Bayesian statistics with R
7. Contrast scientific hypotheses with model selection

Olivier Gimenez

November-December 2023

Model selection

How to select a best model?

- Is there any effect of rain or temperature or both on breeding success?

How to select a best model?

- Is there any effect of rain or temperature or both on breeding success?
- The proportion of explained variance R^2 is problematic, because the more variables you have, the bigger R^2 is.

How to select a best model?

- Is there any effect of rain or temperature or both on breeding success?
- The proportion of explained variance R^2 is problematic, because the more variables you have, the bigger R^2 is.
- Idea: **penalize models with too many parameters**.

Akaike information criterion (AIC)

$$AIC = -2\log(L(\hat{\theta}_1, \ldots, \hat{\theta}_K)) + 2K$$

with L the likelihood and K the number of parameters θ_i.

Akaike information criterion (AIC)

$$\text{AIC} = -2\log(L(\hat{\theta}_1, \ldots, \hat{\theta}_K)) + 2K$$

A measure of goodness-of-fit of the model to the data: the more parameters you have, the smaller the deviance is (or the bigger the likelihood is).

Akaike information criterion (AIC)

$$\text{AIC} = -2\log(L(\hat{\theta}_1, \ldots, \hat{\theta}_K)) + 2K$$

A penalty: twice the number of parameters K

Akaike information criterion (AIC)

- AIC makes the balance between *quality of fit* and *complexity* of a model.

Akaike information criterion (AIC)

- AIC makes the balance between *quality of fit* and *complexity* of a model.
- Best model is the one with lowest AIC value.

Akaike information criterion (AIC)

- AIC makes the balance between *quality of fit* and *complexity* of a model.
- Best model is the one with lowest AIC value.
- Two models are difficult to distinguish if $\Delta\text{AIC} < 2$.

Bayesian version

- Deviance Information Criteria or DIC, a Bayesian method for model comparison that JAGS can calculate for (m)any models.

$$\mathrm{DIC} = -2 \log L(\text{data} \mid \theta) + 2 p_{\mathrm{D}}$$

Bayesian version

- Deviance Information Criteria or DIC, a Bayesian method for model comparison that JAGS can calculate for (m)any models.

$$\mathrm{DIC} = -2 \log L(\text{data} \mid \theta) + 2p_{\mathrm{D}}$$

- where $-2 \log L(\text{data} \mid \theta)$ is the deviance evaluated at the posterior mean of the parameters (measures fit),

Bayesian version

- Deviance Information Criteria or DIC, a Bayesian method for model comparison that JAGS can calculate for (m)any models.

$$\mathrm{DIC} = -2 \log L(\text{data} \mid \theta) + 2 p_{\mathrm{D}}$$

- where $-2 \log L(\text{data} \mid \theta)$ is the deviance evaluated at the posterior mean of the parameters (measures fit),
- and p_{D} is the 'effective number of parameters (measures complexity); it is the posterior mean of the deviance minus the deviance of the posterior means.

Bayesian version

- Deviance Information Criteria or DIC, a Bayesian method for model comparison that JAGS can calculate for (m)any models.

$$\mathrm{DIC} = -2 \log L(\text{data} \mid \theta) + 2p_{\mathrm{D}}$$

- where $-2 \log L(\text{data} \mid \theta)$ is the deviance evaluated at the posterior mean of the parameters (measures fit),
- and p_{D} is the 'effective number of parameters (measures complexity); it is the posterior mean of the deviance minus the deviance of the posterior means.
- DIC is intended as a generalisation of AIC, and with little prior information, p_{D} should be approximately the true number of parameters.

Bayesian version

- Deviance Information Criteria or DIC, a Bayesian method for model comparison that JAGS can calculate for (m)any models.

$$\mathrm{DIC} = -2 \log L(\text{data} \mid \theta) + 2 p_{\mathrm{D}}$$

- where $-2 \log L(\text{data} \mid \theta)$ is the deviance evaluated at the posterior mean of the parameters (measures fit),
- and p_{D} is the 'effective number of parameters (measures complexity); it is the posterior mean of the deviance minus the deviance of the posterior means.
- DIC is intended as a generalisation of AIC, and with little prior information, p_{D} should be approximately the true number of parameters.
- The model with the smallest DIC is estimated to be the model that would best predict a replicate dataset with same structure as that observed.

DIC in Jags

```
storks
#> Inference for Bugs model at "code/logistic.txt", fit using jags,
#>  2 chains, each with 2000 iterations (first 1000 discarded)
#>  n.sims = 2000 iterations saved
#>          mu.vect sd.vect    2.5%     25%     50%     75%   97.5%  Rhat n.eff
#> a          1.550   0.085   1.431   1.515   1.553   1.594   1.667 1.161  2000
#> b.rain    -0.148   0.062  -0.273  -0.188  -0.147  -0.106  -0.038 1.005  1500
#> b.temp     0.028   0.064  -0.102  -0.014   0.031   0.071   0.147 1.065    30
#> deviance 206.492  29.698 201.809 202.798 203.991 205.732 212.404 1.083  2000
#>
#> For each parameter, n.eff is a crude measure of effective sample size,
#> and Rhat is the potential scale reduction factor (at convergence, Rhat=1).
#>
#> DIC info (using the rule, pD = var(deviance)/2)
#> pD = 441.0 and DIC = 647.4
#> DIC is an estimate of expected predictive error (lower deviance is better).
```

Further reading

- Hooten, M.B. and Hobbs, N.T. (2015), A guide to Bayesian model selection for ecologists. Ecological Monographs, 85: 3-28. https://doi.org/10.1890/14-0661.1
- Conn, P.B., Johnson, D.S., Williams, P.J., Melin, S.R. and Hooten, M.B. (2018), A guide to Bayesian model checking for ecologists. Ecol Monogr, 88: 526-542. https://doi.org/10.1002/ecm.1314