Bayesian methods for ecological and environmental modelling
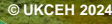—

Trainers:
Lindsay Banin, David Cameron,
Pete Henrys & Peter Levy

UK Centre for
Ecology & Hydrology

© UKCEH 2024

1

**Session 4a
Model selection & comparison**
—

Pete Henrys

UK Centre for
Ecology & Hydrology

2

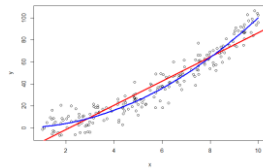### What we will cover in this session

How to assess models
  Important checks and diagnostics

How to compare different models
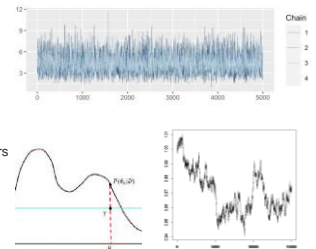 Appropriate metrics

Practical tools

3

Assessing Convergence

What is convergence?
Sampling has reached a point of stationarity

Why is it important?
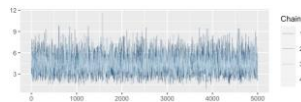Provide robust distributions of parameters that reflect true posterior

What can go wrong?
• Localised optima
• Non-identifiability

4

How to assess convergence

Visual inspection
Look at the trace plots for each parameter

The $\hat{R}$ statistic
Comparison of between chain variation to within chain variation.
If all chains converge to the same region and behave similarly, then the variance between the chains should be approximately equal to the average variance within chains and the estimated Rhat will be close to 1.

5

Other model checking

Effective posterior sample size (ESS)
• MCMC results are not independent.
• The lower the autocorrelation, the more independent pieces of information
• The approximate number of independent draws with the same estimation accuracy as our sample of correlated draws is referred to as the effective sample size.

Monte Carlo standard error (MCSE)
• Approximated by dividing the posterior standard deviation by the square root of the ESS. A low MCSE relative to the estimated posterior standard deviation is desirable
• A large MCSE implies that the sampling error variation masks the posterior standard deviation

6

Other model checking, continued…

Posterior predictive checking (PPC)

• Useful for assessing model fit to the data.
• Simulate data according to the fitted model and comparing the simulations to the observed data to look for important discrepancies.

• If the model fits the data well, we should be able to replicate important features of the observed data in the simulations.
• Simulations generated by the posterior distribution of the model parameters.
• Each MCMC draws from the posterior distribution = a new simulated dataset
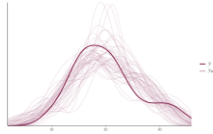
7

shinystan

Using widely available tools

8

Comparing models

• In frequentist modelling, you may have used AIC, BIC or DIC.
  • All of which are some form of penalised goodness of fit

• AIC and BIC are not generally considered fully-Bayesian approaches, as the uncertainty of the posterior distribution is not accounted for in their computations.

• Bayesian researchers have thus pursued relative model-data fit indices that account for the uncertainty associated with Bayesian methodology.

• WAIC is a more fully Bayesian approach for estimating the out-of-sample expectation

9

Cross validation

A family of techniques that try to estimate how well a model would predict previously unseen data by using fits of the model to a subset of the data to predict the rest of the data.

Cross-validation can be used to:
• Assess the predictive performance of a single model
• Assess model misspecification or calibration of the predictive distribution of a single model
• Compare multiple models
• Select a single model from multiple candidates
• Combine the predictions of multiple models
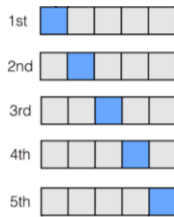
10

Cross validation

How is the data split
• Examples include: leave-one-out; leave-one-group-out;

The utility or loss
• Examples include: RMSE; R^2

The computational method used
• to compute leave-one-out predictive distributions, e.g. K-fold-CV

11

loo package

https://mc-stan.org/loo/

An approximation to Leave-One-Out (LOO) cross-validation

Utility
ELPD: Expected log pointwise predictive density for new observations.
Could observe this with infinite number of future observations.

Computation
PSIS: Pareto smoothed importance sampling leave-one-out cross-validation.

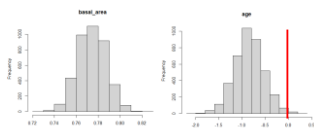Enables comparison of different models

12

Interpreting output

```
> m_multi <- stan_glm(tree_mass ~ basal_area +
+   age + cultivation + tree_height + timber_height,
+   data = df, refresh = 0)
```

Standard output

```
MCMC diagnostics
                mcse Rhat n_eff
(Intercept)     0.7  1.0  7444
basal_area      0.0  1.0  4282
age             0.0  1.0  4465
cultivation     0.0  1.0  3370
tree_height     0.0  1.0  3160
timber_height   0.0  1.0  3235
sigma           0.0  1.0  3765
mean_PPD        0.0  1.0  3782
```

```
Estimates:
                mean   sd   10%   50%   90%
(Intercept)    -1.8.5 13.8 -19.8.8 -1.8.5 -16.8.8
basal_area      0.8   0.0   0.8   0.8   0.8
age            -0.8   0.1  -1.7  -0.9  -0.4
cultivation     1.0   0.7   0.1   1.1   1.9
tree_height     1.9   2.1  -0.7   1.9   4.7
timber_height  20.2   2.3  17.3  20.1  21.1
sigma          74.1   1.2  72.6  74.1  75.6
```

Going further…..

basal_area   **age**

21 out 4000 samples
return a coefficient greater
than 0

13

Practical

- We will fit a series of models in 'rstanarm' together
- Use our diagnostics to assess the model fit
- Use cross validation to compare models and choose the most suitable

14

Recap

In this session we have covered:

- **Some Important checks and diagnostics for assessing models**
- **Ways of comparing different models**
- **Practical tools**

15

Any Questions?

16

This concludes
—
Model selection & comparison

Please be back & ready
to go at 11.30

UK Centre for
Ecology & Hydrology

Image © Pixaby CC01

17

Time for a break

Image © Pixaby CC01

18