Mini-lecture on Likelihoods and AIC

EFB 370: Population Ecology. Lecture 6b

Dr. Elie Gurarie

2022-02-16

A few slides from Lecture 6

```
library(AICcmodavg)

#define list of models
models <- list(model1, model2, model3)

#specify model names
mod.names <- c('disp.hp.wt.qsec', 'disp.qsec', 'disp.wt')

#calculate AIC of each model
aictab(cand.set = models, modnames = mod.names)

Model selection based on AICc:

               K   AICc Delta_AICc AICcWt Cum.Wt     LL
disp.hp.wt.qsec 6 162.43       0.00   0.83   0.83 -73.53
disp.wt         4 165.65       3.22   0.17   1.00 -78.08
disp.qsec       4 173.32      10.89   0.00   1.00 -81.92
```

In general you want the model with the lowest AIC

Maximum Liklihood

- Requires a data set and an explicit model of evolution
- Will look to maximize likelihood (a statistical process that looks to match statistical distributions with observed data) over a tree
- Not only supports different rates of evolution across different branches of the tree
- Will produce multiple equally likely trees which can then be used to bootstrap
- Robust to violations of underlying assumptions (to a point)
- Much slower

What is a **Likelihood**?
What is a **Maximum Likelihood Estimate**?
What is **AIC**?
What does this have to do with **model selection** and **parsimony**?

An example!

Oakie Island

Orange Island

little known fact …

your professor once played in a band called the Black Squirrels.

The question:

Oakie Island

Orange Island

What is the “best” model for squirrel morph distribution?

Data and Models

Data / observations: X_{ij}

island	a. Orange	b. Oakie
squirrel 1	$X_{a,1}$ = light	$X_{b,1}$ =light
squirrel 2	$X_{a,2}$ = light	$X_{b,2}$ dark

Models = Hypotheses

model	n. parameters
M1: $P(X_{ij} = light) = p = 0.5$	1
M2: $p = 0.75$	1
M3: $p_a = 1; p_b = .5$	2
M4: $p_{a,1} = 1;\ \ p_{b,1} = 1$ $p_{a,2} = 1;\ \ p_{b,2} = 0$	4

Likelihoood (of a model)

Product of **probabilities** of **data** given **model**.

$$\mathcal{L}(model) = \prod_{i=1}^{n} Pr(X|model)$$

- We **never** care about the **absolute** value of the likelihood!
- Only the *relative* value of the likelihood.

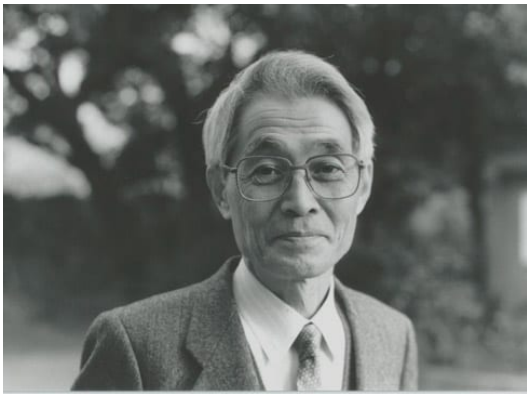
Squirrel Models:

model	likelihood	
M1: $p = 0.5$	$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$	0.0625
M2: $p = 0.75$	$\frac{3}{4} \times \frac{3}{4} \times \frac{3}{4} \times \frac{1}{4} = \frac{27}{256}$	0.1055
M3: $p_a = .5; p_b = 1$	$1 \times 1 \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$	0.25
M4: $p_{a,1} = 1;\ \ p_{b,1} = 1$ $p_{a,2} = 1;\ \ p_{b,2} = 0$	$1 \times 1 \times 1 \times 1$	1

$$\mathcal{L}(\mathcal{M}_4) > \mathcal{L}(\mathcal{M}_3) > \mathcal{L}(\mathcal{M}_2) > \mathcal{L}(\mathcal{M}_1)$$

A(kaike) Information Criterion

A good fit is great! But it is useless if it uses too much information (too many parameters). This is *overfitting*. One parameter per data point is TOO MANY parameters!



Hirotugo Akaike 赤池 弘次 (1927-2006)

Simple formula:

$$AIC = -2\log(\mathcal{L}) + 2k$$

(where k is the number of parameters)

- Better fit = higher $\mathcal{L}$ = lower AIC.
- Too complicated = more k = higher AIC.

Lowest AIC is "best" model

Compute AIC

model	likelihood	log-likelihood	k	AIC
M1:	0.0625	-2.77	1	7.55
M2:	0.1055	-2.25	1	**6.50**
M3:	0.25	-1.39	2	6.77
M4:	1	0	4	8

$AIC_2 < AIC_3 < AIC_1 < AIC_4$

Most *parsimonious* model is M2!

