

Bayes Factor

Zeyong Jin

October 17, 2022

Contents

1	Bayesian Hypothesis Testing	2
1.1	Notation	2
1.2	Bayesian Approach	2
1.3	Expression	2
1.4	Difficulty	2
1.5	Loss Function	2
1.5.1	Example: 0 - 1 Loss	3
2	Bayes Factor	3
2.1	Precondition	3
2.2	Definition	3
2.3	Interpretation by Weight	3
2.4	Alternative Interpretation by Odds	3
2.5	Candidate's Formula	4
2.5.1	Interpretation	4
2.6	BMA	4
3	Controlling for look-elsewhere effects	4
3.1	Multiple Hypotheses	4
3.1.1	Key Idea	4
3.1.2	Process	4
3.2	The difficulty in frequentist control of the look-elsewhere effect	5
4	Bayesian Inference	5
4.1	Conceptualizing Hypothesis Testing via Bayes Factors	5
4.2	Properties	5
4.3	Objectives	5
4.4	Recommendation	6
4.4.1	Robustness Region	6
4.5	Comparison	6
4.5.1	Explanation	6
5	References	6

1 Bayesian Hypothesis Testing

1.1 Notation

Unknowns are \mathcal{H}_0 and \mathcal{H}_1 .

Put a prior on the actual hypotheses/models, that is, on $\pi(\mathcal{H}_0) = \Pr(\mathcal{H}_0 = \text{True})$ and $\pi(\mathcal{H}_1) = \Pr(\mathcal{H}_1 = \text{True})$.

1.2 Bayesian Approach

1. Put a prior on θ , $\pi(\theta) = \mathcal{N}(\theta_0, 1/\tau_0^2)$.
2. Compute posterior $\theta | y^{(n)} \sim \mathcal{N}(\theta_n, 1/\tau_n^2)$ for updated parameters θ_n and τ_n^2 .

1.3 Expression

Using Bayes Theorem, we have:

$$\pi(\mathcal{H}_1|Y) = \frac{p(y^{(n)}|\mathcal{H}_1)\pi(\mathcal{H}_1)}{p(y^{(n)}|\mathcal{H}_0)\pi(\mathcal{H}_0) + p(y^{(n)}|\mathcal{H}_1)\pi(\mathcal{H}_1)} \quad (1)$$

where $p(y^{(n)}|\mathcal{H}_1)$ and $p(y^{(n)}|\mathcal{H}_0)$ are the marginal likelihoods hypotheses.

If for example we set $\pi(\mathcal{H}_1) = 0.5$ and $\pi(\mathcal{H}_0) = 0.5$ as a *priori*, then

$$\begin{aligned} \pi(\mathcal{H}_1|Y) &= \frac{0.5p(y^{(n)}|\mathcal{H}_1)\pi(\mathcal{H}_1)}{0.5p(y^{(n)}|\mathcal{H}_0) + 0.5p(y^{(n)}|\mathcal{H}_1)} \\ &= \frac{p(y^{(n)}|\mathcal{H}_1)}{p(y^{(n)}|\mathcal{H}_0) + p(y^{(n)}|\mathcal{H}_1)} \\ &= \frac{1}{\frac{p(y^{(n)}|\mathcal{H}_0)}{p(y^{(n)}|\mathcal{H}_1)} + 1} \end{aligned}$$

1.4 Difficulty

A major difficulty in the implementation of Bayesian hypothesis testing is the choice of the needed prior distributions of unknown parameters. [Berger, 2011]

For a simple normal model, the only unknown parameter is θ

- Under $\mathcal{H}_0, \theta = 0$ with probability 1.
- Under $\mathcal{H}_0, \theta \in \mathbb{R}$ Could take $\pi(\theta) = \mathcal{N}(\theta_0, \frac{1}{\tau_0^2})$.
- Compute marginal likelihoods for each hypothesis, that is, $\mathcal{L}(\mathcal{H}_0)$ and $\mathcal{L}(\mathcal{H}_1)$.
- Obtain posterior probabilities of \mathcal{H}_0 and \mathcal{H}_1 via Bayes Theorem.

