Bayesian Model Comparison and Bayesian Model Averaging

June 14, 2011

The Plan

Today:

- Model comparison
- Model uncertainty
- Marginal Likelihood
- Bayesian Model Averaging

Wright (2003) and Hoetin, Madigan, Raftery and Volinsky (1999).

We will also have a quick review of course

Model uncertainty

Bayesian statistics lets us consider parameter uncertainty

- Remember: *Parameter uncertainty* is a statement about our knowledge

But we also do not know the "true" model.

- Model comparison: Which model is more likely to have generated the data and how much more likely?
- Model averaging: We do not have to choose between models
  - Weight prediction from different model based on the relative fit as measured by the marginal likelihood

For both model comparison and model averaging we need to compute the marginal likelihood

- Marginal likelihood: How likely is a model to have generated the data?
- Measure naturally "punish" more densely parameterized models

The marginal likelihood

Consider a model M_1. The marginal likelihood of model M_1 is simply

$$p(Z \mid M_1) = \int_{\Theta} p\left(Z \mid \theta, M_1\right) p\left(\theta \mid M_1\right) d\theta$$

i.e. we integrate the likelihood over the parameter space to compute the weighted "average" probability that the model has generated the data.

- Punishes models with many parameters. But why?

The prior probability of many vs few parameter

Consider two models that fit the data equally well so that $p(Z \mid \theta, M_1) = p(Z \mid \theta, M_2)$

- Model 1 has 1 parameter and model 2 has two parameters. Let's say prior probabilities are $U(0,5)$ for all parameters
- $p(\theta \mid M_1) = p(\theta_1 \mid M_1)$
- $p(\theta \mid M_2) = p(\theta_1 \mid M_2)p(\theta_2 \mid M_2)$

Since all priors are $U(0,5)$m we have that $p(\theta_1 \mid M_1) = p(\theta_1 \mid M_2)$ and since all probabilities are in the interval $(0,1)$ we must have that $p(\theta \mid M_1) > p(\theta \mid M_2)$ so that that $p(Z \mid M_1) > p(Z \mid M_2)$.

Computing marginal likelihoods

Simulate J draws from the posterior density of θ using the M-H algorithm. Let each draw in the Markov chain be indexed by j.

1. Draw S integers on a uniform distribution between 1 and J
2. Compute
$$p\,(Z \mid \theta_j, M_1)\,p\,(\theta_j \mid M_1)\,d\theta$$
for each j that was drawn in step 1.
3. The estimate of the marginal likelihood is then given by
$$p(Z \mid M_1) = \frac{1}{S}\sum_j p\,(Z \mid \theta_j, M_1)\,p\,(\theta_j \mid M_1)\,d\theta$$

Model comparison

One way to compare models is to compute the posterior odds ratio

$$\frac{p(Z \mid M_1)}{p(Z \mid M_2)}$$

which gives the relative probability of model M_1 having generated the data Z as compared to model M_2 .

Bayesian Model Averaging

Sometimes there are many plausible models that all fit the data reasonably well. Do we have to choose between them?

- No, we can use what is called Bayesian Model Averaging to assign weights to each model.

Example: Forecasting using multiple models.
Let the forecast of generated by model M_i be denoted as $E\left(y_{t+h} \mid M_i\right)$. The BMA weighted forecast is then given by

$$\frac{1}{n}\sum_j p(M_j \mid Z)E\left(y_{t+h} \mid M_j\right)$$

where the probability of model i conditional on the data is given by

$$p(M_i \mid Z) = \frac{p(Z \mid M_i)p\left(M_i\right)}{\sum_j^n p(Z \mid M_j)p\left(M_j\right)}$$

That's it for the course.

Now it's time for review.