Chapter 10: Model comparison and Hierachical Modelling

Shubham Gupta

May 13, 2019

1 Introduction

- When we have multiple models describing the same data, we need to assign credibilities to each model.
- Bayesian model comparison reallocates credibility across models given the data.
- Model comparison $\implies$ bayesian estimatation of hierachical models where the top-level is the index of the models.

2 Bayes Factor

2.1 General Formula

- Assume we have data D with parameters θ.
- Prior distribution is $p(\theta)$
- Parameter m to specify the index of the model.
- Hence, we will get

$$likelihood = p_m(y|\theta_m, m).$$

$$prior = p(\theta_m|m).$$

- Priors have different subscripts because they might have different distributions for each model.
- Assume each model is given a prior probability of $p(\theta)$. Then, for all possible models $\theta_1, \theta_2 \dots m$, we have:

$$p(\theta_1, \theta_2 \dots |D) = \frac{P(D|\theta_1, \theta_2 \dots m) * p(\theta_1, \theta_2 \dots m)}{\sum_m \int d\theta_m p(D|\theta_1, \theta_2 \dots m) p(\theta_1, \theta_2 \dots m)}.$$

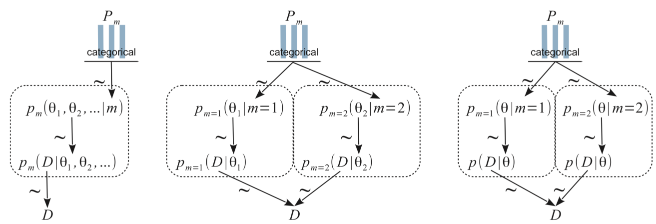

Figure 1: Model comparisons as a hierachical model

- To get the relative probabilities of the models, we will divide their posterior outputs.

$$\frac{p(m=1|D)}{p(m=2|D)} = \frac{p(D|m=1) * p(m=1)/\sum_m P(D|m) * p(m)}{p(D|m=2) * p(m=2)/\sum_m P(D|m) * p(m)}.$$

- The above equation is called the Bayes Factor
- We can use the below table for reference on figuring out when to report a model is better than the alternative model.

K	**dHart**	**bits**	**Strength of evidence**
$< 10^0$	0	—	Negative (supports M_2)
10^0 to $10^{1/2}$	0 to 5	0 to 1.6	Barely worth mentioning
$10^{1/2}$ to 10^1	5 to 10	1.6 to 3.3	Substantial
10^1 to $10^{3/2}$	10 to 15	3.3 to 5.0	Strong
$10^{3/2}$ to 10^2	15 to 20	5.0 to 6.6	Very strong
$> 10^2$	> 20	> 6.6	Decisive

Figure 2: bayes factor

3 Head biased vs tail biased factories

- Two factories that produce head-biased and tail biased factories. Given we have seen some tosses, which factory did the coin come from?
- We have the following hierachy

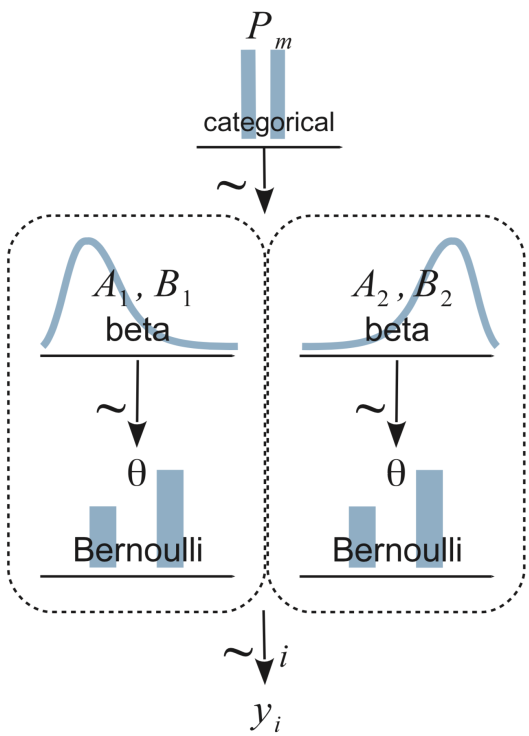

Figure 3: Coin Model Hierachy

3.1 MCMC Method: Individual models

- Main formula to compute the probability of the data is:

$$\frac{1}{p(D)} = \frac{1}{N} \sum_{n=\theta_i \tilde{p}(\theta|D)}^{N} \frac{h(\theta_i)}{p(D|\theta_i)p(\theta_i)}.$$

- $h(\theta_i$ is a probability density function. There is a complex derivation to this formula, which we are skipping for now. Refer to page 275.

3.2 MCMC Method: Hierachical model

- Similar to pymc3 models you have used. Use a index ‘m‘ to indicate for which model are the parameters being specified.
-