

# Abordagens e métodos estatísticos em Ecologia

## Seleção de modelos



*Prof. Marcus Vinícius Vieira – Instituto de Biologia UFRJ*

# Seleção de modelos

- Como trabalhar e comparar hipóteses múltiplas ? (= diferentes modelos)
- Como medir a evidência a favor de uma hipótese em comparação com outra(s)?

# Verossimilhança Máxima

- Se a probabilidade de um evento  $x$  depende de parâmetros  $p$  de um modelo, escrevemos

$$P(x | p)$$

- mas quando falamos de verossimilhança

$$L(p | x)$$

(a verossimilhança dos parâmetros considerando o evento  $x$ )

- **Probabilidade**

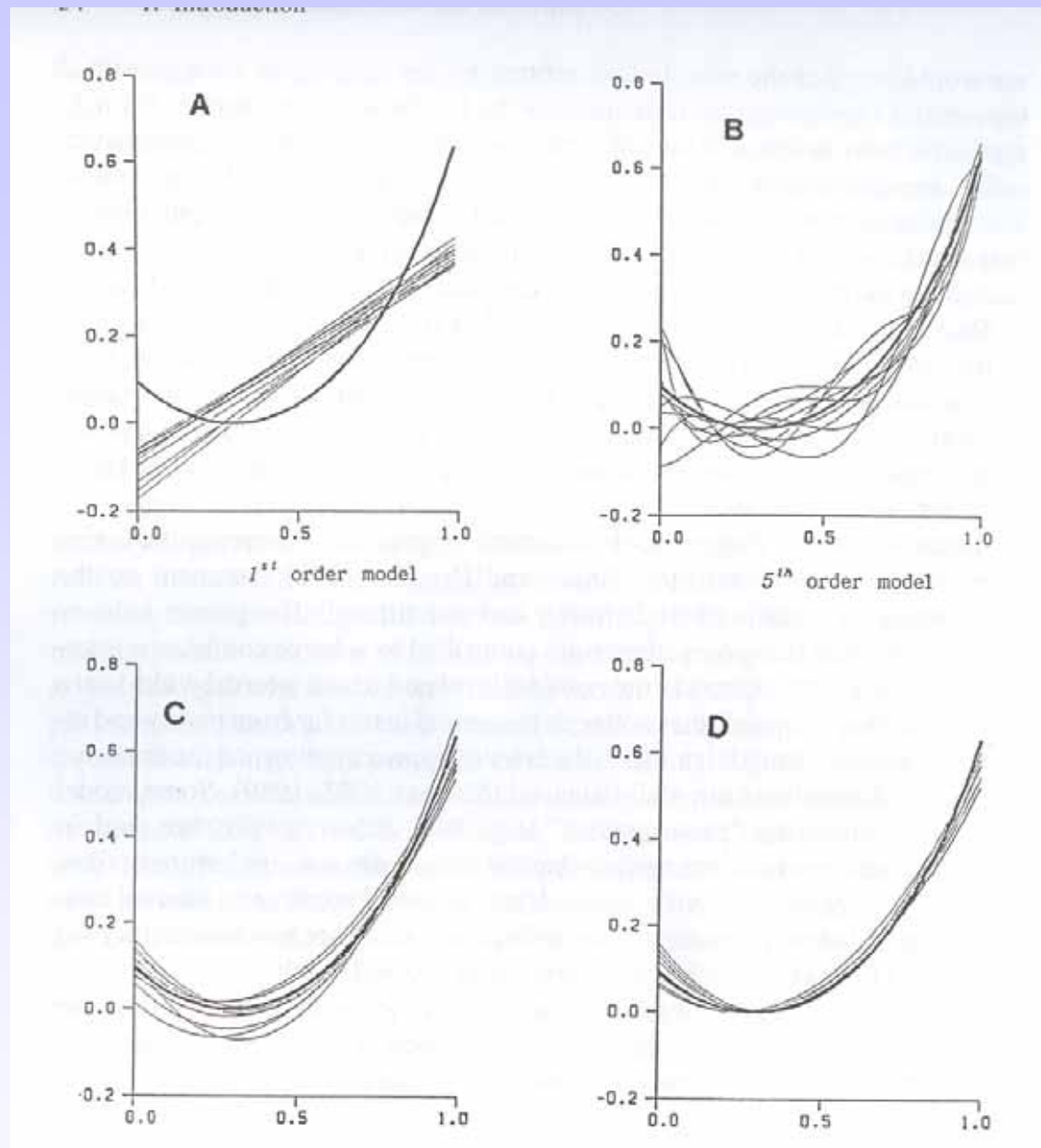
Sabendo os parâmetros -> Previsão de resultado

