

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
 Maths 2020; 5(3): 32-36
 © 2020 Stats & Maths
www.mathsjournal.com
 Received: 12-03-2020
 Accepted: 14-04-2020

Onrina Chandra
 Ballygunge Science College,
 University of Calcutta,
 Ballygunge Circular, Ballygunge,
 Kolkata, West Bengal, India

Corresponding Author:
Onrina Chandra
 Ballygunge Science College,
 University of Calcutta,
 Ballygunge Circular, Ballygunge,
 Kolkata, West Bengal, India

Testing of hypothesis using the bayes factor

Onrina Chandra

Abstract

In any testing problem, the most popular procedure to draw a conclusion regarding the null and alternative hypotheses is to use the p-value of the test. If the p-value is below a certain level of significance “the test is rejected” and if not “we fail to reject the null hypothesis based on the observed data”. What if we wanted to know how much more favoured the alternative hypothesis was, based on the data observed, than the null hypothesis. The p-value only taking into account the distribution under the null setup fails to answer this question. This is primarily why we use the Bayes Factor. The purpose of this paper is to provide a brief overview of the Bayes Factor. Using two simple examples the use of Bayes Factor in testing problems is depicted and conclusions drawn. The paper tries to establish the Bayes Factor as another practical tool for testing of hypotheses.

Keywords: bayes factor, testing of hypothesis, prior odds, posterior odds

1. Introduction

The Bayesian approach uses a fully probabilistic framework to draw conclusions based on the observed data. Given a null and an alternative statistical model, Bayesian analysis uses the idea of Bayes theorem to compute the posterior probability of one of the models being better, in light of the observed data. The Bayes Factor is a straightforward consequence of the Bayes theorem that allows us to do just so, that is, compare the plausibility of the two competing hypotheses relatively. Intuitively, it can be regarded as the weight of evidence provided by the data in favour of one of the hypotheses. It is the ratio of the odds of observing two competing hypotheses after investigating relevant data compared to the odds of observing those hypotheses before investigating the data.

The Bayes Factor for the alternative versus the null is a ratio comparing how likely the data is under the alternative model versus the null model. As the Bayes Factor in favour of the alternative increases, the probability of observing the alternative hypothesis increases, relative to the null. In case of a simple versus simple testing problem the Bayes Factor reduces to the likelihood ratio.

For example, Let the hypotheses H_1 and H_2 be deemed equally probable a priori, that is $p(H_1)=p(H_2)=0.5$. Then if the Bayes Factor in favour of the H_2 , $BF_{2,1}=8$, it means that after conducting the experiment, H_2 is deemed eight times more likely than H_1 . Bayes Factor being a ratio of probabilities take values between 0 and infinity. If it is greater than 1 we conclude that the alternative model is more likely but a natural question now is how do we draw conclusions for the testing problem based on the value of the Bayes Factor? A commonly used scale for drawing conclusions is the Jeffreys scale (1961) ^[1]

Table 1: This table shows Jeffreys scale

Jeffreys scale	
$BF_{2,1}$	Evidence against H_1
1 to 3	Not worth a bare mention
3 to 20	Moderate Evidence
20 to 150	Strong Evidence
> 150	Very strong Evidence

2. Method of calculating the Bayes Factor

Let the null hypothesis be denoted by H_0 and the alternative hypothesis be denoted by H_1 . Before the testing procedure is carried out it is assumed that prior probability of observing H_0 , or the prior probability of observing the parameter of interest under H_0 is $P(H_0)$. Similarly the prior probability of assuming H_1 is $P(H_1)$. We now define the prior odds and posterior odds. The prior odds is defined as ratio of observing the two hypotheses before examining the data. It is the ratio of the prior probabilities of the hypotheses H_1 to H_0 . In mathematical terms, $O_{10} = P(H_1)/P(H_0)$.

