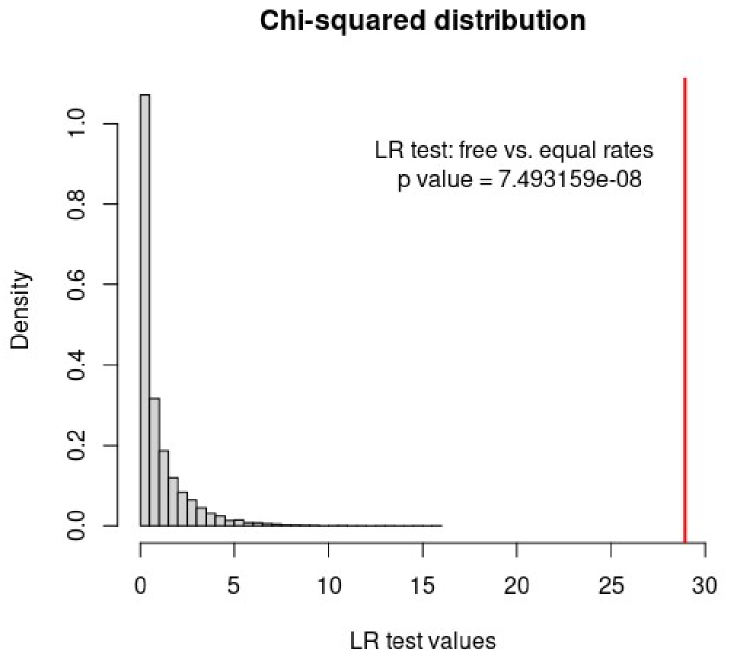
free_fit ← two free parameters

equal_rates_fit ← one free parameter

Likelihood ratio test

$LR = -2 * (\log(lik_{nested}) - \log(lik_{free}))$

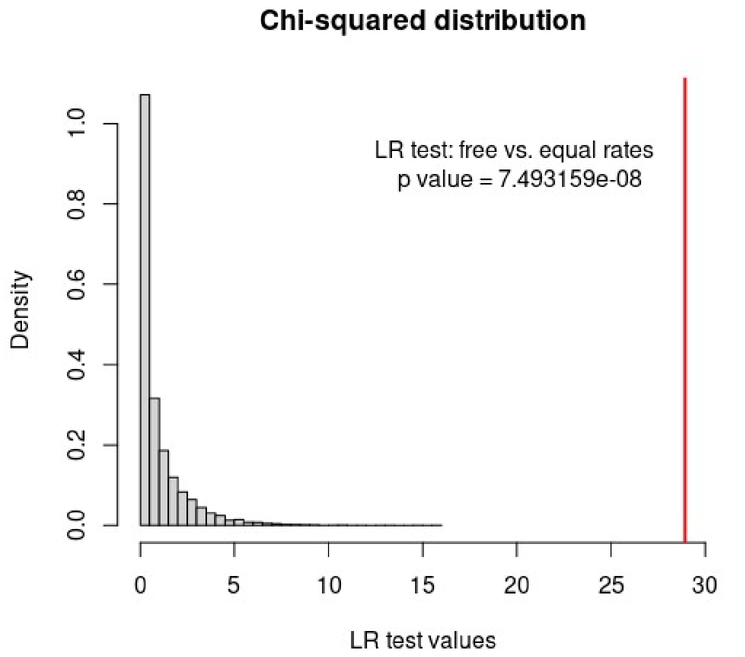
$LR = 28.93$; p valor < 0.01



Definindo – Likelihood ratio test

- No caso de modelos aninhados, podemos testar qual é o melhor modelo usando a razão das verossimilhanças.
- A razão das verossimilhanças testa se temos suporte para aceitar o modelo mais complexo ao invés do modelo mais simples (nulo).

Tutorial



Definindo – Likelihood ratio test

- Limitações do likelihood ratio test:
  - Somente par à par (problemas com testes múltiplos).
  - Unidirecionalidade (H_0 é sempre o modelo simples).
  - Somente para modelos aninhados.

Definindo – Akaike Information Criteria

- O **Critério de Informação de Akaike (AIC)** é uma forma de selecionar um modelo entre um grupo (pré determinado) de modelos.
- Diferente do teste de razão de verossimilhança, o AIC é usado para ranquear modelos (não tem um p valor).
- **NOTA:** Eu nunca vi a sigla CIA ser usada!

Likelihood ratio	**AIC**
nested models	any model
directional test	ranking factor
same data	same data

Definindo – Akaike Information Criteria

- O modelo escolhido usando AIC será aquele que minimiza a distância de Kullback-Leibler entre o modelo e a verdade.
- Essa distância mede o quanto o modelo deixa de “acomodar” os dados observados. Quantos “bits” de informação teríamos que incluir no modelo para ele explicar tanto quanto o modelo verdadeiro.
- Se essa distância for alta, então o modelo atual deixa de explicar muito dos dados e não é o melhor modelo para os dados.