- **Verossimilhança**

Observação de resultado -> Estimativa de parâmetros

# Verossimilhança Máxima

- Pode ser estendida para comparar hipóteses
- Princípio da verossimilhança (controverso)  
(*Likelihood principle*):  
“Toda a *informação* de uma amostra está contida em sua função de verossimilhança”
- Lei de verossimilhança (*Likelihood law*):  
“A razão entre verossimilhanças seria uma medida da *evidência* de uma hipótese em relação a outra”



*Prof. Marcus Vinícius Vieira – Instituto de Biologia UFRJ*

# Entropia de Boltzmann

- A entropia termodinâmica: constante de proporcionalidade  $k$  (de Boltzmann) multiplicada pelo logaritmo do número de micro-estados independentes ( $w$ ) disponíveis ao sistema:

$$S = k \log (w)$$

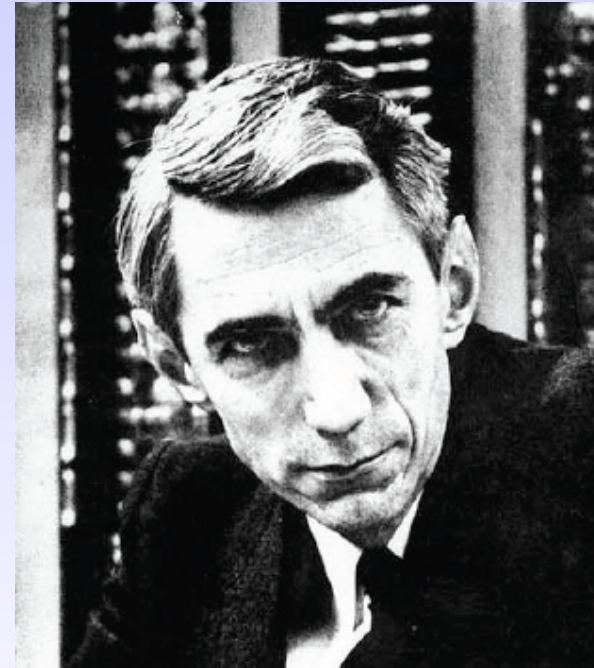
- Mede o grau de desorganização de um sistema termodinâmico
- Quando a concentração de partículas é homogênea a desorganização e a entropia são máximas
- Shannon utilizou o mesmo conceito para o grau de entropia "de informação" ( $H$ ) a partir da fórmula termodinâmica de Boltzmann ( $S$ ).



Figure 1 - Boltzmann's Theorem Engraved on his Tomb in Vienna (1906)

# Entropia de Claude Shannon

- $H = - \sum p_i \log p_i$
- Incerteza: quantidade de respostas possíveis que conhecemos, mas não sabemos qual delas é verdadeira
- Informação: redução da incerteza quando se obtém resposta a uma pergunta
- $-\log p_i$ : quanto menor  $p_i$ , maior o valor
- Quanto mais raro for a probabilidade de uma categoria, mais a incerteza será reduzida caso a categoria ocorra
- Ex. palavras que começam com z e palavras que começam com a



# Informação

- Com uma moeda viciada, que sempre dá cara, quanta informação é necessária para prever o resultado?
  - Nenhuma informação, já que o resultado é sempre o mesmo.
- Com uma moeda “justa”, é preciso mais informação?
  - Sim, mais informação já que os resultados têm igual probabilidade de ocorrência
- E uma moeda com algum viés?
  - Qualquer moeda com algum viés exigirá menos informação

# Perda de Informação dos modelos

- A realidade em sua totalidade não pode ser medida devido às nossas limitações sensoriais
- Temos apenas modelos
- Um modelo da realidade é apenas uma aproximação desta realidade
- Quanta informação é perdida por esta aproximação representada no modelo?
- Distância de Kullback-Leibler

