Bayesian Model Selection and Extrasolar Planet Detection

E.B. Ford

Discussion by Jayanta Ghosh
Purdue, I.S.I.

This is a lovely description of how Bayesian Model Selection helps Extrasolar Planet Detection.

Data available on
(1) Amplitude of jitter
(2) Velocity offset
(3) Orbital period
(4) Velocity semi-amplitude
(5) Orbital eccentricity
(6) Orbital phase + (7) One other variabl[e]

The catch is that even for moderately high dimensional $\underset{\sim}{\theta}$ (as here) the integrals are not easy to calculate.

This is surprising since

(1) $p(\underset{\sim}{\theta} \mid \text{data}, M_j)$ is "easy" to sample from by MCMC

(2) $p(\underset{\sim}{\theta} \mid M_j)$, $L(\underset{\sim}{\theta} \mid M_j)$ are usually easy to calculate and

(3) $$I = \frac{p(\underset{\sim}{\theta} \mid M_j) \cdot L(\underset{\sim}{\theta} \mid M_j)}{p(\underset{\sim}{\theta} \mid \text{data}, M_j)}.$$

If we could calculate or estimate the posterior density (not just sample it) we could calculate or estimate I.

Suppose there are two models M_1, M_2 with M_j postulating n_j planets.

Each model specifies a likelihood

$$L(\underset{\sim}{\theta} \mid M_j) = \text{p.d.f. of data}$$
$$= p(\text{data} \mid \underset{\sim}{\theta}, M_j)$$

The likelihood is assumed to be normal (eqn (1), p 2).

The priors for diffferent variables are given in Table 1, with a good discussion of reasons for choice.

The marginal likelihood of data given M_j is

$$I = p(\text{data} \mid M_j) = \int_{\Theta} L(\underset{\sim}{\theta} \mid M_j)\, p(\underset{\sim}{\theta} \mid M_j)\, d\underset{\sim}{\theta}$$

Bayes Factor $= p(\text{data} \mid M_2) / p(\text{data} \mid M_1)$

An estimate of I due to Raftery is (dropping M_j as index)

$$\left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{1}{p(\underset{\sim}{\theta}_i) L(\underset{\sim}{\theta}_i)} \right\}^{-1} = \text{a Harmonic Mean}$$

where $\underset{\sim}{\theta}_i$'s are a sample from the posterior.

It doesn't work because the mean square

$$\int \left\{ \frac{1}{p(\underset{\sim}{\theta}) L(\underset{\sim}{\theta})} \right\}^2 p(\underset{\sim}{\theta}) L(\underset{\sim}{\theta}) d\underset{\sim}{\theta}$$

is typically infinite.

What about the new work of Raftery and Newton (Valencia 8)?

The paper discusses various other methods, e.g., ~~using~~ Laplace approximation, and a modification of Raftery's method due to Gelfand & Dey. Graphs and theoretical arguments show none of them perform well.

Two methods, Defensive Importance sampling and a new method of Berger — "Crazy Importance Sampling" — perform well.

Berger's method is based on the identity

$$\underset{\text{def}}{I =} \int p(\underset{\sim}{\theta}) L(\underset{\sim}{\theta}) \, d\underset{\sim}{\theta} = \frac{\int p(\underset{\sim}{\theta}) L(\underset{\sim}{\theta}) h(\underset{\sim}{\theta}) \, d\underset{\sim}{\theta}}{\int h(\underset{\sim}{\theta}) \, p(\underset{\sim}{\theta} \mid \text{data}) \, d\underset{\sim}{\theta}}$$

The trick is to sample $\underset{\sim}{\tilde{\theta}}_i$'s from $h(\underset{\sim}{\theta})$ for the numerator and $\underset{\sim}{\bar{\theta}}_i$'s from posterior for the denominator. The estimate is

$$\frac{\sum p(\underset{\sim}{\tilde{\theta}}_i) L(\underset{\sim}{\tilde{\theta}}_i) / n'}{\sum p(\underset{\sim}{\bar{\theta}}_i) L(\underset{\sim}{\bar{\theta}}_i) / n}$$

(two imp samples)

I end this discussion with three new, possibly naive, methods.

1. The paper has a method of tracking multiple modes of $p(\underset{\sim}{\theta})L(\underset{\sim}{\theta})$. Given the location of the modes, it is possible to calculate Laplace Approximation around each mode and sum.

2. Choose a suitable constant c, which should be like the midpoint of the range of $p(\underset{\sim}{\theta})L(\underset{\sim}{\theta})$.

Let

$$I = \int_{p(\underset{\sim}{\theta})L(\underset{\sim}{\theta}) \le c} p(\underset{\sim}{\theta})L(\underset{\sim}{\theta})\,d\underset{\sim}{\theta} + \int_{p(\underset{\sim}{\theta})L(\underset{\sim}{\theta}) > c} p(\underset{\sim}{\theta})L(\underset{\sim}{\theta})\,d\underset{\sim}{\theta} = I_1 + I_2$$

Estimate I_1 by Berger's and I_2 by Raftery's method.

Note that the problem of infinite variance of R's method will not arise here

3. Let c be as in (2). Estimate 8 I_2 as in (2), i.e, by Raftery's method.

Now estimate

$$I_3 = \int_{p(\underset{\sim}{\theta})L(\underset{\sim}{\theta}) \le c} p(\underset{\sim}{\theta} \mid \text{data})\, d\underset{\sim}{\theta}$$

by $\frac{1}{n} \sum \text{Indicator}\{p(\underset{\sim}{\theta}_i)L(\underset{\sim}{\theta}_i) \le c\}$ where $\underset{\sim}{\theta}_i$'s are samples from the posterior by MCMC.

Now estimate I from the fact

$$I_3 = 1 - \frac{I_2}{I}$$

This is reminiscent of a method due to Chib.