Similarly, the posterior odds is the ratio of the odds of observing two competing hypotheses after examining relevant data written as

$$PO_{10} = P(H_1|data)/P(H_0|data).$$

Using Bayes' theorem, the posterior odds can be written in terms of the prior odds as:

$$PO_{10} = \frac{P(H_1|data)}{P(H_0|data)} = \frac{(P(data|H_1) \times P(H_1))/P(data)}{(P(data|H_0) \times P(H_0))/P(data)} = \frac{(P(data|H_1) \times P(H_1))}{(P(data|H_0) \times P(H_0))} = \left(\frac{P(data|H_1)}{P(data|H_0)}\right) \times \left(\frac{P(H_1)}{P(H_0)}\right)$$

The Bayes Factor is the posterior odds in favour of one hypothesis when the prior probabilities on the two hypotheses are equal. Hence, the Bayes Factor is the ratio of posterior odds to prior odds.

3. Applications

The use of Bayes Factor is denoted using the following two examples.

3.1. Example 1. Suppose a manufacturer of fluorescent lamps claims that no more than 30% of his products is defective. A sample of 20 lamps is found to contain 4 defective lamps. We need to obtain whether the manufacturer's claim is justified. Here the null hypothesis is $H_0: p=0.3$ and the alternative hypothesis $H_1: p>0.3$.

We summarize this data using a Binomial(20, 4) likelihood function.. The Beta prior on p is appropriate to describe hypotheses as it is continuously defined on the interval between 0 and 1 and is not always symmetric. However looking at the value of p under null we need to choose a distribution which places the bulk of the probability density at or around 0.3. Therefore, we use a Beta(3,7) prior distribution to describe the null hypothesis. Similarly, we use a Beta(5,5) prior distribution to describe the alternative hypothesis as this skewed distribution places the bulk of the probability density at or around 0.5 (i.e., higher probability of defectives) and places less density on values less than 0.3.

3.1.1 Results

Assuming the prior odds to be 1 the Bayes Factor BF_{10} for H_1 against H_0 is 0.088914. As this value does not lie on Jeffrey's scale, we take the reciprocal of BF_{10} , which is nothing but BF_{01} , the Bayes Factor for H_0 against H_1 . Taking the reciprocal BF_{01} is 11.24671.

3.1.2 Discussion

Thus we draw the conclusion that there is moderate evidence in favour of the null hypothesis. Based on the data the null seems 11 times as likely as the alternative hypothesis.

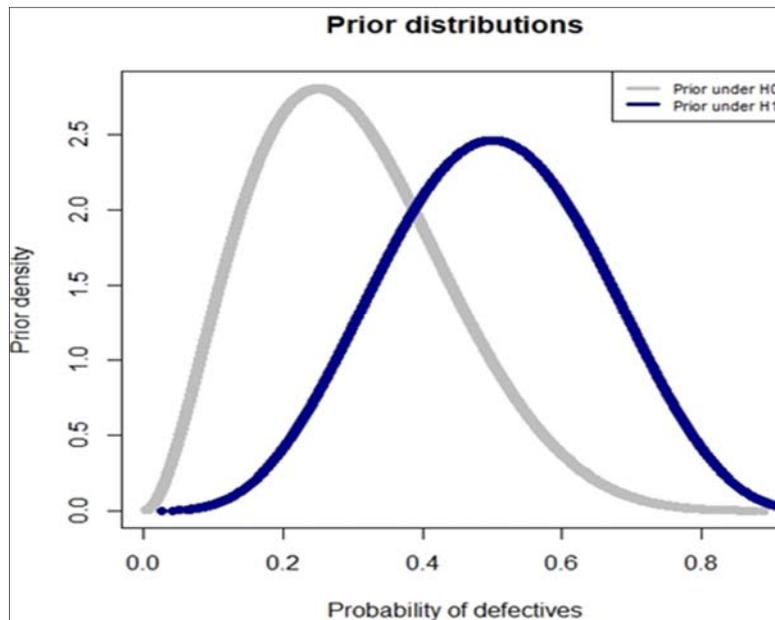


Fig 1: Graph showing the prior density under the competing hypotheses.

The Bayes factor and as a consequence the conclusion drawn depends on the choice of the prior distribution of the parameter of interest. If non informative priors are chosen we might mistakenly accept the wrong hypotheses. This often

requires a priori knowledge about the process and parameters to be tested. We now check how taking different Beta priors under the alternative would change the Bayes factor.