# Distância de Kullback-Leibler

- Suponha que a realidade seja  $f(x)$ , isto é a probabilidade de  $x$  dada a realidade
- Podemos aproximar por  $g(x|p)$
- Se  $(f,g)$  é a informação perdida nesta aproximação de  $f$  com  $g$ ,
- $(f,g) \neq 0$  quase sempre
- $(f,g) = 0$  somente se  $f = g$

$$I(f, g) = \int f(x) \log \left( \frac{f(x)}{g(x|p)} \right) dx$$

# Distância K-L para dados categóricos

- Com binomiais
- Suponha que a probabilidade real seja  $p_1$  e  $p_2$
- As probabilidades segundo um modelo sejam  $\pi_1$  e  $\pi_2$

$$I(f, g) = \sum_{i=1}^2 p_i \log \left( \frac{p_i}{\pi_i} \right) = p_1 \log \left( \frac{p_1}{\pi_1} \right) + p_2 \log \left( \frac{p_2}{\pi_2} \right)$$

# Informação

$$I(f, g) = p_1 \log(p_1/\pi_1) + p_2 \log(p_2/\pi_2)$$

- Suponha que tenhamos estimativas perfeitas:
- $p_{i1} = p_1$
- $p_{i2} = p_2$

$$I(f, g) = p_1 \log(1) + p_2 \log(1) = p_1(0) + p_2(0) = 0$$

# Informação

- Suponha que  $p_1$  e  $p_2$  sejam ambos 0,20 maiores:

$$\begin{aligned} I(f, g) &= p_1 \log(1/1.2) + p_2 \log(1/1.2) = \log(1/1.2)(p_1 + p_2) = \log(1/1.2) \\ &= - 0,1823 \end{aligned}$$

- E se forem 0,40 maiores:

$$\begin{aligned} I(f, g) &= p_1 \log(1/1.4) + p_2 \log(1/1.4) = \log(1/1.4) \\ &= - 0,3365 \end{aligned}$$

# Informação

- Informação e entropia são aditivas -> escala logarítmica
- Probabilidade é multiplicativa
- Se  $P_1 = 0,3$ ,  $P_2 = 0,2$

Assumindo que são independentes:  $P_1 * P_2 = 0,06$

Em escala logarítmica:

$$\text{Log}(P_1) = -1,204$$

$$\text{Log}(P_2) = -1,609$$

$$e^{(-1,204 + -1,609)} = 0,06$$

# Mas não é possível conhecer a verdade!

- Verdade =  $p_1, p_2, f(x)$
- Não é preciso conhecê-la!
- Lembre-se:

$$\log(a/b) = \log(a) - \log(b)$$

$$I(f, g) = \sum_{i=1}^2 p_i \log \left( \frac{p_i}{\pi_i} \right) = p_1 \log \left( \frac{p_1}{\pi_1} \right) + p_2 \log \left( \frac{p_2}{\pi_2} \right)$$

$$I(f, g) = p_1 \log(p_1) - p_1 \log(\pi_1) + p_2 \log(p_2) - p_2 \log(\pi_2)$$

$$= \sum_i p_i \log(p_i) - \sum_i p_i \log(\pi_i)$$

# A verdade é constante, modelos não

- Diferenças em  $I(f,g)$  entre  $g$ 's nos dizem o que precisamos: qual modelo mais se aproxima da realidade

$$I(f, g_1) = \sum_i p_i \log(p_i) - \sum_i p_i \log(\pi_{i,1})$$

$$I(f, g_2) = \sum_i p_i \log(p_i) - \sum_i p_i \log(\pi_{i,2})$$

$$I(f, g_1) - I(f, g_2) = - \sum_i p_i \log(\pi_{i,1}) - \sum_i p_i \log(\pi_{i,2})$$

- A realidade pode ser cancelada então ...
- Não precisamos conhecê-la!

# A verdade é relativa



- Só precisamos das estimativas  $p_i$
- Hirotugu Akaike (1971) demonstrou que a distância relativa à realidade (incomensurável),  $I(f,g)$ , é proporcional:
  - ao logaritmo da verossimilhança máxima do modelo
  - subtraída de um valor proporcional ao número de parâmetros no modelo ( $K$ )

$$\sum_i p_i \log(\pi_i) \propto \log(\text{Pr}(\text{data}|\hat{p})) - K$$

# A verdade é relativa



- Akaike definiu um critério de informação, AIC, “Akaike Information Criterion”

$$AIC = -2 \log(\text{Pr}(\text{data}|\hat{p})) + 2K$$

- AIC estima a distância relativa de um modelo na sua verossimilhança máxima a processo real
- Valores menores indicam modelos mais próximos à realidade, ou ...
- que têm menor perda de informação em relação a realidade

