

559

Welcome

Fish 559; Day1: 13h00-13h30

Likelihoods are a reoccurring them.

- This workshop is centered around three principles of modeling
 - Formulation
 - Building models from first principles using biologically meaningful parameters
 - Likelihoods
 - Implementation
 - Fitting your model to data using TMB.
 - Likelihoods
 - Evaluation
 - Likelihoods
 - Residual diagnostics
 - Management risk

What Does TMB Do?

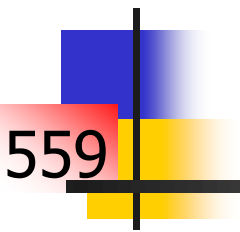
- Parameters are hypotheses describing nature.
- TMB finds the hypotheses that best describe your system.
- It can find the value of the parameter vector that minimizes a (complex) non-linear function of many variables and of many data sources.
- Uncertainty for models fitted using TMB can be computed by:
 - estimating asymptotic variance-covariance matrices;
 - computing likelihood profiles for parameters and model outputs - relatively how much better is one hypothesis over another; and
 - sampling parameter vectors from Bayesian posteriors using the Markov Chain Monte Carlo (MCMC) algorithm.

Why Use TMB?

- It is fast because the minimization algorithm uses analytical derivatives and a quasi-Newton algorithm.
- There is no need to supply the derivatives of the function to be minimized with respect to the parameters; these are computed automatically using Reverse Mode Autodifferentiation.
- TMB code is essentially C++ but includes:
 - commands to specify the function to be minimized; and
 - support for additional data structures (in particular it can deal with matrices).

Aims of this Workshop

- Introduce students to TMB.
- Conduct several in-class examples.
- Motivate projects for later in the class.



559

A: Maximum Likelihood, Likelihood Profiles, and Model Selection

Fish 559; Day1: 13h30-14h30

Maximum Likelihood

Given a set of data, and a probability model, maximum likelihood “chooses” the values for the parameters that make the data “most likely”.

$$L(D \mid p) \approx \text{Prob}(D \mid p)$$

The likelihood of the data D given the parameter(s) p is the probability of the data given the parameter(s).

Note:

Probability: knowing parameters \rightarrow predicting data.

Likelihood: knowing data \rightarrow parameter estimation.

Basics of Likelihood

If the data consists of two parts D_1 and D_2 , the likelihood of the data $D=D_1+D_2$ is the product of the likelihoods of each of D_1 and D_2 :

$$L(D \mid p) = L(D_1 \mid p)L(D_2 \mid p)$$

Data point 1

Data point 2

Further Likelihood Considerations

To compute the likelihood for a given set of parameters we need:

- ❖ The deterministic (model) relationship between the input variables (covariates) and the (expected) data:

$$L(a) = \ell_{\infty}(1 - e^{-\kappa a})$$

- ❖ How the data pertain to the model predictions – the “sampling” distribution (in this case normal), i.e.:

$$L = \frac{1}{\sqrt{2\pi}\sigma} e^{-(L(a) - L_a)^2 / 2\sigma^2}$$

Model prediction

Data

How we got here

- Remember:
 $L(D \mid p) \approx \text{Prob}(D \mid p)$
- Normal distribution:

$$P(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-(x-\mu)^2 / (2\sigma^2)}$$

- So: $L = \frac{1}{\sqrt{2\pi}\sigma} e^{-(L(a)-L_a)^2 / 2\sigma^2}$

Maximum Likelihood

Given a likelihood function, we find the maximum likelihood estimates for the parameters of the model by maximizing the log-likelihood, or minimizing the negative log-likelihood, i.e.:

$$\frac{\partial}{\partial p} \ln L(D | p) = 0$$

It is seldom the case that one can solve this equation, but it is possible (such as a linear regression).