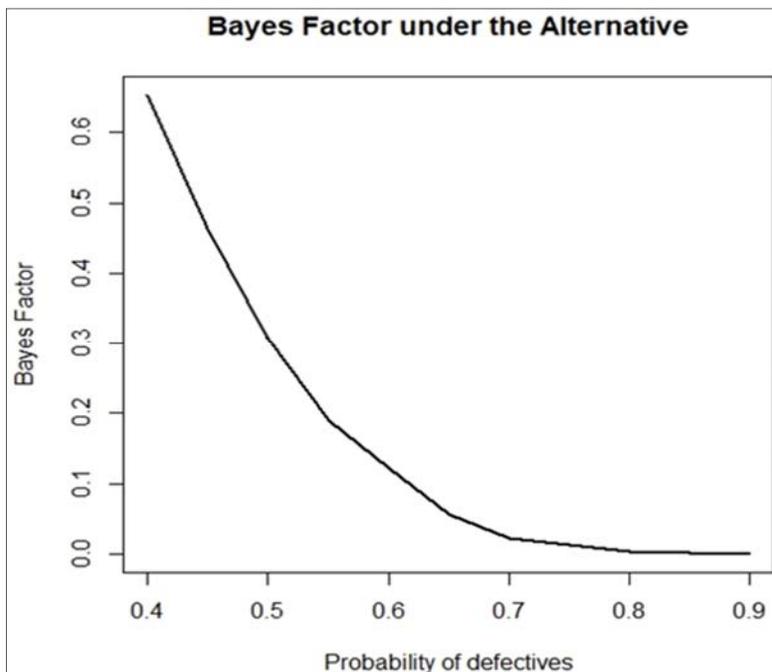


Fig 2: Graph showing variation in Bayes Factor due to the choice of Beta priors

Taking various prior Beta densities with increasing probabilities for defective items, as observed in Fig.2, the Bayes Factor BF_{01} decreases. The null hypothesis becomes more and more likely given the data as the probability of defective items under the alternative increases (which is reasonable!).

Reiterating, as seen in the above example the Bayes Factor is sensitive to the choice of the prior distribution. As the parameters of the Beta prior vary, so does the Bayes factor! Nevertheless, there are a few default priors which give fairly good results in most cases one such being the Jeffrey-Zellner-Siow prior which is used in the next example.

3.2. Example 2. Suppose we were given the values of the ages of having the first heart attack for a few people and we were required to test whether the average age for the first heart attack differed for men and women. The dataset used here is the Statlog Dataset on Heart Failure^[15]. We calculate the mean age for women and denote it by μ_{women} and that for men as μ_{men} . Let the Null Hypothesis be denoted by H_0 and the Alternative Hypothesis be denoted by H_1 . In mathematical terms we test

$$H_0: \mu_{\text{women}} = \mu_{\text{men}} \text{ against the alternative } H_1: \mu_{\text{women}} \neq \mu_{\text{men}}$$

We will view the data as a random sample from two populations, one for men and women. Let the total number of women= n_1 and the total number of men= n_2 . The two groups are modeled as:

$$X_i \sim N(\mu + \epsilon, \sigma^2) \text{ iid for all } i=1(1) n_1$$

$$Y_i \sim N(\mu - \epsilon, \sigma^2) \text{ iid for all } i=1(1) n_2$$

The model X_i stands for mean age of having the first heart attack for the i^{th} woman and the model Y_i stands for mean age of having the first heart attack for the i^{th} man. It is assumed that the observations are independent and identically distributed, with means $\mu + \epsilon$ and $\mu - \epsilon$ and variance σ^2 . Hence, the testing problem reduces to

$H_0: \epsilon = 0 \Leftrightarrow H_0: \mu = 0$ (The means are not different) versus $H_1: \epsilon \neq 0$ (The means are different)

To test the hypothesis, we require assumptions about prior distributions for ϵ, μ, σ^2 under both the hypotheses. For the continuous case, when the value of the parameter can take any value within a given range, the Bayes Factor is obtained by integrating over the entire range of the parameters under the models. In this case the Bayes Factor

$$BF_{01} = \frac{\iint p(\text{data}|\epsilon=0, \mu, \sigma^2) p(\mu, \sigma^2|H_0) d\mu d\sigma^2}{\iint p(\text{data}|\epsilon, \mu, \sigma^2) p(\epsilon|\sigma^2) p(\mu, \sigma^2|H_1) d\mu d\sigma^2 d\epsilon}$$