Definindo – Akaike Information Criteria

- Em teoria da informação, temos que assumir que existe um modelo que gerou os dados. Isso é importante para formalizar a definição estatística do AIC.
- Note que isso também acontece com a estimativa MLE. O método de MLE retorna um único vetor de parâmetros. Sem nenhuma estimativa de incerteza.
- Isso é **o contrário** do que pensamos. Sabemos que o modelo é uma simplificação da realidade e que existe uma variância associada às estimativas.

Definindo – Akaike Information Criteria

- Em teoria da informação, temos que assumir que existe um modelo que gerou os dados. Isso é importante para formalizar a definição estatística do AIC.
- Note que isso também acontece com a estimativa MLE. O método de MLE retorna um único vetor de parâmetros. Sem nenhuma estimativa de incerteza.
- Isso é **o contrário** do que pensamos. Sabemos que o modelo é uma simplificação da realidade e que existe uma variância associada às estimativas.

Estatística Bayesiana resolve esse problema condicionando o modelo na observação dos dados.

Nota sobre “verdade” e “modelo verdadeiro”

LB: p628 state “Most would concede that it is unlikely that Truth is in our model set.” Of course not, only models are in the model set -- truth is not a model. One cannot have a model set such as $\{g_1, g_2, g_3, Truth, g_4, \ldots, g_R\}$. Truth cannot be in a model set – this statement obstructs clear thinking about this science philosophy issue. Rather, consider the model set $\{g_1, g_2, g_3, g_4, \ldots, g_R\}$; now ask if any one of these *models* is an exact representation of truth or full reality. That is, “is the (supposed) true *model* in the model set?” It is critical here to clearly distinguish between the concepts of truth or full reality and a mathematical model.

Fonte: http://sites.warnercnr.colostate.edu/kenburnham/model-selection-issues/

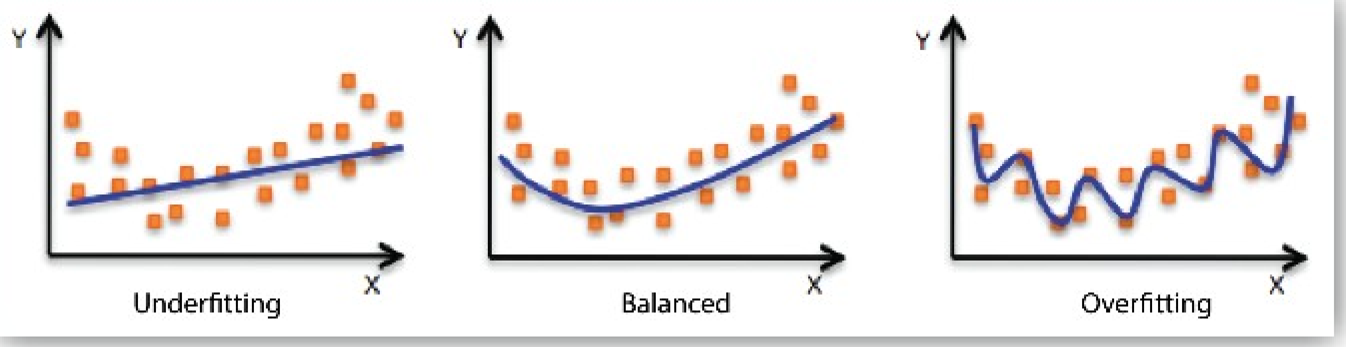
Nota sobre “verdade” e “modelo verdadeiro”

We have argued (Burnham and Anderson 2002: 20) that there are no true models in the biological sciences, certainly not in the sense that real data literally are produced from such a model. Real data do not come from (parametric) mathematical models as is often assumed in some statistical literature. The unfounded notion of a true model producing the data is the thinking of a mathematician not of an applied scientist.

Fonte: http://sites.warnercnr.colostate.edu/kenburnham/model-selection-issues/

Definindo – Akaike Information Criteria

- O AIC busca selecionar o modelo mais próximo do modelo verdadeiro levando em conta o número de parâmetros.
- Imagine que o “modelo” que tem o ajuste máximo aos dados terá número de parâmetros igual ao número de observações.



O AIC na prática

$$\textbf{AIC} = -2 \log(\textit{lik}) + 2\,k$$

- $\log(lik)$ = logaritmo da verossimilhança
- k = número de parâmetros livres

O AIC na prática

$$\mathbf{AIC = -2\ log(\mathit{lik}) + 2\ \mathit{k}}$$

- log(*lik*) = logaritmo da verossimilhança
- *k* = número de parâmetros livres

Rank	**Model**	$\sigma^2_{1,1}$	$\sigma^2_{1,2}$	$\sigma^2_{2,1}$	$\sigma^2_{2,2}$	r_1	r_2	$\log(L)$	AIC	w
1	model 3	0.23	–	0.06	–	0.34	-0.31	55.11	-98.22	0.28
2	model 3c	0.23	–	0.05	0.10	0.33	-0.31	56.04	-98.08	0.26
3	model 3b	0.21	0.27	0.06	–	0.33	-0.32	55.32	-96.63	0.13
4	model 4	0.21	0.28	0.05	0.10	0.32	-0.34	56.29	-96.58	0.12

