Chapter 4. Bayesian Analysis (C1)

♠ Bayesian Inference (continued)

• Hypothesis Testing

Consider null hypothesis H_0: $\theta \in \Theta_0$ versus alternative hypothesis H_1: $\theta \in \Theta_1$, where $\Theta_0 \cap \Theta_1 = \emptyset$ and $\Theta_0 \cup \Theta_1 = \Theta$.

Let $\alpha_0 = P(\Theta_0|\boldsymbol{x})$ and $\alpha_1 = P(\Theta_1|x)$ denote the posterior probabilities and $\pi_0 = P(\Theta_0)$ and $\pi_1 = P(\Theta_1)$ denote the prior model probabilities.

Definition:

The ratio α_0/α_1 is called the *posterior odds ratio* of H_0 to H_1, and π_0/π_1 is called the *prior odds ratio.* The quantity

$$B = \frac{\text{posterior odds ratio}}{\text{prior odds ratio}} = \frac{\alpha_0/\alpha_1}{\pi_0/\pi_1} = \frac{\alpha_0\pi_1}{\alpha_1\pi_0}$$

is called the *Bayes factor* in favor of Θ_0.

Note:

Bayes factor can be viewed as the “odds for H_0 to H_1 that are given by the data.

Simple Hypothesis Case:

Suppose $\Theta_0 = \{\theta_0\}$ and $\Theta_1 = \{\theta_1\}$. Then

$$\alpha_i = \frac{\pi_i f(\boldsymbol{x}|\theta_i)}{\pi_0 f(\boldsymbol{x}|\theta_0) + \pi_1 f(\boldsymbol{x}|\theta_1)}, \quad i = 0, 1,$$

$$\frac{\alpha_0}{\alpha_1} = \frac{\pi_0 f(\boldsymbol{x}|\theta_0)}{\pi_1 f(\boldsymbol{x}|\theta_1)},$$

and

$$B = \frac{\alpha_0 \pi_1}{\alpha_1 \pi_0} = \frac{f(\boldsymbol{x}|\theta_0)}{f(\boldsymbol{x}|\theta_1)}.$$

Thus, the Bayes factor B is just the likelihood ratio of H_0 to H_1.

General Case:

Let

$$\pi(\theta) = \begin{cases} \pi_0 g_0(\theta) & \text{if } \theta \in \Theta_0, \\ \pi_1 g_1(\theta) & \text{if } \theta \in \Theta_1, \end{cases}$$

so that g_0 and g_1 are (proper) densities which describe how the prior mass is spread out over the two hypotheses. Let $f(\boldsymbol{x}|\theta)$ denote the likelihood function. Then, the posterior distribution is given by

$$\begin{aligned} \pi(\theta|\boldsymbol{x}) &= \frac{f(\boldsymbol{x}|\theta)\pi(\theta)}{\int_\Theta f(\boldsymbol{x}|\theta)\pi(\theta)d\theta} \\ &= \frac{f(\boldsymbol{x}|\theta)\pi(\theta)}{\int_{\Theta_0} f(\boldsymbol{x}|\theta)\pi_0 g_0(\theta)d\theta + \int_{\Theta_1} f(\boldsymbol{x}|\theta)\pi_1 g_1(\theta)d\theta} \\ &= \begin{cases} \frac{f(\boldsymbol{x}|\theta)\pi_0 g_0(\theta)}{m(\boldsymbol{x})} & \text{if } \theta \in \Theta_0, \\ \frac{f(\boldsymbol{x}|\theta)\pi_1 g_1(\theta)}{m(\boldsymbol{x})} & \text{if } \theta \in \Theta_1, \end{cases} \end{aligned}$$

where

$m(\boldsymbol{x}) = \int_{\Theta_0} f(\boldsymbol{x}|\theta)\pi_0 g_0(\theta)d\theta + \int_{\Theta_1} f(\boldsymbol{x}|\theta)\pi_1 g_1(\theta)d\theta.$