1.5 Loss Function

Loss function:

$$L(\hat{\mathcal{H}}, \mathcal{H}) = w_1 1(\hat{\mathcal{H}} = 1, \mathcal{H} = 0) + w_2 1(\hat{\mathcal{H}} = 0, \mathcal{H} = 1) \quad (2)$$

- w_1 weights how bad making a Type I error.
- w_2 weights how bad making a Type II error.

1.5.1 Example: 0 - 1 Loss

0 - 1 loss is the loss function $L(h, d) = \mathbf{1}(h \neq d)$, i.e., you lose 1 if wrong, 0 if right.

- The posterior expected loss in this case is

$$\begin{aligned} E(L(H, d) | x) &= \sum_h L(h, d) p(H = h | x) \\ &= \sum_h \mathbf{1}(h \neq d) p(H = h | x) \\ &= 1 - p(H = d | x). \end{aligned}$$

- So, to minimize our posterior expected loss, the optimal decision d^* (under 0 - 1 loss) is the hypothesis with highest posterior probability $p(H = d | x)$.

- In the case of two hypotheses, H_0 and H_1 ,

$$d^* = \begin{cases} H_0 & \text{if } p(H_0 | x) > 1/2 \\ H_1 & \text{if } p(H_1 | x) > 1/2 \\ \text{either} & \text{otherwise.} \end{cases}$$

2 Bayes Factor

2.1 Precondition

To calculate a Bayes factor, we must first specify a model of the data assuming there is a true effect. [Lakens et al., 2018]

2.2 Definition

$$\text{Bayes Factor} = \frac{p(y^{(n)} | \mathcal{H}_0)}{p(y^{(n)} | \mathcal{H}_1)}, \quad (3)$$

where

$$p(y^{(n)} | \mathcal{H}_k) = \int f(y^{(n)} | \theta, \sigma^2, \mathcal{H}_k) \pi(\theta | \sigma^2, \mathcal{H}_k) \pi(\sigma^2 | \mathcal{H}_k) d\theta d\sigma^2 \quad (4)$$

is the marginal probability of the data under model \mathcal{H}_k . [Chen et al., 2018]

(2) is known as the **Bayes Factor** in favor of \mathcal{H}_0 , and often written as \mathcal{BF}_{01} .

When estimating the posterior distribution of θ under each model, we find the hyperparameters, namely λ, μ_0, a_0 , and b_0 , have little effects on the posterior distributions in general. We use $\lambda = 1$ for precision parameter following the recommendations in [Kass and Raftery, 1995].

2.3 Interpretation by Weight

Bayes factor is a ratio of marginal likelihoods and it provides a **weight of evidence** in the data in favor of one model over another. It is often used as an alternative to the frequentist p-value. $\mathcal{BF}_{01} > 10$ is strong, $\mathcal{BF}_{01} > 100$ is decisive (Rule of thumb).

2.4 Alternative Interpretation by Odds

Bayes factor can be thought of as the factor by which our **prior odds change** (towards the posterior odds) in the light of the data.