O AIC na prática

$$\mathbf{AIC = -2\ log(\mathit{lik}) + 2\ \mathit{k}}$$

- log(*lik*) = logaritmo da verossimilhança
- *k* = número de parâmetros livres

Rank	**Model**	$\sigma^2_{1,1}$	$\sigma^2_{1,2}$	$\sigma^2_{2,1}$	$\sigma^2_{2,2}$	r_1	r_2	$\log(L)$	AIC	w
1	model 3	0.23	–	0.06	–	0.34	-0.31	55.11	-98.22	0.28
2	model 3c	0.23	–	0.05	0.10	0.33	-0.31	56.04	-98.08	0.26
3	model 3b	0.21	0.27	0.06	–	0.33	-0.32	55.32	-96.63	0.13
4	model 4	0.21	0.28	0.05	0.10	0.32	-0.34	56.29	-96.58	0.12

O AIC na prática - ΔAIC

- Geralmente usamos o ΔAIC para reportar o AIC entre modelos.
- ΔAIC é a diferença entre o menor AIC e o modelo atual.

```
> aic <- c(-98.22, -98.08, -96.63, -96.58)
> min_aic <- min( aic )
> delta_aic <- aic - min_aic
> min_aic
[1] -98.22
> delta_aic
[1] 0.00 0.14 1.59 1.64
> 
```

O AIC na prática - ΔAIC

- Geralmente usamos o ΔAIC para reportar o AIC entre modelos.
- ΔAIC é a diferença entre o menor AIC e o modelo atual.

```
> aic <- c(-98.22, -98.08, -96.63, -96.58)
> min_aic <- min( aic )
> delta_aic <- aic - min_aic
> min_aic
[1] -98.22
> delta_aic
[1] 0.00 0.14 1.59 1.64
> 
```

Melhor modelo terá ΔAIC de 0.0

O AIC na prática - ΔAIC

- Geralmente usamos o ΔAIC para reportar o AIC entre modelos.
- ΔAIC é a diferença entre o menor AIC e o modelo atual.

```
> aic <- c(-98.22, -98.08, -96.63, -96.58)
> min_aic <- min( aic )
> delta_aic <- aic - min_aic
> min_aic
[1] -98.22
> delta_aic
[1] 0.00 0.14 1.59 1.64
>
```

Distância relativa dos modelos

Melhor modelo terá ΔAIC de 0.0

Definindo – ΔAIC

```
> delta_aic
[1] 0.00 0.14 1.59 1.64
>
```

Distância relativa dos modelos

$$\Delta_i = \mathrm{AIC}_i - \mathrm{AIC}_{\min},$$

Hence, Δ_i is the information loss experienced if we are using fitted model g_i rather than the best model, $g_{\min}$, for inference. These Δ_i allow meaningful interpretation without the unknown scaling constants and sample size issues that enter into AIC values.

Definindo – ΔAIC

The Δ_i are easy to interpret and allow a quick strength-of-evidence comparison and ranking of candidate hypotheses or models. The larger the Δ_i, the less plausible is fitted model i as being the best approximating model in the candidate set. It is generally important to know which model (hypothesis) is second best (the ranking), as well as some measure of its standing with respect to the best model. Some simple rules of thumb are often useful in assessing the relative merits of models in the set: Models having $\Delta_i \leq 2$ have substantial support (evidence), those in which $4 \leq \Delta_i \leq 7$ have considerably less support, and models having $\Delta_i > 10$ have essentially no support. These rough guidelines have similar counterparts in the Bayesian literature (Raftery 1996).

Definindo – ΔAIC

The Δ_i are easy to i … nce
comparison and rank … The
larger the Δ_i, the les … est
approximating model … to
know which model (h … ell
as some measure of it … me
simple rules of thumb are often useful in assessing the relative merits of models in the set: Models having $\Delta_i \leq 2$ have substantial support (evidence), those in which $4 \leq \Delta_i \leq 7$ have considerably less support, and models having $\Delta_i > 10$ have essentially no support. These rough guidelines have similar counterparts in the Bayesian literature (Raftery 1996).

Infelizmente eu não conheço nenhum trabalho fazendo uma avaliação formal desses limites quando aplicados para situações reais em Evolução e Ecologia.

(onde modelos são geralmente complexos, tamanho amostral reduzido, e os modelos estão bem longe da verdade).

Definindo – ΔAIC

Naive users often question the importance of a $\Delta_i = 10$ when the two AIC values might be, for example, 280,000 and 280,010. The difference of 10 here might seem trivial. In fact, large AIC values contain large scaling constants, while the Δ_i are free of such constants. Only these differences in AIC are interpretable as to the strength of evidence.

Definindo – ΔAIC

Definindo – ΔAIC