Thus,

$$\begin{aligned}\alpha_0 =& P(\Theta_0|x) = \int_{\Theta_0} \pi(\theta|\boldsymbol{x})d\theta \\ =& \int_{\Theta_0} [f(\boldsymbol{x}|\theta)\pi_0 g_0(\theta)/m(\boldsymbol{x})]d\theta \\ =& \pi_0 \int_{\Theta_0} f(\boldsymbol{x}|\theta)dF^{g_0}(\theta)/m(\boldsymbol{x})\end{aligned}$$

and

$$\begin{aligned}\alpha_1 =& P(\Theta_1|x) = \int_{\Theta_1} \pi(\theta|\boldsymbol{x})d\theta \\ =& \int_{\Theta_1} [f(\boldsymbol{x}|\theta)\pi_1 g_1(\theta)/m(\boldsymbol{x})]d\theta \\ =& \pi_1 \int_{\Theta_1} f(\boldsymbol{x}|\theta)dF^{g_1}(\theta)/m(\boldsymbol{x}).\end{aligned}$$

It follows that

$$\frac{\alpha_0}{\alpha_1} = \frac{\pi_0 \int_{\Theta_0} f(\boldsymbol{x}|\theta)dF^{g_0}(\theta)}{\pi_1 \int_{\Theta_1} f(\boldsymbol{x}|\theta)dF^{g_1}(\theta)},$$

and the Bayes factor is

$$B = \frac{\int_{\Theta_0} f(\boldsymbol{x}|\theta)dF^{g_0}(\theta)}{\int_{\Theta_1} f(\boldsymbol{x}|\theta)dF^{g_1}(\theta)}.$$

Thus, the Bayes factor is the ratio of "weighted" (by g_0 and g_1) likelihoods of Θ_0 and Θ_1. Because of the involving of g_0 and g_1, this cannot be viewed as a measure of the relative support for the hypotheses provided solely by the data. Sometimes, however, B will be relatively insensitive to reasonable choices of g_0 and g_1, and then such an interpretation is reasonable. The main operational advantage of having such a "stable" Bayes factor is that a scientific report could include this Bayes factor, and anyone could then determine his/her personal posterior odds by simply multiplying the reported Bayes factor by his/her personal prior odds, i.e.,

$$\text{posterior odds} = \text{Bayes factor} \times \text{prior odds}.$$

Note:

In both simple and general hypothesis cases, the Bayes factor does not depend on the prior probabilities π_0 and π_1.

Example 1: Assume $X \sim N(\theta, \sigma^2)$, where θ is unknown and σ^2 is known. Let $\pi(\theta)$ be a $N(\mu, \tau^2)$ density. Then

$$\theta | x \sim N(\mu(x), 1/\rho),$$

where

$$\mu(x) = \frac{\sigma^2}{\sigma^2 + \tau^2}\mu + \frac{\tau^2}{\sigma^2 + \tau^2}x = x - \frac{\sigma^2}{\sigma^2 + \tau^2}(x - \mu),$$

and $\rho = \text{precision} = \frac{1}{\sigma^2} + \frac{1}{\tau^2}$.

Consider the situation wherein a child is given an intelligence test. Assume that the test result $X \sim N(\theta, 100)$, where θ is the true IQ (intelligence) level of the child, as measured by the test. In other words, if the child were to take a large number of independent similar tests, his/her average score would be about θ. Assume also that, in the population as a whole, $\theta \sim N(100, 225)$. Using the above calculation, it follows that, marginally, $X \sim N(100, 325)$, while the posterior distribution of θ given x is normal with mean

$$\mu(x) = \frac{100(100) + x(225)}{100 + 225} = \frac{400 + 9x}{13}$$

and variance

$$\rho^{-1} = \frac{100(225)}{100 + 225} = \frac{900}{13} = 69.23.$$

Thus, if a child scores $x = 115$ on the test, his/her IQ θ has a $N(110.39, 69.23)$ posterior distribution.

The child taking the IQ test is to be classified as having below average IQ (less than or equal to 100) or above average (greater than 100). Formally, it is thus desired to test

$$H_0\colon\ \theta \leq 100 \text{ versus } H_1\colon\ \theta > 100.$$

Since

$$\theta | x = 115 \sim N(110.39, 69.23),$$

we have

$$\alpha_0 = P(\theta \leq 100 | x = 115) = 0.106,$$

$$\alpha_1 = P(\theta > 100 | x = 115) = 0.894,$$

and hence the posterior odds ratio is

$$\alpha_0 / \alpha_1 = 1/8.44.$$

Also, the prior is $N(100, 225)$, so that

$$\pi_0 = P^{\pi}(\theta \leq 100) = \frac{1}{2} = \pi_1$$

and the prior odds ratio is 1. Note that a prior odds ratio 1 indicates that H_0 and H_1 are viewed as equally plausible initially. The Bayes factor is thus

$$B = \frac{\alpha_0 \pi_1}{\alpha_1 \pi_0} = \frac{1}{8.44}.$$

A Question:

What are g_0 and g_1 for this case?