Compute AIC in R

```
L <- c(1/2^4, 27/256, 1/4, 1)
k <- c(1,1,2,4)

data.frame(L, k, AIC = - 2 * log(L) + 2 * k)
```

```
##           L k      AIC
## 1 0.0625000 1 7.545177
## 2 0.1054688 1 6.498681
## 3 0.2500000 2 6.772589
## 4 1.0000000 4 8.000000
```


Let's add one more observation ...

Oakie Island

Orange Island

$$X_{b,3} = dark$$

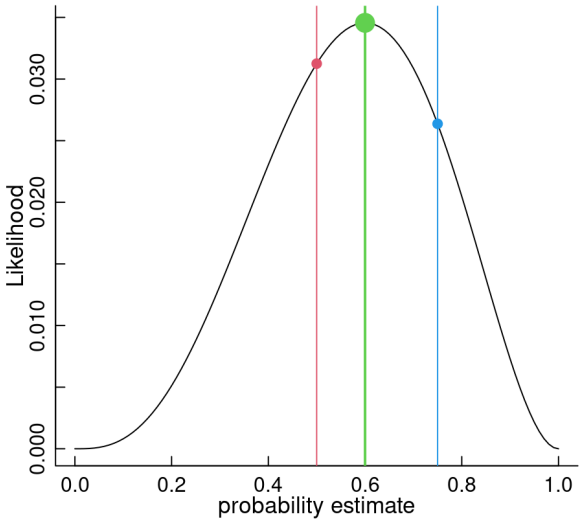
Updated squirrel models:

model	probs	$\mathcal{L} = \Pi P(X\|M)$	$\mathcal{L}$	k	AIC
M1	$p = \frac{1}{2}$	$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}$	0.03125	1	8.93
M2	$p = \frac{3}{4}$	$\frac{3}{4} \times \frac{3}{4} \times \frac{3}{4} \times \frac{1}{4} \times \frac{1}{4}$	0.02637	1	9.27
M2b	$p = \frac{3}{5}$	$\frac{3}{5} \times \frac{3}{5} \times \frac{3}{5} \times \frac{2}{5} \times \frac{2}{5}$	0.0346	1	8.73
M3	$p_a = .5; p_b = 1$	$1 \times 1 \times \frac{1}{2} \times \frac{1}{2} \times 0$	0 (!!)	2	∞
M3b	$p_a = \frac{1}{2}; p_b = \frac{2}{3}$	$\frac{1}{2} \times \frac{1}{2} \times \frac{2}{3} \times \frac{2}{3} \times \frac{1}{3}$	0.037	2	10.6

Updated (1 parameter) squirrel models:

$\hat{p} = 3/5$ is the **maximum likelihood estimate** of the probability that a squirrel is light morph.

model	probs	$\mathcal{L} = \Pi P(X\|M)$	$\mathcal{L}$
M1	$p = \frac{1}{2}$	$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}$	0.03125
M2	$p = \frac{3}{4}$	$\frac{3}{4} \times \frac{3}{4} \times \frac{3}{4} \times \frac{1}{4} \times \frac{1}{4}$	0.02637
M2b	$p = \frac{3}{5}$	$\frac{3}{5} \times \frac{3}{5} \times \frac{3}{5} \times \frac{2}{5} \times \frac{2}{5}$	0.0346



Likelihoods are the *SOUL of INFERENCE*!

They allow you to FIT models (i.e estimate **parameters**) via maximization;

They are essential for **model selection**, e.g. with AIC;

They underlie **Bayesian approaches**;

They are used **throughout** Ecology!

- **Phylogenies**, **Coalescence Trees**
- **Population dynamics**
- **Species Distributions**
- **Habitat Selection**
- **Movement Ecology**
- **Survival Analysis**
- and on and on and on and on and on

```
library(AICcmodavg)

#define list of models
models <- list(model1, model2, model3)

#specify model names
mod.names <- c('disp.hp.wt.qsec', 'disp.qsec', 'disp.wt')

#calculate AIC of each model
aictab(cand.set = models, modnames = mod.names)

Model selection based on AICc:

               K   AICc Delta_AICc AICcWt Cum.Wt     LL
disp.hp.wt.qsec 6 162.43       0.00   0.83   0.83 -73.53
disp.wt         4 165.65       3.22   0.17   1.00 -78.08
disp.qsec       4 173.32      10.89   0.00   1.00 -81.92
```

In general you want the model with the lowest AIC