Bayesian model comparison (continued)

STAT 206
4 Mar 21

Lecture

Let M be either (a) a frequentist model (with no prior layer in the hierarchy) or (b) the sampling model [SM] part of a Bayesian model (ignoring the prior). (frequentist)

The deviance of M, as defined in the 3 Mar 21 discussion section notes, is

$$\text{deviance}_F(M \mid \mathbb{D}\, \mathcal{B}) = -2 \cdot \ell\ell(\tilde{\theta}_{MLE} \mid \mathbb{D}\, [SM]\, \mathcal{B}),$$

where $\tilde{\theta}$ is the model's parameter vector and $\mathbb{D}$ is the dataset, in which the log likelihood function is defined by taking the log of {the likelihood function obtained directly from the SM, including the SM's normalizing constant}.

Notice that $\text{deviance}_F(M \mid \mathbb{D}\, \mathcal{B})$ is a single real number

The Bayesian way of looking at the deviance is different: $\mathbb{D}_B(\underset{\sim}{\theta} \mid M\,\mathbb{D}\,B) = -2 \cdot ll(\underset{\sim}{\theta} \mid \mathbb{D}\,[M]\,B)$ (the usual likelihood function)

In other words, deterministic $\mathbb{D}_B$ is a function of the parameter vector $\underset{\sim}{\theta}$

($\mathbb{D}_F$ is just the maximum value attained by $\mathbb{D}_B$ over the parameters) This means that

$\mathbb{D}_B(\underset{\sim}{\theta} \mid M\,\mathbb{D}\,B)$ has a (marginal) posterior distribution of its own (!), which we can monitor as a derived quantity in our (MC)MC (!)

simple example $(Y_i \mid \theta\, B) \overset{IID}{\sim} \text{Bernoulli}(\theta)$, $(i = 1, \ldots, n)$

$\theta \in (0,1)$ $\quad \underset{\sim}{y} = (y_1, \ldots, y_n)$

$p(\underset{\sim}{y} \mid \theta\, B) = \theta^s (1-\theta)^{n-s}$ (note: in this case the normalizing constant is 1)

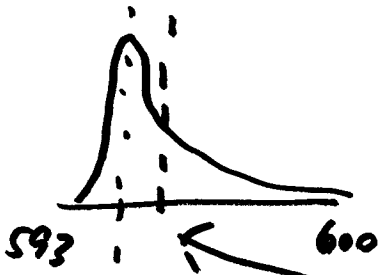
$\ell(\theta \mid \underset{\sim}{y}\, \theta\, B) = \theta^s (1-\theta)^{n-s}$

$$ll(\theta \mid x, B) = s \log\theta + (n-s)\log(1-\theta)$$

$$D_B(\theta \mid M, B) = -2\, ll(\theta \mid x, B)$$

$$= -2\left[s \log\theta + (n-s)\log(1-\theta)\right]$$

(R code)



posterior distribution of D_B

posterior mean

Frequentist deviance (posterior mode)

The Deviance Information Criterion (DIC) was developed by David Spiegelhalter & colleagues in the early 2000s & published in 2002 by ~~including Nicky Best &~~ Angelika Van der Linde

As described in the document camera notes from the morning discussion section on 3 Nov 21, the deviance itself

uses the entire dataset to evaluate the model's quality, which encourages overfitting, so it can't serve by itself as a good model comparison tool; it needs to be penalized by a term that discourages overly-complex models.

So a good model comparison method looks like this: $f(\text{model fit}, \text{model complexity})$

(model fit: can use the deviance for this; what should we use for this?)

The simplest such $f(\cdot,\cdot)$ would be additive: $f(\text{model fit}, \text{model complexity})$

(want small)

$= \text{Deviance} + (\text{model complexity})$, but

this needs measuring (model complexity) on the same scale as the deviance,

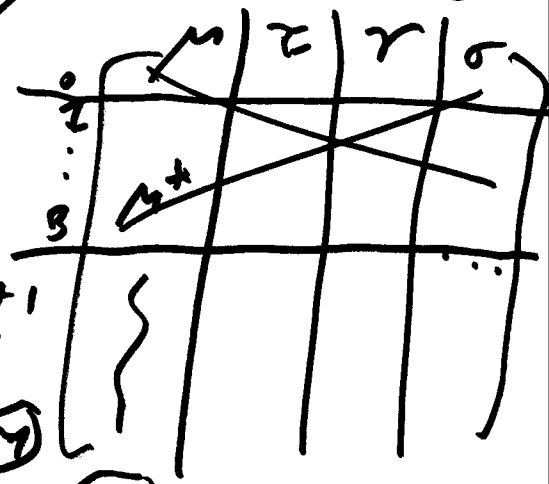
to get apples to apples

how should this be accomplished?

it turns out that Laplace approximations can help.

First, a note on MCSE values in MCMC

If a column in the MCMC data set behaves like an AR(1), then



$$\widehat{MCSE}(\bar{\mu}^*) = \frac{s_{\mu^*}}{\sqrt{M}} \sqrt{\frac{1+\hat{\rho}}{1-\hat{\rho}}}$$

IID MCSE

MC est. of post. mean for μ

as $\hat{\rho} \uparrow +1$, MCSE $\to +\infty$

the MCMC algorithm works by finding a Markov chain whose equilibrium (stationary) distribution is $p(\underset{\sim}{\theta} | D -)$

Markov chain

$$\theta^*_{B+2} = f(\theta^*_{B+1})$$

$B+1$: $B+M$ [θ^*_{B+1} θ^*_{B+2} ⋮]

SD 0.47

$p(\mu | D -)$

404.3 ± 0.0018

NBIO case study

Using the cheating approach, we identified 2 interesting models:

- Gaussian (M_1)
- t (M_2)

$A_1 \rightarrow q_1$ | If M_2 better than M_1?

$A_1 \rightarrow q_2$ | Better for what purpose?

Pursuing A_2

turns model comparison into a decision problem (Bayesian decision theory)

utility

	θ	$\mathcal{D}(\theta \mid \text{---})$
1	θ_1^*	$-2[s \log \theta_1^* + (b-s) \log (1-\theta_1^*)$
2	θ_2^*	
⋮	⋮	
M	θ_M^*	

M_C