Marginal Likelihoods and Bayes Factors

By : Mehdi Hamedi, MD

Psychiatrist

Marginal likelihood is the core component of Bayesian model comparison. So, in this article, we are going to discuss about Marginal likelihood and model comparison based on Bayes Factor.

Bayes theorem

Let’s to restate Bayes theorem:

$$\underbrace{\boxed{P(\theta \mid y)}}_{\textit{Posterior}} = \left(\underbrace{\boxed{P(y \mid \theta)}}_{\textit{Likelihood}} \times \underbrace{\boxed{P(\theta)}}_{\textit{prior}}\right) / \underbrace{\boxed{P(y)}}_{\substack{\textit{Marginal} \\ \textit{Likelihood}}} \quad (1)$$

In Bayesian inference, the last part or Marginal likelihood treated as a constant value and P(y) is dropped. So, the mentioned equation usually states as:

$$P(\theta \mid y) \propto P(y \mid \theta) \times P(\theta)) \quad (2)$$

Marginal Likelihood definition

P(y) is nominated as normalizing factor, marginal likelihood or model evidence plays a critical role in Bayesian model comparison because it quantifies the evidence the data y provide for the model.

To understand why, let's to remind that all of the components of the equation 1 are conditional on particular model (M). So the equation should be restated as:

$$\underbrace{\boxed{P(\theta|y,M)}}_{\text{Posterior}}=(\underbrace{\boxed{P(y|\theta,M)}}_{\text{Likelihood}}\times\underbrace{\boxed{P(\theta|M)}}_{\text{prior}})/\underbrace{\boxed{P(y|M)}}_{\text{evidence}} \quad (3)$$

The model evidence or marginal likelihood or $P(y|M)$ actually is the probability of observed data y under the model M.

How to calculate Marginal Likelihood?

The evidence is obtained by calculating the marginal likelihood of the data given the model:

$$P(y|M) = \int P(y|\theta, M)P(\theta|M)d\theta \quad (4)$$

Actually we are calculating how likely the data are for each point in parameter space and the averaging the resulting values.

Why is marginal likelihood important?

One property of marginal likelihood is that it accounts for the complexity of the model (M). Complex models tend to produce different patterns of data for different parameter values, so only a subset of parameter values will produce predictions similar to any given set of data. Accordingly, when we average $p(y|\theta,M)$ across the parameter space, the complex model will tend to return a lower average. If the simple model produces predictions close to the data, it will tend to do this irrespective of the parameter value, resulting in a larger average of $p(y|\theta,M)$. However, if the simple model is incapable of giving a good fit to the data, its average will be small. This means that the marginal likelihood does not simply punish complex models, but also rewards models for a good fit.

Importance of prior distribution

Prior distribution is another important part of marginal likelihood calculation. The prior represents our knowledge about different value of our parameter (θ). So according to equation 4, the prior distribution plays an important role in marginal likelihood calculation. This average is highly influenced by values of our parameter in prior distribution. This means that a more complex model can also return a high marginal likelihood if our priors turn out to heavily weight those parameter values that give the best fit to the data.

What is Bayes Factor?

To compare models using marginal likelihood, the ratio of marginal likelihoods can determine whether one model's evidence is over than other one's. This ratio is Bayes Factor:

$$BF_{ij} = \frac{P(y|M_i)}{P(y|M_j)} = \frac{\int P(y|\theta, M_i)P(\theta|M_i)d\theta}{\int P(y|\theta, M_j)P(\theta|M_j)d\theta}$$

The subscripts to the Bayes factor, BF_{ij} denote the two models being compared, the model i in the numerator of the Bayes factor ratio, and model j in the denominator.

Accordingly, $BF_{ij} > 1$ indicates that the data provide evidence for model i over model j, while $BF_{ij} < 1$ provides evidence for model j.

Bayes Factor is a continues metric of one model in comparison to another one. There is no arbitrary threshold for significance. But some authors have recommended some threshold for Bayes Factors values.

For example, Jeffreys (1961) suggested

- $1 \leq BF < 3.2$ is worth no more than a bare mention
- $3.2 \leq BF < 10$ offers substantial evidence
- $10 \leq BF < 100$ strong evidence
- $BF \geq 100$ is decisive

In either case, the exact values are not critical, and other authors have suggested other heuristics with different breakpoints.