STAT40380/STAT40390/STAT40850 Bayesian Analysis

Dr Niamh Russell

School of Mathematics and Statistics
University College Dublin

niamh.russell@ucd.ie

April 2016

Model checking and improvement

- The first two steps in a Bayesian analysis are:
  1. Choose the likelihood(s) and prior(s)
  2. Fit the model (using MCMC or another method)
- Once we have a model or models, we then want to:
  - Choose between them (relative model checks)
  - Determine which models are fit for purpose
- There will very rarely be just a single model we want to fit
- Remember that there is usually no such thing as a *true* model. We simply want to find a good model that we can use for whatever purpose it was built.

Example: Tumours in rats

Example

Suggest different likelihoods and prior models for the rat tumours data. Think about why these different methods might be used.

0/20	0/20	0/20	0/20	0/20	0/20	0/20	0/19	0/19	0/19
0/19	0/18	0/18	0/17	1/20	1/20	1/20	1/20	1/19	1/19
1/18	1/18	2/25	2/24	2/23	2/20	2/20	2/20	2/20	2/20
2/20	1/10	5/49	2/19	5/46	3/27	2/17	7/49	7/47	3/20
3/20	2/13	9/48	10/50	4/20	4/20	4/20	4/20	4/20	4/20
4/20	10/48	4/19	4/19	4/19	5/22	11/46	12/49	5/20	5/20
6/23	5/19	6/22	6/20	6/20	6/20	16/52	15/47	15/46	9/24

Relative model checks

- We usually want to test the effect on the posterior of *changing the likelihood or the prior distributions*
- (Remember that the likelihood is another subjective choice in our modelling assumptions just like the prior)
- In the case of informative prior distributions, we may want to check whether these prior assumptions were correct by contrasting the model with that fitted with an uninformative prior
- We may also want to *compare models of differing size and complexity* (eg using more explanatory variables) to see whether the improvement in fit for a more complex model is large enough to justify the extra computation and understanding

Relative model checks 2

- We have already the Bayes Factor which measures the relative probabilities of observing the data given the model is correct
- In many cases (eg with improper priors) the Bayes Factor is hard if not impossible to compute
- If we accept that model is correct, then we may be better off by simply calculating the distance between the data and the model
- Define the
- Deviance to be:

$$D(\theta) = -2 \log p(\mathbf{x}|\theta)$$

- A (relatively) low deviance means the model fits well. However, a more complex model will almost always have a lower deviance

Measuring model complexity

- After fitting the model, we have posterior samples of $\theta|\mathbf{x}$
- But which values of θ should we use to compute D? We have 2 options:
  1. Compute the deviance for some 'best' values for the parameters, eg the mean $\bar{\theta}$. WinBUGS calls this `Dhat`$= D(\hat{\theta}) = -2 \log p(\mathbf{x}|\hat{\theta})$
  2. Average the deviance over the parameter values. WinBUGS calls this `Dbar` $= \hat{D}(\theta) = \frac{1}{L}\sum_{i=1}^{L} D(\theta^i)$ where θ^i are our posterior sample of L final iterations
- `Dhat` is almost always smaller than `Dbar` because the best values of θ tend to give the best fitting model
- The quantity

$$p_D = \hat{D}(\theta) - D(\hat{\theta}) \tag{1}$$

(equivalently `Dbar-Dhat`) is the *effective number of parameters*

The effective number of parameters p_D

- In a normal, linear, fixed effects model, p_D will be the number of parameters in the model
- More generally, p_D can be thought of as *the number of unconstrained parameters in the model*. A parameter will count as 1 if it is estimated with no constraints or prior information; 0 if it is fully constrained or if all the information comes from the prior distribution and the data have some effect on its value
- Example: $x \sim N(\theta|1)$ with $\theta \sim (0, \infty)$. If an observation x is close to zero then the effective number of parameters is approximately 1/2. If the value of x is large then p_D is approximately 1.
- p_D is thus a good measure of *model complexity*. We can use it to penalise the deviance so as to create *parsimonious models*

The Deviance Information Criterion

- Define the *Deviance Information Criterion* as:

$$\begin{aligned} DIC &= \hat{D}(\theta) + p_D \\ &= D(\bar{\theta}) + 2p_D \\ &= 2\hat{D}(\theta) - D(\bar{\theta}) \end{aligned}$$

- ie DIC = 'goodness of fit' + 'complexity'. Lower DIC means a 'better' model
- Very simple and quick to compute, no need for Bayes Factor-like marginal likelihoods
- Monitored in WinBUGS through Inference > Samples menu
- Warning: p_D not invariant to re-parametrisation, and can also be negative!

Relationship to other model performance criteria

- Deviance Information Criterion:

$$DIC = D(\bar{\theta}) + 2p_D$$

- Akaike Information Criterion (where ψ are hyperparameters, and p_ψ # of hyperparameters):

$$AIC = -2 \log p(\mathbf{x}|\hat{\psi}) + 2p_\psi$$

- Bayesian Information Criterion

$$BIC = -2 \log p(\mathbf{x}|\hat{\psi}) + p_\psi \log n$$

- Each is useful for different purposes. *DIC* best when top level parameters fixed, *AIC* when hyperparameters fixed, *BIC* as an approximation to the Bayes Factor, ie no interest in parameters at all

Example 1: DIC

Example

Suppose data are available as $y_i \sim Po(\lambda)$ for $i = 1, \ldots, 10$ with prior $\log(\lambda) \sim N(0, 100)$. We observe $\sum y_i = 22$ and $\sum y_i! = 17.877$. We fit the model using WinBUGS and find that $\bar{\lambda} = 2.2$ and $\bar{D}(\lambda) = 46.08$. Find the DIC and p_D.

Example 2: DIC

Example

A more complicated model is proposed for the data in Example 1: $y_i \sim Po(\lambda_i)$ for $i = 1, \ldots, 10$ with prior $\log(\lambda_i) \sim N(0, 100)$. After fitting the model using WinBUGS we find that $\sum \bar{\lambda}_i = 22.16, \sum(y_i \log \bar{\lambda}_i) = 29.89$ and $\bar{D}(\lambda) = 27.78$. Find the DIC and p_D. Is this model to be preferred over that fitted in the first example?

Absolute model checks

- Once we have settled on some 'best' models we might like to see if they individually meet the assumptions we have made and check that they are fit for prediction. This is *just as important* as the relative model checking
- We can still use traditional methods such as residual checks to see if distributional assumptions are reasonable
- We can create predictive distributions for new data given the old:

$$p(\mathbf{x}^*|\mathbf{x}) = \int p(\mathbf{x}^*|\theta)p(\theta|\mathbf{x})d\theta$$

  (This is very easy to do in WinBUGS)
- These could be run to compare with existing data, or as predictions for new data not yet collected
- Alternatively we can run cross-validations, where we leave out small amounts of the data, re-fit the model and compare our new predictions with the missing values