Interpretation of Bayes Factor:

Bayes factor is a summary of the evidence provided by the data in favor of one scientific theory, represented by a statistical hypothesis, as opposed to another. Jeffreys (1961) suggested interpreting B in half-units on the $\log_{10}$ scale. Kass and Raftery (1995, JASA) suggested the following interpretation of the Bayes factor.

$\log_{10}(B)$	B	Evidence against H_0
0 to 1/2	1 to 3.2	Not worth more than a bare mention
1/2 to 1	3.2 to 10	Substantial
1 to 2	10 to 100	Strong
> 2	> 100	Decisive

A modified version of the above interpretation was offered by Kass and Raftery (1995, JASA) based on the natural logarithm scale. Then, we have the following rule of thumb.

$2\log_e(B)$	B	Evidence against H_0
0 to 2	1 to 3	Not worth more than a bare mention
2 to 6	3 to 20	Positive
6 to 10	20 to 150	Strong
> 10	> 150	Very Strong

One-Sided Testing:

One-sided hypothesis testing occurs when $\Theta \subset R^1$ and Θ_0 is entirely to one side of Θ_1.

Example 2: When $X \sim N(\theta, \sigma^2)$ and θ has the noninformative prior $\pi(\theta) = 1$, we have

$$\theta|x \sim N(x, \sigma^2).$$

Consider the situation of testing H_0: $\theta \leq \theta_0$ versus H_1: $\theta > \theta_0$. Then

$$\alpha_0 = P(\theta \leq \theta_0|x) = \Phi((\theta_0 - x)/\sigma),$$

where Φ is the standard normal c.d.f.

The classical P-value against H_0 is the probability, when $\theta = \theta_0$, of observing an X "more extreme" than the actual data x. Here the P-value would be

$$P\text{-value} = P(X \geq x|\theta = \theta_0) = 1 - \Phi((\theta_0 - x)/\sigma).$$

Because of the symmetry of the normal distribution, it follows that α_0 equals the P-value against H_0.

Testing a Point Null Hypothesis:

Consider a test of H_0: $\theta = \theta_0$ versus H_1: $\theta \neq \theta_0$. The marginal density of X is

$$m(\boldsymbol{x}) = \int f(\boldsymbol{x}|\theta) dF^{\pi}(\theta) = f(\boldsymbol{x}|\theta_0)\pi_0 + (1-\pi_0) m_1(\boldsymbol{x}),$$

where

$$m_1(\boldsymbol{x}) = \int_{\{\theta \neq \theta_0\}} f(\boldsymbol{x}|\theta) dF^{g_1}(\theta)$$

is the marginal density of $\boldsymbol{X}$ with respect to g_1.
Hence the posterior probability that $\theta = \theta_0$ is

$$\pi(\theta_0|\boldsymbol{x}) = \frac{f(\boldsymbol{x}|\theta_0)\pi_0}{m(\boldsymbol{x})} = \frac{f(\boldsymbol{x}|\theta_0)\pi_0}{f(\boldsymbol{x}|\theta_0)\pi_0 + (1-\pi_0) m_1(\boldsymbol{x})}$$
$$= \left[1 + \frac{(1-\pi_0)}{\pi_0} \cdot \frac{m_1(\boldsymbol{x})}{f(\boldsymbol{x}|\theta_0)} \right]^{-1}.$$

Note that this is α_0, the posterior probability of H_0, and that $\alpha_1 = 1 - \alpha_0$ is hence the posterior probability of H_1. The posterior odds ratio is

$$\frac{\alpha_0}{\alpha_1} = \frac{\pi(\theta_0|\boldsymbol{x})}{1 - \pi(\theta_0|\boldsymbol{x})} = \frac{\pi_0}{\pi_1} \cdot \frac{f(\boldsymbol{x}|\theta_0)}{m_1(\boldsymbol{x})},$$

so that the Bayes factor for H_0 versus H_1 is

$$B = \frac{f(\boldsymbol{x}|\theta_0)}{m_1(\boldsymbol{x})}.$$

Example 3 (Lindley's paradox):

