# Bayesian Hypothesis Testing & the Bayes Factor

Kenneth Chu Study Notes January 1, 2013

## 1 The Bayes Factor

Suppose we are interested in a certain physical phenomenon. We have designed and carried out an experiment to study it. In particular, suppose that we would like to compare two statistical models for the underlying data generation mechanism of the experiment. Let $\Omega$ be observation space of the experiment.

In order to perform a Bayesian statistical model comparison, the following entities must be given:

1. Model 1
   a) parameter space $\Theta_1$.
   b) prior density $\pi_1 : \Theta_1 \longrightarrow [0,\infty)$.
   c) sampling density (conditional probability): $f_1 : \Omega \times \Theta_1 \longrightarrow [0,\infty)$.
2. Model 2
   a) parameter space $\Theta_2$.
   b) prior density $\pi_2 : \Theta_2 \longrightarrow [0,\infty)$.
   c) sampling density (conditional probability): $f_2 : \Omega \times \Theta_2 \longrightarrow [0,\infty)$.
3. prior densities $P(M_1), P(M_2) \geq 0$, with $P(M_1) + P(M_2) = 1$.

**Remark 1.1** *Note:*

$$\int_{\Theta_1} \pi_1(\theta_1)d\theta_1 = \int_{\Theta_2} \pi_2(\theta_2)d\theta_2 = 1$$

$$\int_{\Omega} f_1(y|\theta_1)dy = 1, \text{ for each } \theta_1 \in \Theta_1, \quad \text{and} \quad \int_{\Omega} f_2(y|\theta_2)dy = 1, \text{ for each } \theta_2 \in \Theta_2$$

With the above given entities, we may define the following function:

$$f : \Omega \times \{1,\, 2\} \longrightarrow [0,\infty), \quad f(y,i) := P(M_i) \cdot \int_{\Theta_i} f_i(y|\theta_i)\pi_i(\theta_i)d\theta_i$$

**Lemma 1.2** $f : \Omega \times \{1,\, 2\} \longrightarrow [0,\infty)$ *integrates to* 1 *over* $\Omega \times \{1,\, 2\}$*, and hence can be regarded as a probability density function on* $\Omega \times \{1,\, 2\}$.

Proof

$$\begin{aligned}
\int_{\Omega\times\{1,2\}} f &= \sum_{m=1}^{2} \int_{\Omega} f(y,m)dy = \sum_{m=1}^{2} \int_{\Omega} \left( P(M_m) \cdot \int_{\Theta_m} f_m(y|\theta_m)\pi_m(\theta_m \right) d\theta_m \\
&= \sum_{m=1}^{2} P(M_m) \cdot \int_{\Theta_m} \left( \int_{\Omega} f_m(y|\theta_m)dy \right) \pi_m(\theta_m)d\theta_m \\
&= \sum_{m=1}^{2} P(M_m) \int_{\Theta_m} 1 \cdot \pi_m(\theta_m)d\theta_m \\
&= \sum_{m=1}^{2} P(M_m) = 1
\end{aligned}$$

□

With respect to the joint probability distribution defined on $\Omega \times \{1,\,2\}$ via the preceding Lemma, we have

$$P(M = m | Y = y) \;\; := \;\; \frac{f(y,m)}{f(y)} \;\; = \;\; \frac{f(y,m)}{f(y,1) + f(y,2)}$$

Hence,

$$\begin{aligned}
\frac{P(M = 1 | Y = y)}{P(M = 2 | Y = y)} \;\; &= \;\; \frac{f(y,1)/f(y)}{f(y,2)/f(y)} \;\; = \;\; \frac{f(y,1)/(f(y,1)+f(y,2))}{f(y,2)/(f(y,1)+f(y,2))} \;\; = \;\; \frac{f(y,1)}{f(y,2)} \\
&= \;\; \frac{P(M_1)\int_{\Theta_1} f_1(y|\theta_1)\pi_1(\theta_1)d\theta_1}{P(M_2)\int_{\Theta_2} f_2(y|\theta_2)\pi_2(\theta_2)d\theta_2} \\
&= \;\; \frac{P(M_1)}{P(M_2)} \cdot \left(\frac{\int_{\Theta_1} f_1(y|\theta_1)\pi_1(\theta_1)d\theta_1}{\int_{\Theta_2} f_2(y|\theta_2)\pi_2(\theta_2)d\theta_2}\right) \\
&= \;\; \frac{P(M_1)}{P(M_2)} \cdot B\left(M_1 : M_2\right),
\end{aligned}$$

where

$$B\left(M_1 : M_2\right) \;\; := \;\; \frac{\int_{\Theta_1} f_1(y|\theta_1)\pi_1(\theta_1)d\theta_1}{\int_{\Theta_2} f_2(y|\theta_2)\pi_2(\theta_2)d\theta_2}$$

is called the *Bayes factor* in favour of the model $M_1$ against the model $M_2$.

**Remark 1.3** *Note that*

$$m_{M_i}(y) \;\; := \;\; \int_{\Theta_i} f_i(y|\theta_i)\pi_i(\theta_i)d\theta_i$$

*is simply the marginal density of the observed data $y$ under model $M_i$. Hence, computationally speaking, the Bayes factor $B(M_1 : M_2)$ is simply the ratio of the marginal density of $y$ under $M_1$ to that under $M_2$.*

**Remark 1.4** *On the other hand, we also have:*

$$B(M_1 : M_2) \;\; = \;\; \frac{P(M_1|Y = y)/P(M_2|Y = y)}{P(M_1)/P(M_2)}.$$

*Hence, the Bayes factor $B(M_1 : M_2)$ is, probability-theoretically, the odds ratio of the posterior odds $\dfrac{P(M_1|Y = y)}{P(M_2|Y = y)}$ to the prior odds $\dfrac{P(M_1)}{P(M_2)}$. It is this probabilistic interpretation of the Bayes factor that enables it to be used in Bayesian model comparison.*

**Remark 1.5** *Recall that the map*

$$g_i : \Omega \times \Theta_i \longrightarrow [0,\infty) : (y,\theta_i) \longmapsto f_i(y|\theta_i)\pi_i(\theta_i)$$

*integrates to 1 over $\Omega \times \Theta_i$, and hence defines a probability density on $\Omega \times \Theta_i$. And,*

$$m_{M_i}(y) \;\; := \;\; \int_{\Theta_i} f_i(y|\theta_i)\pi_i(\theta_i)d\theta_i$$

*is thus the marginal density on the observation space $\Omega$ with respect to the joint density $g_i$ on $\Omega \times \Theta_i$. We now verify that $g_i$ indeed integrates to 1:*

$$\begin{aligned}\int_{\Omega\times\Theta_i} g_i &= \int_{\Theta_i}\int_{\Omega} g_i(y,\theta_i)dy\,d\theta_i \;=\; \int_{\Theta_i}\int_{\Omega} f_i(y|\theta_i)\pi_i(\theta_i)dy\,d\theta_i \;=\; \int_{\Theta_i}\left(\int_{\Omega} f_i(y|\theta_i)dy\right)\pi_i(\theta_i)d\theta_i \\ &= \int_{\Theta_i} 1\cdot\pi_i(\theta_i)d\theta_i \;=\; 1\end{aligned}$$

## References