$$\begin{aligned}
\frac{\pi(\mathcal{H}_0 | Y)}{\pi(\mathcal{H}_1 | Y)} &= \frac{p(y^{(n)} | \mathcal{H}_0) \pi(\mathcal{H}_0)}{p(y^{(n)} | \mathcal{H}_0) \pi(\mathcal{H}_0) + p(y^{(n)} | \mathcal{H}_1) \pi(\mathcal{H}_1)} \div \frac{p(y^{(n)} | \mathcal{H}_1) \pi(\mathcal{H}_1)}{p(y^{(n)} | \mathcal{H}_0) \pi(\mathcal{H}_0) + p(y^{(n)} | \mathcal{H}_1) \pi(\mathcal{H}_1)} \\
&= \frac{p(y^{(n)} | \mathcal{H}_0) \pi(\mathcal{H}_0)}{p(y^{(n)} | \mathcal{H}_0) \pi(\mathcal{H}_0) + p(y^{(n)} | \mathcal{H}_1) \pi(\mathcal{H}_1)} \times \frac{p(y^{(n)} | \mathcal{H}_0) \pi(\mathcal{H}_0) + p(y^{(n)} | \mathcal{H}_1) \pi(\mathcal{H}_1)}{p(y^{(n)} | \mathcal{H}_1) \pi(\mathcal{H}_1)} \\
&\therefore \underbrace{\frac{\pi(\mathcal{H}_0 | Y)}{\pi(\mathcal{H}_1 | Y)}}_{\text{posterior odds}} = \underbrace{\frac{\pi(\mathcal{H}_0)}{\pi(\mathcal{H}_1)}}_{\text{prior odds}} \times \underbrace{\frac{p(y^{(n)} | \mathcal{H}_0)}{p(y^{(n)} | \mathcal{H}_1)}}_{\text{Bayes factor } \mathcal{BF}_{01}}
\end{aligned}$$

2.5 Candidate's Formula

Alternative expression for Bayes Factor

$$\frac{p(y^{(n)} | \mathcal{H}_0)}{p(y^{(n)} | \mathcal{H}_1)} = \frac{\pi_\theta(0 | y^{(n)}, \mathcal{H}_1)}{\pi_\theta(0 | \mathcal{H}_1)} \quad (5)$$

2.5.1 Interpretation

Ratio of the prior to posterior densities for θ evaluated at 0.

2.6 BMA

The BMA posterior relies on the Bayes factor as the weighting factor, favoring one model over using weights based on \mathcal{BF}_{01} .

Given an established association, we expect the Bayes factor provide evidence supporting one of the two models. Intuitively, if \mathcal{BF}_{01} then we have **more support** for \mathcal{H}_0 from the data and vice versa when $\mathcal{BF}_{01} < 1$. [Chen et al., 2018]

3 Controlling for look-elsewhere effects

A major strength of Bayesian analysis is that it easily (and often automatically) adjusts for look-elsewhere effects.

3.1 Multiple Hypotheses

3.1.1 Key Idea

The point here is that the marginal likelihood of hypothesis H_i gets automatically **down-weighted** by $\frac{1}{m}$, the 'cost' of looking in m different bins. [Berger, 2011]

3.1.2 Process

Suppose N_j of the X_i are in bin $B_j, j = 1, \dots, m$, and that we assume we have only densities $f_s(B_j)$ and $f_b(B_j)$. Then

$$B_{10} = \frac{\int_0^\infty \int_0^\infty b^N e^{-(b+s)} \prod_{i=1}^m \left[1 + \frac{sf_s(B_j)}{bf_b(B_j)} \right]^{N_j} \pi_0(b) \pi_1(s | b) ds db}{\int_0^\infty b^N e^{-b} \pi_0(b) db}. \quad (6)$$

Suppose, in addition, that $f_s(B_j)$ gives probability one to some unknown bin B , with each bin being equally likely. This is equivalent to saying that we are testing the mutually exclusive hypotheses: H_j : signal is only in bin B_j , with the hypotheses having equal prior probability. Then

$$\begin{aligned}
B_{10} &= \frac{E^B \left[\int_0^\infty \int_0^\infty b^N e^{-(b+s)} \prod_{i=1}^m \left[1 + \frac{s f_s(B)}{b f_b(B_j)} \right]^{N_j} \pi_0(b) \pi_1(s | b) ds db \right]}{\int_0^\infty b^N e^{-b} \pi_0(b) db} \\
&= \frac{\frac{1}{m} \sum_{j=1}^m \int_0^\infty \int_0^\infty b^N e^{-(b+s)} \left[1 + \frac{s}{b f_b(B_j)} \right]^{N_j} \pi_0(b) \pi_1(s | b) ds db}{\int_0^\infty b^N e^{-b} \pi_0(b) db}.
\end{aligned} \tag{7}$$

3.2 The difficulty in frequentist control of the look-elsewhere effect

Type I error that would result from a single test. Thus the needed frequentist control for multiple testing ranges from the drastic Bonferroni correction to none, depending on the correlations among the data.