Thoughts on likelihoods

It is traditional to assume that count data are Poisson distributed and yes-no data are binomial, but actual data are usually overdispersed. In these cases one should assume:

Negative binomial likelihood:

$$L(k | n, m, r) = \frac{\Gamma(r+k)}{k! \Gamma(r)} \left(\frac{r}{r+m} \right)^r \left(\frac{m}{r+m} \right)^k$$

Beta-binomial likelihood:

$$L(k | n, \alpha, \beta) = \frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k+1)} \frac{\Gamma(k+\alpha)\Gamma(n-k+\beta)}{\Gamma(n+\alpha+\beta)} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

Likelihood Profile-I

Likelihood profile is a way to estimate a confidence interval for the estimate of a model parameter when the model is fitted using maximum likelihood.

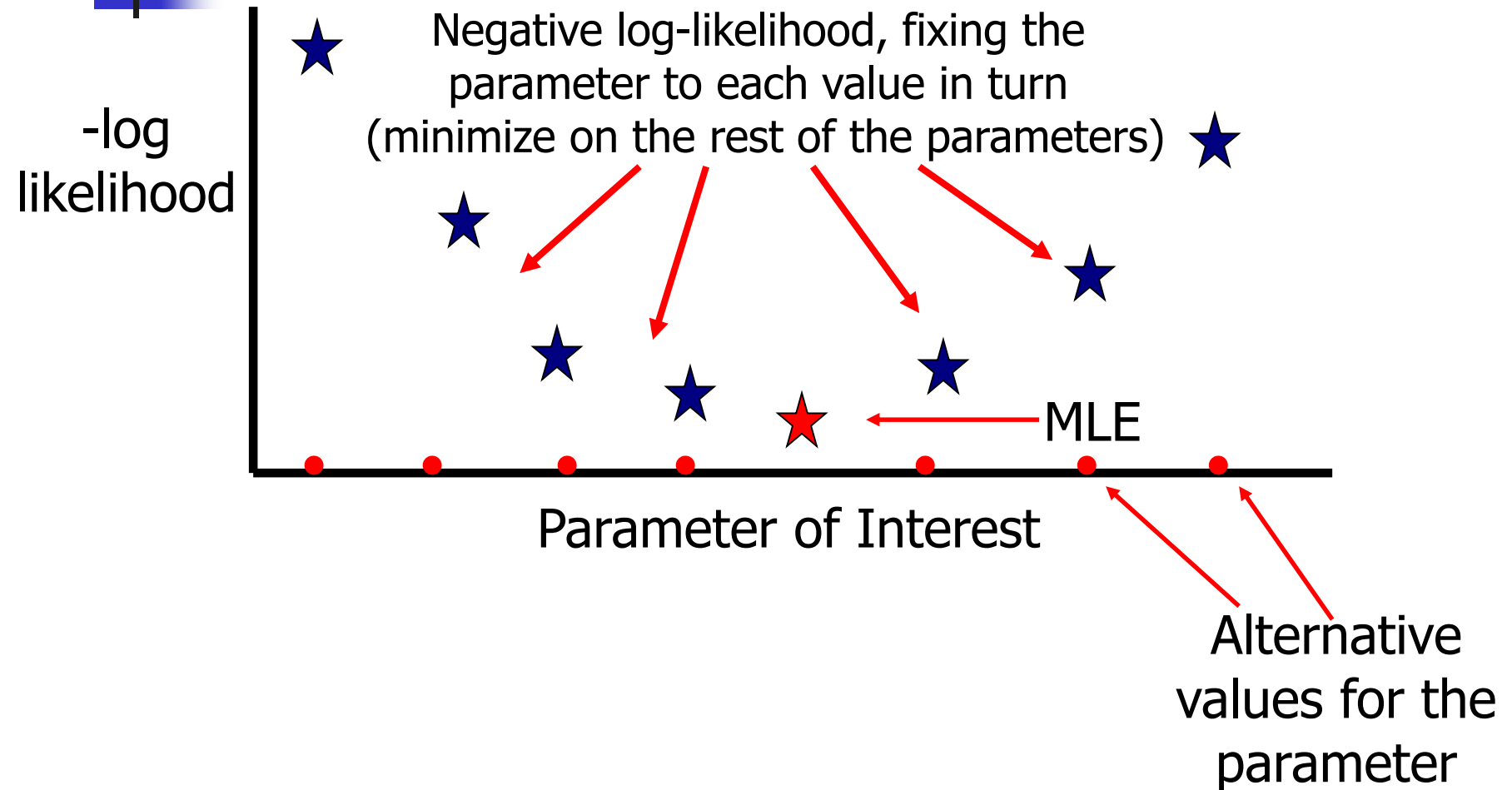
The problem. We have a model based on a vector of parameters $\underline{\theta}$. One of these parameters is " p " and we want a $x\%$ confidence interval for it. The negative log-likelihood function for this problem is $-\ell \ln L(D | \underline{\theta})$ or (separating out p), $-\ell \ln L(D | \underline{\theta}', p)$. Note: $\underline{\theta} = \{\underline{\theta}', p\}$.

