Bayesian statistics with R

7. Contrast scientific hypotheses with model selection

Olivier Gimenez

March 2021

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 $\Delta AIC < 2$.

Bayesian version

- Watanabe-Akaike Information Criteria or WAIC:

$$\mathrm{WAIC} = -2\sum_{i=1}^{n} \log E[p(y_i \mid \theta)] + 2p_{\mathsf{WAIC}}$$

- where $E[p(y_i \mid \theta)]$ is the posterior mean of the likelihood of the ith observation and
- p_{WAIC} is the effective number of parameters computed using the posterior variance of the likelihood.
- Relatively new and not yet available in Jags in routine.

WAIC in Jags

```
# calculate wAIC with JAGS
# https://sourceforge.net/p/mcmc-jags/discussion/610036/thread/8211df61/#e
samples <- jags.samples(storks$model,c("WAIC","deviance"), type = "mean",
                        n.iter = 2000,
                        n.burnin = 1000,
                        n.thin = 1)
```

```
samples$p_waic <- samples$WAIC
samples$waic <- samples$deviance + samples$p_waic
tmp <- sapply(samples, sum)
waic <- round(c(waic = tmp[["waic"]], p_waic = tmp[["p_waic"]]),1)
waic
#>    waic p_waic
#>   218.0   13.3
```

Your turn: Practical 7