# AICc

- A demonstração de que AIC é uma medida relativa da distância à realidade vale somente para grandes quantidades de dados
- Quando existem muitos parâmetros a serem estimados em relação à quantidade de evidência (dados), AIC tem um viés
- Correção de “2a. Ordem”: AICc

$$AIC_c = AIC + \frac{2K(K + 1)}{n - K - 1}$$

- $n$  = número de observações (dados)
- Conforme  $n$  aumente, AICc  $\rightarrow$  AIC

# Seleção de modelos na prática



- (1) Defina o conjunto de modelos a serem comparados
- (2) Obtenha estimativas de ajuste de cada um aos dados através de verossimilhança máxima
- (3) Calcule o AIC ou AICc de cada um
- (4) Calcule o ajuste relativo:  $\Delta_i = AIC_i - AIC_{\min}$

# Seleção de modelos na prática

- É possível converter valores  $\Delta_i$  em estimativas de verossimilhança do modelo considerando os dados:

$$\Pr(g_i|\text{data}) \propto \exp\left(-\frac{1}{2}\Delta_i\right)$$

- $\Pr(g_i|\text{data})$  é a probabilidade que o modelo  $i$  é o melhor no conjunto de modelos comparado
- AIC presume que todos os modelos são “errados”, não são a realidade
- Estas verossimilhanças relativas são mais fáceis de interpretar, mas não há um limiar de “significância” baseado na teoria da informação e de seleção de modelos

# Seleção de modelos na prática

- Pesos de Akaike são bastante úteis:

$$w_i = \frac{\exp(-\frac{1}{2}\Delta_i)}{\sum_j \exp(-\frac{1}{2}\Delta_j)}$$

- $w_i$  varia de 0 a 1 e estima o peso de evidência a favor de um modelo  $i$ , considerando o conjunto de modelos comparado
- Pesos de Akaike podem ser usados para previsões baseadas na média dos modelos

## Scaling body mass and use of space in three species of marsupials in the Atlantic Forest of Brazil

MARCUS VINÍCIUS VIEIRA<sup>1\*</sup> AND ANDRÉ DE ALMEIDA CUNHA<sup>2</sup>

<sup>1</sup>Laboratório de Vertebrados, Departamento de Ecologia, Universidade Federal do Rio de Janeiro, CP 68020, Rio de Janeiro RJ, CEP 21941-590, (Email: [mrvieira@biologia.ufrj.br](mailto:mrvieira@biologia.ufrj.br), [mrvieira@gmail.com](mailto:mrvieira@gmail.com)), and <sup>2</sup>Pós-Graduação em Ecologia, Conservação e Manejo da Vida Silvestre, ICB, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

As a measurement of area, and assuming a circular shape  
DHR would be proportional to the square of its linear distance

Hence proportional to the square of day range

$$DHR \approx DR^2$$

If Day range  $\approx M^{0.25}$  (Carbone *et al.* 2005), then

$$DHR \approx \left( M^{0.25} \right)^2 \approx M^{0.5}$$

# Los marsupiales



*Didelphis aurita* (ca. 1800 g)



*Philander frenatus*

(ca. 400 g)



*Metachirus nudicaudatus* (ca. 450 g) Prof. Marcus Vinícius Vieira – Instituto de Biologia UFRJ

# Métodos

- Gradeados de trampas y animales liberados con carretel de rastreo



## Vieira & Cunha (2008)

**Table 1.** Performance of models predicting daily home range (DHR) and its intensity of use (IU) of individuals of three species of didelphid marsupials.

	Model	Variables	d.f.	<i>K</i>	AICc	$\Delta_i$	$w_i$
DHR	1	Thread + Body mass	2	4	185.229	0.000	0.501
	2	Thread + Body mass + Species	4	5	186.544	1.315	0.260
	3	Thread + Species	3	4	188.290	3.061	0.109
	4	Thread + Body mass + Species + 2*3	6	6	190.686	5.457	0.033
	5	Thread + Body mass + Species + 1*3	6	6	190.858	5.629	0.030
	6	Thread + Body mass + Species + 1*2*3	6	6	190.899	5.670	0.029
	7	Thread	1	3	192.517	7.288	0.013
IU	1	Thread + Body mass	2	4	82.643	0.000	0.501
	2	Thread + Body mass + Species	4	5	83.958	1.315	0.260
	3	Thread + Species	3	4	85.704	3.061	0.109
	4	Thread + Body mass + Species + 2*3	6	6	88.100	5.457	0.033
	5	Thread + Body mass + Species + 1*3	6	6	88.272	5.629	0.030
	6	Thread + Body mass + Species + 1*2*3	6	6	88.313	5.670	0.029
	7	Thread	1	3	89.931	7.288	0.013