Suppose a random sample $X_1, \ldots, X_n$ is from $N(\theta, \sigma^2)$ (σ^2 known). Reduction to the sufficient statistic $\overline{X}$ yields the effective likelihood function

$$f(\bar{x}|\theta) = \frac{1}{\sqrt{2\pi\sigma^2/n}} \exp\left\{-\frac{n}{2\sigma^2}(\theta - \bar{x})^2\right\}.$$

Consider a test of H_0: $\theta = \theta_0$ versus H_1: $\theta \neq \theta_0$. Suppose g_1 is a $N(\mu, \tau^2)$ density on $\theta \neq \theta_0$. $m_1(\bar{x})$ is a $N(\mu, \tau^2 + \sigma^2/n)$ density. Thus,

$$m_1(\bar{x}) = \frac{1}{\sqrt{2\pi(\tau^2 + \sigma^2/n)}} \exp\left\{-\frac{n}{2(n\tau^2 + \sigma^2)}(\bar{x} - \mu)^2\right\},$$

and the Bayes factor is

$$\begin{aligned} B =& \frac{f(\bar{x}|\theta_0)}{m_1(\bar{x})} \\ =& \sqrt{\frac{n\tau^2 + \sigma^2}{\sigma^2}} \exp\left\{-\frac{n}{2\sigma^2}(\theta_0 - \bar{x})^2\right\} \\ & \times \exp\left\{\frac{n}{2(n\tau^2 + \sigma^2)}(\bar{x} - \mu)^2\right\}. \end{aligned}$$

It is easy to check that, *for any fixed* $\bar{x}$, $B \to \infty$ as $\tau^2 \to \infty$, so that evidence in favor of H_0 becomes overwhelming as the prior precision in H_1 gets vanishingly small, and hence $\alpha_0 = P(H_0|\bar{x}) \to 1$. In particular, this is true for $\boldsymbol{x}$ such that $|\bar{x} - \theta_0|/\sigma$ is large enough to cause the "null hypothesis" to be "rejected" at any arbitrary, prespecified level using a conventional significant test. This "paradox" was first discussed in detail by Lindley (1957) and has since occasioned considerable debate: see Smith (1965), Bernardo (1980), Berger and Delampady (1987), and many others.

Theorem 1: *For any distribution g_1 on $\theta \neq \theta_0$,*

$$\alpha_0 = \pi(\theta_0|\boldsymbol{x}) \geq \left[1 + \frac{1-\pi_0}{\pi_0} \cdot \frac{r(\boldsymbol{x})}{f(\boldsymbol{x}|\theta_0)}\right]^{-1},$$

where

$$r(\boldsymbol{x}) = \sup_{\theta \neq \theta_0} f(\boldsymbol{x}|\theta).$$

(Usually, $r(\boldsymbol{x}) = f(\boldsymbol{x}|\hat{\theta})$, where $\hat{\theta}$ is a maximum likelihood estimate of θ.) The corresponding bound on the Bayes factor for H_0 versus H_1 is

$$B = \frac{f(\bar{x}|\theta_0)}{m_1(\bar{x})} \geq \frac{f(\boldsymbol{x}|\theta_0)}{r(\boldsymbol{x})}.$$

Please read pages 152 – 156 for examples and discussions.

Multiple Hypothesis Testing:

For the multiple hypothesis testing problems, we simply calculate the posterior probability of each hypothesis and choose the one with the highest posterior probability.

- **Predictive Inference**

Suppose we consider the situation to predict a random variable $Z \sim g(z|\theta)$ based on the observation of $\boldsymbol{X} \sim f(\boldsymbol{x}|\theta)$. We assume $\boldsymbol{X}$ and Z are independent.

Then, the *predictive density* of Z given $\boldsymbol{x}$, when the prior for θ is π, is defined by

$$p(z|\boldsymbol{x}) = \int_\Theta g(z|\theta) dF^{\pi(\theta|\boldsymbol{x})}(\theta).$$