Likelihood Profile-II

The steps in computing a likelihood profile:

- Fit the model to find the maximum likelihood estimates for the parameters and also the lowest negative log-likelihood $-\ell n L_{\text{Best}}$.
- Select a wide range of values for p .
- For each value of $p=A$, find the values for the parameters $\underline{\theta}'$ such that $-\ell n L(D | \underline{\theta}', p = A)$ is minimized. Record the negative log-likelihood for each value of p , $-\ell n L(D | \underline{\hat{\theta}}', p = A)$
- Plot $-\ell n L(D | \underline{\hat{\theta}}', p = A) - (-\ell n L_{\text{Best}})$ against p .

Likelihood Profile-I



Laplace approximation-I

For models with random effects (parameters u) and fixed effects (θ), we maximize the marginal likelihood, i.e.:

$$L(\theta) = \int_u e^{-f(u, \theta)} du$$

This integral is usually intractable so we use the Laplace approximation of the negative log of the marginal likelihood

Laplace approximation-II

$$-\ell n L(\theta) \sim -n \ell n \sqrt{2\pi} + 0.5 \log \det [H(\hat{u}(\theta), \theta)] + f(\hat{u}(\theta), \theta)$$

where $\hat{u}(\theta)$ is the maximum likelihood estimate of u conditional on the current value for θ , and $H[\cdot]$ is the hessian matrix for $f(u, \theta)$ with respect to u .

TMB Tricks

TMB “tricks” (according to Kasper Kristensen):

- The sparsity of the Hessian is detected and accounted for.
- The Laplace approximation is calculated using automatic differentiation.
- Automatic bias-corrections are applied when reporting a non-linear function of the random effects.
- The model can be parallelized.
- You can check the Laplace approximation using the “checkConsistency” function in TMB and the SIMULATE aspects (we will not cover this here).

Model Selection-I

Given N models that are all fitted to the same data set (denoted D), we can identify the best approximating model by calculating the AIC, BIC, etc.

$$AIC_i = -2\ell\text{n}L_i(D | \hat{\theta}) + 2p_i$$

$$BIC_i = -2\ell\text{n}L_i(D | \hat{\theta}) + p_i\ell\text{n}(n)$$

where p_i is the number of parameters in model i and n is the number of data points

Model Selection-II

All things equal, BIC will select less complex models than AIC. AIC_c takes small sample size effects into account:

$$AIC_{c,i} = -2\ell n L_i(D | \hat{\theta}) + 2p_i + \frac{2p_i(p_i + 1)}{n - p_i - 1}$$

I always use AIC_c but its effect is small for $n \gg p_i$

Model Selection-III

Models can be weighted with AIC and BIC, e.g.:

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

where $\Delta_i(AIC)$ is:

$$\Delta_i(AIC) = AIC_i - \min_j AIC_j$$

References

- Wagenmakers, E-J and S. Farrell. 2004. AIC model selection using Akaike weights. *Psychonomic Bulletin & Review* 11: 192-196.