ADVANCED BAYESIAN MODELING

Bayes Factors

Appealing idea for comparing models:

1. Assign models prior probabilities.
2. Use Bayes' rule to compute posterior probabilities.
3. Choose model(s) with highest posterior probability.

Can we compare in a way that isn't too dependent on prior probabilities?

Want a pure measure of strength of evidence from the data.

Two Models

For data y, consider two different models: H_1 and H_2

Assign them prior probabilities

$$p(H_1) > 0 \qquad\qquad p(H_2) > 0$$

Bayes' rule formally gives posterior probabilities

$$p(H_1 \mid y) = \frac{p(H_1)\, p(y \mid H_1)}{c} \qquad\qquad p(H_2 \mid y) = \frac{p(H_2)\, p(y \mid H_2)}{c}$$

where normalizing factor c is the same in both cases (depends only on y).

Follows that

$$\frac{p(H_2 \mid y)}{p(H_1 \mid y)} = \frac{p(H_2)}{p(H_1)} \times \frac{p(y \mid H_2)}{p(y \mid H_1)}$$

The final factor is the **Bayes factor in favor of** H_2 **versus** H_1:

$$\mathrm{BF}(H_2; H_1) = \frac{p(y \mid H_2)}{p(y \mid H_1)}$$

Also,

$$\frac{p(H_2)}{p(H_1)} = \textbf{prior odds} \qquad\qquad \frac{p(H_2 \mid y)}{p(H_1 \mid y)} = \textbf{posterior odds}$$

in favor of H_2.

So

$$\mathrm{BF}(H_2; H_1) \;=\; \frac{\text{posterior odds favoring } H_2}{\text{prior odds favoring } H_2}$$

Represents how much the odds of H_2 relative to H_1 change after seeing data y:

- $\mathrm{BF}(H_2; H_1) \approx 1$ indicates data do not distinguish well between them.
- $\mathrm{BF}(H_2; H_1) \gg 1$ indicates strong support for H_2 over H_1.

Bayes factor represents strength of evidence coming from the data.

A possible interpretive scale (Kass & Raftery, 1995):

$\mathrm{BF}(H_2; H_1)$	data evidence for H_2 vs. H_1
1 to 3	Barely mentionable
3 to 20	Positive
20 to 150	Strong
> 150	Very Strong

Simple Case: No Parameters

Suppose H_1 and H_2 each *fully* specify a distribution for y, without needing any parameters (or priors).

Then $p(y \mid H_m)$ is the *likelihood* ($m = 1, 2$), and

$$\mathrm{BF}(H_2; H_1) \;=\; \frac{p(y \mid H_2)}{p(y \mid H_1)} \;=\; \textit{likelihood ratio}$$

This is a classical (non-Bayesian) measure of evidence in favor of H_2 over H_1.

Example

$$H_2 \quad = \quad \text{you have rare disease} \qquad\qquad H_1 \quad = \quad \text{you don't}$$

$$y \quad = \quad 1 \text{ if you test positive, } 0 \text{ if not}$$

Supposing

$$p(y \mid H_2) \quad = \quad \begin{cases} 0.99, & y = 1 \\ 0.01, & y = 0 \end{cases} \qquad\qquad p(y \mid H_1) \quad = \quad \begin{cases} 0.05, & y = 1 \\ 0.95, & y = 0 \end{cases}$$

then, if you test positive ($y = 1$)

$$\mathrm{BF}(H_2; H_1) \quad = \quad \frac{0.99}{0.05} \quad \approx \quad 20$$

representing (almost) strong evidence from the data in favor of disease.

Even without reference to prior probabilities (e.g., background rates), a positive test greatly increases your odds of having the disease.

Nonetheless, if your prior probability of having the disease is sufficiently small, your posterior probability will also be small.

General Case

If model H_m has sampling (data) distributions specified according to parameter θ_m, with densities

$$p(y \mid \theta_m, H_m)$$

then

$$\begin{aligned} p(y \mid H_m) &= \textit{marginal} \text{ data density under model } H_m \\ &= \int p(y \mid \theta_m, H_m)\, p(\theta_m \mid H_m)\, d\theta_m \end{aligned}$$

where $p(\theta_m \mid H_m)$ is the prior for model H_m.

Thus, Bayes factor depends on priors chosen for H_1 and for H_2.

Note: Both priors must be proper – otherwise marginal densities will depend on arbitrary scaling factors, and the Bayes factor will be undefined.

Application

Consider model with parameter θ and two disjoint sub-models

$$H_1 : \theta \in \Theta_1 \qquad\qquad H_2 : \theta \in \Theta_2$$

each of positive prior probability.

Then the Bayes factor favoring H_2 is

$$\frac{\Pr(\theta \in \Theta_2 \mid y)}{\Pr(\theta \in \Theta_1 \mid y)} \Bigg/ \frac{\Pr(\theta \in \Theta_2)}{\Pr(\theta \in \Theta_1)}$$

which can be used instead of a classical hypothesis test.

Drawbacks

In the general case, Bayes factors

- Can be sensitive to aspects of models that shouldn't matter.
- Can give paradoxical results, especially when parameter spaces have different dimensions (BDA3, Sec. 7.4).