Under H_0 we assume ϵ is degenerate at 0 and under H_1 we use a Cauchy prior for ϵ . In Bayesian testing, the commonly used method is to use effect sizes (Cohen's d) to test the hypotheses. Effect sizes simplify comparison using the idea of plausibility of relative effects under the models. Here we use the standardized effect δ with the scale of r where $\delta = \epsilon / \sigma^2 \sim \text{Cauchy}(0, r^2)$. In terms of effect sizes, under H_0 gender is assumed to have no effect. Hence the entire probability of δ is concentrated at 0. Under the alternative model the effect size is assumed to take both positive and negative values with equal probability.

Now, under both the hypotheses, we use the Jeffreys-Zellner-Siow (JZS) default prior for μ and σ^2 . As mentioned above the JZS prior does not vary much with the choice of hyperparameters in the prior distributions. It also simplifies the computation of the Bayes Factor. Here we assume prior probability of both hypotheses to be equal, that is $P(H_0) = 0.5$ and $P(H_1) = 0.5$

3.2.1 Method and material used

We use R to calculate the Bayes factor.

R code

```
library(stats)
w=read.csv("C:/Users/ICLUSER/Desktop/Bookbayes.csv")
bayes_inference(y=difference,data=w,type="ht",statistic="mean",alternative="twosided", null=0)
```

3.2.2 Results

Running the code we obtained the value of $BF_{10}=8.9233$. That means the alternative is approximately 9 times as likely as the null set after examining the dataset. Therefore, based on the data we conclude that the mean ages for the first heart attack of men and women are not the same. There is moderate evidence that the mean ages are different for men and women.

3.2.3 Discussion

The posterior probabilities are obtained as $P(H_0|data) = 0.1008$ and $P(H_1|data) = 0.8992$. Thus we can estimate the likelihood of the posterior probability given the two models. This is another advantage of the Bayes Factor. The Bayes Factor helps us to compare between models in mathematical terms, using posterior probabilities.

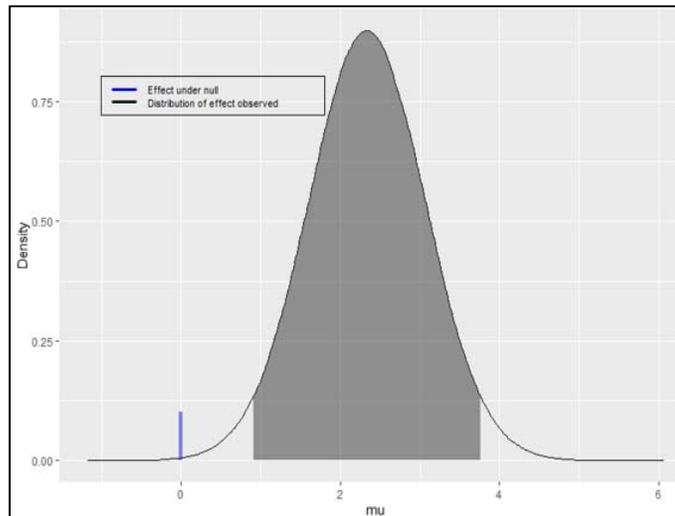


Fig 3: Plot showing the density of mean difference observed and under Null Hypothesis

The above plot gives us the likelihood of the standardized effect size under the data which is depicted by the grey shaded region and the distribution of effect size under H_0 which is assumed to be degenerate at 0. From the plot it can be concluded that the effect observed as given by the data set is far removed from the value zero. Therefore, on the basis of

the data it could be concluded that the effect is highly unlikely to have a distribution which has the bulk of its probability concentrated at zero.

As mentioned earlier, Bayes Factor is susceptible to the choice of the prior distribution. It varies according to the choice of the parameter r in our example.