*K* is number of parameters of the model, AICc is the Akaike Information Criteria corrected for small samples,  $\Delta_i$  is the difference in AICc between the given model and the best model, and  $w_i$  is the Akaike weight, the relative likelihood of a model given the data and the set of models analyzed. Asterisks indicate interactions between variables, coded by numbers (Thread = 1, Body mass = 2, Species = 3).

## Vieira & Cunha (2008)

**Table 2.** Parameters of the best models predicting daily home range (DHR) and its intensity of use (IU) of individuals of three species of didelphid marsupials.

Model	Variables	DHR			IU			Wald <i>d.f.</i>	statistic	<i>P</i>
		Estimate	<i>SE</i>	<i>CL</i>	Estimate	<i>SE</i>	<i>CL</i>			
1	Intercept	-3.257	1.104	2.164	1.629	0.552	1.082	1	8.7	0.003
	Thread length	1.474	0.116	0.228	0.263	0.058	0.114	1	161.0	<0.001
	Body mass	0.435	0.136	0.267	-0.218	0.068	0.123	1	10.2	0.001
	Scale	0.809	0.067	0.130	0.405	0.033	0.065	1	148.0	<0.001
2	Intercept	-2.696	1.293	2.533	1.348	0.646	1.267	1	4.4	0.037
	Thread length	1.450	0.115	0.225	0.275	0.057	0.112	1	160.0	<0.001
	Body mass	0.359	0.176	0.345	-0.179	0.088	0.172	1	4.2	0.041
	Species							2	3.0	0.218
	<i>D. aurita</i>	0.070	0.161	0.316	-0.035	0.081	0.158	1	0.2	0.665
	<i>M. nudicaudatus</i>	0.198	0.145	0.283	-0.099	0.072	0.142	1	1.9	0.171
Scale	0.793	0.065	0.128	0.396	0.033	0.064	1	148.0	<0.001	
3	Intercept	-0.308	0.564	1.105	0.154	0.282	0.552	1	0.3	0.585
	Thread length	1.406	0.116	0.227	0.297	0.058	0.113	1	147.6	<0.001
	Species							2	9.0	0.011
	<i>D.aurita</i>	0.281	0.127	0.249	-0.141	0.064	0.125	1	4.9	0.027
	<i>M.nudicaudatus</i>	0.145	0.146	0.287	-0.073	0.073	0.143	1	1.0	0.320
	Scale	0.815	0.067	0.131	0.407	0.033	0.066	1	148.0	<0.001

Wald statistic and associated probability (*P*) estimate the significance of regression coefficients. For models that include thread length as a covariate, Wald statistic for DHR and IU is the same. In Models 2 and 3, *P. frenatus* is the reference species to which the other two are compared

# Cálculo de AIC na prática

- Modelos lineares generalizados (Generalized linear models):

$$\text{AIC} = -2\ln(\text{verossimilhança}) + 2K$$

$$-2\ln(\text{verossimilhança}) = \text{desviança}$$

$$\text{AICc} = -2\ln(\text{verossimilhança}) + 2K(K+1) / (n - K - 1)$$

K = número de parâmetros

$$\text{AICc} = \text{AIC} * (K + 1) / (n - K - 1)$$

- Regressão por quadrados mínimos comum (Ordinary least squares regression)

$$\text{AIC} = n * \ln(\text{RSS}/n) + 2K$$

onde n = número de observações (dados)

RSS = residual sum of squares = Soma dos Quadrados dos Resíduos

# Referências

- Burnham, K.P., & D. R. Anderson. 2002. Model selection and multimodel inference: A practical information-theoretic approach, 2nd ed (Caps. 1 e 2). Springer-Verlag, Heidelberg.
- Epstein, I. 1986. Teoria da Informação. Coleção Primeiros Passos, Editora Ática, São Paulo.
- McElreath, R. Statistical thinking in evolutionary ecology (<http://xcelab.net/rm/?p=198>).
- Vieira & Cunha (2008). Austral Ecology 33:872-879