In contrast, the Bayesian adjustment for multiple testing does not depend on correlations among the data, and occurs only through the choice of prior probabilities of hypotheses. In the above scenario, for instance, one might assign prior probability $\frac{1}{2}$ to no signal, and assign each of the possible alternative hypotheses prior probability of $\frac{1}{2m}$. The ensuing Bayesian analysis correctly controls for the look-elsewhere effect, **regardless of the data distribution**.

4 Bayesian Inference

4.1 Conceptualizing Hypothesis Testing via Bayes Factors

Simulating data many times from the posterior distribution will ideally yield representative samples of the unknown parameter that we can interpret to answer the research question. [Baig, 2019]

In the context of Bayesian inference, hypothesis testing can be framed as a special case of model comparison where a model refers to a likelihood function and a prior distribution.

The combination of the likelihood function for the observed data with each of the prior distributions yields hypothesis-specific models. For each of the hypothesis-specific models, averaging (ie, integrating) the likelihood with respect to the prior distribution across the entire parameter space yields the probability of the data under the model and, therefore, the corresponding hypothesis. This quantity is more commonly referred to as the marginal likelihood and represents the average fit of the model to the data. The ratio of the marginal likelihoods for both hypothesis-specific models is known as the Bayes factor.

The Bayes factor represents how we should update our knowledge about the hypotheses after examining data.

4.2 Properties

1. Bayes factors can provide direct evidence for the common null hypothesis of no difference.
2. Bayes factors can reveal when experimental data is insensitive to the null and alternative hypotheses, clearly suggesting that the researcher should withhold judgment
3. Bayes factors can be interpreted **continuously** and thus provide an indication of the strength of the evidence for the null or alternative hypothesis.

4.3 Objectives

To specify a model for the data if there is a true effect would be to consider the range of possible results based on the scale used in the research. The maximum possible effect when calculating proportions is a difference of 1. [Lakens et al., 2018]

Smaller differences should be more plausible. We can model this prior belief about smaller differences being more likely than larger differences by using a half-normal distribution, with an SD of 0.5.

4.4 Recommendation

Reporting robustness regions for each Bayes factor that is reported. [Lakens et al., 2018]

4.4.1 Robustness Region

A robustness region specifies the range of expected effect sizes used when specifying the alternative model that support the same conclusion (e.g., evidence for H_1 , evidence for H_0 , or inconclusive outcomes). Robustness regions are reported as Rob. Reg. [L, U], and give the lower and upper effect size for the alternative model that leads to the same conclusion, given a certain Bayes factor threshold.

4.5 Comparison

Bayes factors calculated for different alternative hypotheses can be compared in such a way when they have been calculated using the same data and against the same model of H_0 ($B_{H_1/H_0}/B_{H_2/H_0} = B_{H_1/H_2}$).

4.5.1 Explanation

Based on the best estimate of which effect sizes would be reasonable (based on related earlier research), the data are nonevidential.

5 References

References

- [Baig, 2019] Baig, S. A. (2019). Bayesian Inference: An Introduction to Hypothesis Testing Using Bayes Factors. *Nicotine Tobacco Research*, 22(7):1244–1246.
- [Berger, 2011] Berger, J. (2011). The Bayesian Approach to Discovery. pages 17–26.
- [Chen et al., 2018] Chen, B., Craiu, R. V., and Sun, L. (2018). Bayesian model averaging for the X-chromosome inactivation dilemma in genetic association study. *Biostatistics*, 21(2):319–335.
- [Kass and Raftery, 1995] Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.
- [Lakens et al., 2018] Lakens, D., McLatchie, N., Isager, P. M., Scheel, A. M., and Dienes, Z. (2018). Improving Inferences About Null Effects With Bayes Factors and Equivalence Tests. *The Journals of Gerontology: Series B*, 75(1):45–57.