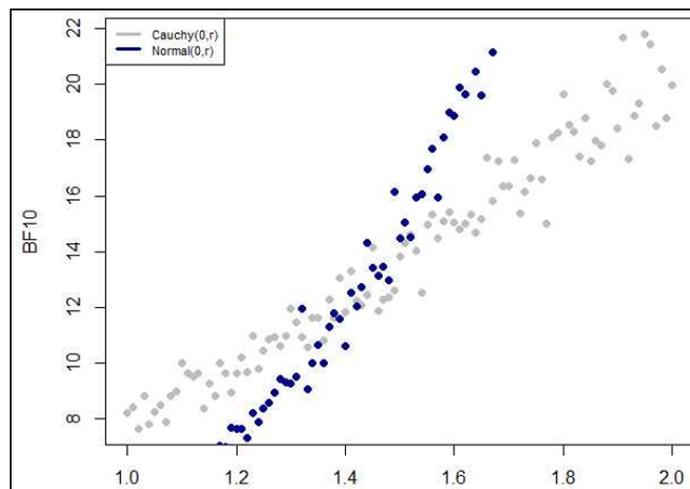


Fig 4: Plot showing the variation in Bayes factor due to the Prior.

However under few priors the Bayes Factor is known to be fluctuate less. For example, if we had assumed the standardized effect followed a Normal prior instead of a Cauchy Prior then the Bayes Factor as seen in Fig.4 would vary a lot more with the choice of the parameters of the prior distribution.

4. Conclusion

The advantage of the Bayes Factor over conventional methods of testing is that it incorporates information at all stages of the problem. It takes into account the prior probabilities of the

parameters and includes the likelihood of the data as well. As opposed to the p-value which considers the distribution under the null only, it allows us to make a comparative study of the null and alternative distributions.

Bayes factor assists in making conclusions regarding the unknown parameters of interest in concrete mathematical terms. It enables us to draw conclusion in terms of posterior and prior probabilities. However all of this can be done if prior distributions are carefully chosen. This requires some preconceived knowledge regarding the parameters, source of the data and the process of data collection.

Nevertheless, Bayes factor is an effective tool for applied statistics having many practical applications in the testing of hypotheses.

5. References

1. Jeffreys H. Theory of Probability, Third Edition. Oxford University Press, 1961.
2. Berger JO. Statistical decision theory and Bayesian analysis. Springer Science & Business Media, 2013.
3. Jeon M, De Boeck P. Decision qualities of Bayes factor and p value-based hypothesis testing. *Psychological Methods*. 2017; 22(2):340-360.
4. Berger JO, Pericchi LR. The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association*. 1996; 91(433):109-122.
5. Edwards W, Lindman H, Savage LJ. Bayesian statistical inference for psychological research. *Psychological Review*. 1963; 70(3):193-242.
6. Quintana DS, Williams DR. Bayesian alternatives for common null-hypothesis significance tests in psychiatry: a non-technical guide using JASP. *BMC Psychiatry*. 2018; 18:178.
7. Dienes Z. Using Bayes to get the most out of non-significant results. *Front Psychol*. 2014; 5(781):1-17.
8. Kass RE. Bayes factors in practice. *Journal of the Royal Statistical Society: Series D (The Statistician)*. 1993; 42(5):551-560.
9. Florens JP, Mouchart M. Bayesian testing and testing Bayesians. Université Catholique de Louvain. Center for Operations Research and Econometrics [CORE], 1992.
10. Kadane JB, Dickey JM. Bayesian decision theory and the simplification of models. In *Evaluation of econometric models*. Academic Press, 1980, 245-268.
11. Aitkin Murray. "Posterior bayes factors." *Journal of the Royal Statistical Society: Series B (Methodological)*. 1991; 53(1):111-128.
12. Kass Robert E, Adrian E Raftery. "Bayes factors and model uncertainty, 1995.
13. Jeffrey N, Rouder Richard, Morey D. Default Bayes Factors for Model Selection in Regression. *Multivariate Behavioral Research*. 2012; 47:877-903.
14. Robert Kass E, Adrian Raftery E. Bayes Factors, *Journal of the American Statistical Association*. 1995; 90:430, 773-7953.
15. Statlog, "Heart Disease Data Set", Available at: <https://archive.ics.uci.edu/ml/datasets/Statlog+%28Heart%29>