

# Lecture IX: Model Comparison

Charles B. Moss

May 16, 2013

- I. Typically we have several possible models to characterize random variables. The question is then what do Bayesian methods bring to the table in making choices between the alternatives.
  - A. First, assume that we have  $J$  alternative models  $j = 1, 2, \dots, J$ . Each model  $j$  has a vector of unobservable variables  $\theta_{A_j}$ . In addition, each model has a prior density function  $p(\theta_{A_j} | A_j)$ , an observables density function  $p(y, \theta_{A_j}; A_j)$  and density function for the variables of interest  $p(\omega | y, \theta_{A_j}, A_j)$ .
    1. These  $J$  models are linked by their predictions on the common set of observable variables  $y$  and variables of interest  $\omega$ .
    2. In this formulation we have a collection of priors for the  $J$  models  $p(A_j)$  such that  $\sum_{j=1}^J p(A_j) = 1$ .
    3. "There is no essential conceptual difference between model and prior – we could just as well regard the entire collection as a single model, with  $\{p(A_j), p(\theta_{A_j} | A_j), d\nu_j(\theta_{A_j})\}_{j=1}^J$  providing the prior distributions of unobservables.
  - B. The density for the marginal variable of interest  $p(\omega | y^0)$  can be defined for the combined model as

$$p(\omega | y^0) = \sum_{j=1}^J p(\omega | y^0, A_j) p(A_j | y^0) \quad (1)$$

This formulation is known as *model averaging*.

1. In this model

$$p(A_j | y^0) = \frac{p(y^0 | A_j) p(A_j)}{p(y^0)} \propto p(y^0 | A_j) p(A_j). \quad (2)$$

2. **Definition 2.6.1** In favor of the Model  $A_j$  versus the model  $A_k$ , the *prior odds ratio* is  $P(A_j)/P(A_k)$ ; the *Bayes factor* is  $p(y^0 | A_j)/p(y^0 | A_k)$ ; and the *posterior odds ratio* is  $P(A_j | y^0)/P(A_k | y^0)$ .
3. The basic idea is that we have two different models  $A_j$  and  $A_k$ . We can compare the "acceptableness" of each model by comparing

$$\frac{P(A_j | y^0)}{P(A_k | y^0)} = \frac{P(A_j)}{P(A_k)} \cdot \frac{p(y^0 | A_j)}{p(y^0 | A_k)} \quad (3)$$

4. **Definition 2.6.2** In choosing between two models, the ratio of the loss functions  $L(A_1, A_2)/L(A_2, A_1)$  is the *Bayes critical values*.

$$\frac{P(A_1 | y^0)}{P(A_2 | y^0)} = \frac{P(A_1) p(y^0 | A_1)}{P(A_2) p(y^0 | A_2)} > \frac{L(A_1, A_2)}{L(A_2, A_1)} \quad (4)$$

## II. Predictive Densities

- A. "The marginal likelihood of model  $A_j$  ( $p(Y_T^0 | A_j)$ ) is the measure of how well model  $A_j$  predicted the data  $Y_T^0$  that are relevant for the comparison of model  $j$  with other models."
- B. **Definition 2.6.3** The *predictive density* of  $y_{T+1}, \dots, y_F$  conditional on  $Y_T^0$  and model  $A$  is

$$p(y_{T+1}, \dots, y_F | Y_T^0, A). \quad (5)$$

"The predictive density is relevant after the formulation of model  $A$  and observing  $Y_T = Y_T^0$ , but before observing  $y_{T+1}, \dots, y_F$ ."

- C. **Definition 2.6.4** The *predictive likelihood* of  $y_{T+1}^0, \dots, y_F^0$  conditional on  $Y_T^0$  and the model  $A$  is the real number  $p(y_{T+1}^0, \dots, y_F^0 | Y_T^0, A)$ .
  1. Think about this as a sequence – I have observed  $y_1^0, \dots, y_T^0$  and formulated my unobserved parameters based on these observations (i.e., minimizing the loss function).

2. Given this prior I can then evaluate the empirical likelihood function for  $y_{T+1}^0, \dots, y_F^0$ .

D. **Definition 2.6.5** the *predictive Bayes factor* in favor of model  $A_j$ , versus model  $A_k$  is

$$\frac{p(y_{T+1}^0, \dots, y_T^0 | Y_T^0, A_j)}{p(y_{T+1}^0, \dots, y_F^0 | Y_T^0, A_k)} \quad (6)$$

1. Again, assume that I have two extreme cases – a Bayesian estimation of a yield function under a Gamma distribution function and a yield function under normality.
2. In each case, I observe  $y_1, \dots, y_T$ . Using these estimates I can construct estimates of each set of parameters.
3. Next, I observe a second set of data  $y_{T+1}^0, \dots, y_F^0$ . I can compare the two models by examining the likelihood ratio in Equation 6.

E. **Theorem 2.6.1 Representation of Predicted Likelihood**

The predictive likelihood is a ratio of marginal likelihoods:

$$p(y_{T+1}^0, \dots, y_F^0 | Y_T^0, A) = \frac{p(Y_F^0 | A)}{p(Y_T^0 | A)} \quad (7)$$

F. **Corollary 2.6.1 Decomposition of the Marginal Likelihood**

For any strictly increasing sequence of integers  $\{s_j\}_{j=0}^q$  with  $s_0 = 0$  and  $s_q = T$ , the marginal likelihood can be decomposed

$$p(Y_T^0 | A) = \prod_{\tau=1}^q p(y_{s_{\tau-1}+1}^0, \dots, y_{s_\tau}^0 | Y_{s_{\tau-1}}^0, A). \quad (8)$$

”This result immediately implies that the Bayes factor in favor of model  $A_j$  versus model  $A_k$  can be decomposed in terms of the predictive Bayes factors”

$$\frac{p(Y_T^0 | A_j)}{p(Y_T^0 | A_k)} = \prod_{\tau=1}^q \left[ \frac{p(y_{s_{\tau-1}+1}^0, \dots, y_{s_\tau}^0 | Y_{s_{\tau-1}}^0, A_j)}{p(y_{s_{\tau-1}+1}^0, \dots, y_{s_\tau}^0 | Y_{s_{\tau-1}}^0, A_k)} \right] \quad (9)$$