Model Selection and Inference

February 13, 2007

Model Selection Criteria

Readings Relevant for Today’s Discussion
B&A Chapter 2
Hilborn and Hobbs Ch. 7
Key Objectives for Today’s Class

- Understand the idea behind maximum likelihood estimation and the link to AIC
- Understanding of the basic structure of AIC model selection criteria
Principle of Parsimony

No. of Parameters

Bias^2

Variance

Theoretical optimum

Few

Many

No. of Parameters
Model Selection Approaches to Find “Optimal” Tradeoff

• Stepwise

• Mallows $C_p$

• Information-theoretic approaches
  
  • Bayesian Information Criteria (BIC)
  
  • Takeuchi’s Information Criteria (TIC)

• Akaike’s Information Criteria (AIC)
Focus on Akaike’s Information Criteria: A Useful Approach

- Selection of a (good) approximating model is key to good inference
- Relationship with log-likelihood function and MLE
- Simple, intuitive, useful tool
Which Model is “Closest”?  

Kullback-Leibler Distance Between Two Models

Underlying theory relates to KL Information, amount of information lost from truth to model
Akaike identified an estimator of relative KL Distance, based on Log likelihood function of empirical data.
I Know What You Are Thinking

Let’s Just Understand AIC

So, a guide to “the ABC’s of AIC”

Gail Olson

(a start, anyhow)
Akaike's Information Criteria

\[ \text{AIC} = -2\log \mathcal{L} + 2k \]

- Deviance (a measure of fit)
- \( k \) = number of ALL estimated parameters (guard against too many parameters)

Trade-off of bias and precision
Akaike’s Information Criteria

$$\text{AIC} = -2\log L + 2k$$

The BIG BLACK BOX … for most of us
Likelihood and Maximum Likelihood Estimation

\[ \text{AIC} = -2 \log L + 2k \]

The likelihood at its maximum value
Likelihood and Maximum Likelihood Estimation

• The Likelihood of any value of the parameters is the probability of the actual outcome given those values and the **probability model** of the distribution (example to follow)

• The higher the likelihood, the more likely are those values of the (set of) parameters **(of your biological model)**

• The set of parameter values at the greatest likelihood value are the MLEs- maximum likelihood estimates (e.g., the estimates of the betas in logistic regression)

• Thus, the algorithm chooses the set of parameter values that simultaneously maximizes the likelihood

• Likelihood goes up as fit improves (i.e., obs data == predicted values)
Likelihood and Maximum Likelihood Estimation

\[ L(\text{parameters} \mid \text{data}) \rightarrow \text{likelihood of the data, as a function of the parameters, given a prob. model} \]

- We always start with a probability distribution model (bull trout analysis = logit model)

- The better the correspondence of the data to the model, the greater the likelihood

- We identify through this process those values of the parameters that make the correspondence greatest (MLE) and the associated likelihood is our measure of relative “fit” among models
Likelihood and Maximum Likelihood Estimation

A Simple Example of Estimating a Proportion

We all know the ML estimator for the proportion:

$$\text{Proportion} = \frac{y}{n},$$

$y=$ number of events

$n=$ total trials

If we have 10 nests, and 6 are successful,
The proportion of successful nests $= \frac{6}{10} = 0.6$

Let’s obtain the same results through formal maximum likelihood estimation
Likelihood and Maximum Likelihood Estimation

\( \mathcal{L} \) (parameters | data) \( \rightarrow \) likelihood of the data, given the parameter estimates and underlying model

\( \mathcal{L} \) (p|y,n) =

- use binomial distribution,

\[
\begin{pmatrix} n \\ y \end{pmatrix} \cdot p^y \cdot (1 - p)^{n - y}
\]

Computes number of combinations of “n” taken “y” at a time
Likelihood and Maximum Likelihood Estimation

\[ L(p|y,n) = \binom{n}{y} p^y (1 - p)^{n - y} \]

\[ L(p|6,10) = \binom{10}{6} p^6 (1 - p)^{10 - 6} \]

- we now estimate the proportion by finding the max likelihood value
Estimating the Likelihood and MLE

- Calculus, taking first derivative of log likelihood with respect to $p$, setting result $= 0$, and solve for $p$

$$\frac{6}{p} - \frac{4}{1-p} = 0$$

$p = \frac{y}{n}$, the intuitive estimator is the ML estimator

$p = 0.6$

- Numerical approach: iteratively seek solutions that maximize $L$
Estimating the Likelihood and MLE

\[ \mathcal{L} (p|y, n) = \binom{10}{6} p^y (1 - p)^{n-y} \]

Possible Values of $p$  \[ \mathcal{L} \]

0  1E-18
0.05  2.6726E-06
0.10  0.00013
0.15  7781
0.20
0.30

Let’s use Excel as a simple exercise to compute values and find MLE

1.0
Likelihood Estimation: Small N

-2 * Log Likelihoo

Parameter Value

Likelihood or Log-Likelihood
-max indicates best (MLE)

“Most Likely”
Remains the Same

-2 Log-Likelihood
Now best = least
Shape of likelihood indicates precision; profile likelihood confidence intervals
(Allison 2000, p. 32-33)

An excellent way to illustrate confidence in your estimates without having to choose an arbitrary level.
You must use SAME Data when comparing Models in AIC framework

Here’s why…

Change sample OR response variable, relative Likelihoods are meaningless
Let’s estimate the sex ratio in this class using Likelihood theory?

• Imagine the class is actually a sample from some larger population… say it’s a random sample from the graduate student population

\[ n = ? \]
\[ Y \text{ (female)} = ? \]

\[ -2 \times \log\text{-Likelihood} = \]
\[ \text{MLE of } p = \]
Summary and Assignment

• Is the origin of the Likelihood function reasonably clear?

• Does the MLE make sense?

• I did NOT go over ΔAIC, AICc, QAIC, and Model Weights

Come to class next week with a 1 sentence statement of what these are and mean, and any questions you have on their properties or use based on Ch. 2 in B&A. (plan to turn these in to me)

We will go over this next Tuesday.
Tomorrow’s Lab:
(1) finalize Bull Trout global model
(2) go over reports (come to class with ??)
(3) SAS AIC macro (very useful tool)
(4) make sure you are ready to “fly” with your analysis