ST 740: Goodness of Fit

Alyson Wilson

Department of Statistics
North Carolina State University

November 11, 2013

Key Questions

1. How can I tell if my model is providing an adequate fit to the data? (*goodness of fit*)
2. What aspects of reality are not captured by my model? Are these important given the use I would like to make of the model? (*model checking*)
3. How can I tell if any of the modeling choices I have made are having an undue impact on my results? (*sensitivity analysis*)
4. Which model (or models) should I ultimately choose for the final presentation of my results? (*model selection*)

Penalized Likelihood Criteria

- We want to compare several models.
- Let p denote the number of parameters in the model and n the number of data points.
- Define the *deviance* as

$$D(y, \theta) = -2 \log f(y \mid \theta).$$

- Also define

$$D_{\hat{\theta}}(y) = D(y, \hat{\theta}(y))$$

as the deviance evaluated at $\hat{\theta}$ (usually the posterior mean).

Penalized Likelihood Criteria

The Akaike Information Criterion (AIC) $= D_{\hat{\theta}}(y) + 2p$. The AIC is used to choose models that have good "out-of-sample" predictive capabilities.

It can be hard to figure out p and n. For example,

- Suppose that we have longitudinal data, where we have s_i observations on patient i, where $i = 1, \ldots, m$. Is $n = \sum_{i=1}^{m} s_i$ or is $n = m$? If the observations on each patient are independent, then the former seems most appropriate; if they are perfectly correlated within each patient, then the latter. (Of course, the truth is somewhere in between.)
- Suppose that we have a collection of m parameters and a hierarchical model.

Penalized Likelihood Criteria

Define the posterior mean deviance as:

$$\begin{aligned} D_{avg}(y) &= E_{\theta \mid y}(D(y, \theta) \mid y) \\ &= \int D(y, \theta)\pi(\theta \mid y) d\theta \\ &= \int -2 \log f(y \mid \theta)\pi(\theta \mid y) d\theta \\ \hat{D}_{avg}(y) &\approx \frac{1}{M} \sum_j -2 \log(f(y \mid \theta^{(j)})) \end{aligned}$$

The difference between the posterior mean deviance and $D_{\hat{\theta}}(y)$ represents the effect of model fitting and has been used as a measure of the *effective number of parameters* in a Bayesian model.

The effective number of parameters is

$$p_D = \hat{D}_{\text{avg}}(y) - D_{\hat{\theta}}(y)$$

Penalized Likelihood Criteria

p_D can be thought of as the number of "unconstrained" parameters in the model, where a parameter counts as: 1 if it is estimated with no constraints or prior information; 0 if it is fully constrained or if all the information about the parameter comes from the prior distribution; or an intermediate value if both the data and prior distributions are informative.

Penalized Likelihood Criteria

We then use the *deviance information criterion* (DIC), which is a generalization of AIC. Define

$$\begin{aligned} \text{DIC} &= D_{\hat{\theta}}(y) + 2p_D \\ &= \hat{D}_{avg}(y) + p_D \end{aligned}$$

Lower values of the penalized likelihood criteria are better. These criteria do not have an absolute scale and should be used only to rank models.

DIC

- For non-hierarchical models, p_D should be approximately the true number of parameters.
- In models with negligible prior information, DIC will be approximately equivalent to AIC.
- Different values of $\hat{D}_{avg}(y)$ (and hence p_D and DIC) can be obtained depending on the parameterization used for the prior distribution.

DIC

- The minimum DIC estimates the “best” model in the same spirit as Akaike’s criterion. However, if the difference in DIC is, say, less than 5, and the models make very different inferences, then it could be misleading just to report the model with the lowest DIC.
- DICs are comparable only over models with exactly the same observed data, but there is no need for them to be nested.