Bayes Factor

Karla Alfaro, Ana B. Coronel

Foundations of Statistics

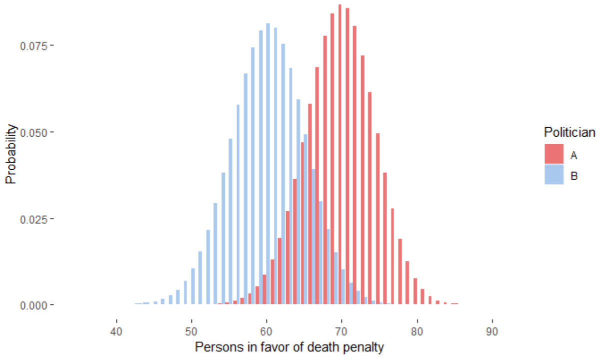
November 5, 2019

Introduction

During 1935 and 1961, Jeffreys developed the Bayesian approach to hypothesis testing by introducing statistical models to represent the probability of the data according to two different theories or beliefs. This is summarized in a quantity known as the Bayes Factor.

Comparing two simple hypothesis

- Suppose that there are two politicians who differ in their beliefs about whether the public supports the death penalty. Politician A believes that 70% of the population is in favor, while Politician B believes only 60% is.
- To test their beliefs, they organized a poll which asks 100 randomly selected people whether they support the death penalty. Based on this, the prior distribution can be defined as $Binomial(100, p_i)$ where $p_1 = 0.7$ and $p_2 = 0.6$ respectively for politician A and B.



Comparing two simple hypothesis

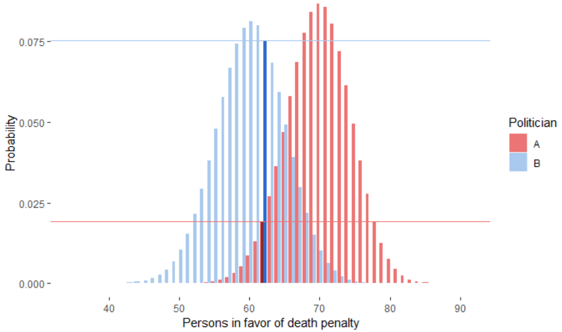
- The result of the poll is that 62 persons support the death penalty.
- It seems that the result supports *Politician B* hypothesis because the value is closer to his prediction. However, *Politician A* points out that he also predicted that 62 was possible.
- Then, in order to find out by how much is Politician B hypothesis favored, we compute the ratio between the probabilities of obtaining 62 according to their beliefs.

Comparing two simple hypothesis

- Politician B predicted that the probability of $y = 62$ was 0.0754 while politician A predicted that it was 0.0191.
- If we compute the ratio between the probabilities:

$$\frac{P(Y|H_B)}{P(Y|H_A)} = \frac{0.0754}{0.0191} = 3.94$$

- The observed data favors Politician B by a factor of 4.



Another example: Marfan syndrome

- Marfan Syndrome is a genetic disease of connective tissue that occurs in 1 out of every 15,000 people. The main ocular features of Marfan syndrome include biltareal lens dislocation, myopia and retinal detachment. About 70% of people with Marfan Syndrome have at least one of these ocular features while only 7% without it does.
- If a person has at least one of these ocular features, what are the odds that they have Marfan Syndrome?

Another example: Marfan syndrome

- Let H_0= *the person has Marfan syndrome*, H_1 = *the person does not have Marfan syndrome* and F=*the person has at least one ocular feature*
- Prior odds:

$$O(H_0) = \frac{P(H_0)}{P(H_1)} = \frac{1/15000}{14999/15000} = 0.000067$$

- Posterior odds:

$$O(H_0|F) = \frac{P(F|H_0)}{P(F|H_1)} = \frac{P(F|H_0)P(H_0)}{P(F|H_1)P(H_1)} \frac{(0.7)(1/15000)}{(0.07)(14999/15000)} = 0.00067$$

- The posterior odds are 10 times larger than the prior odds. In other words, having an ocular feature is a strong evidence in favor of H_0.

Bayes Factor

- Having developed some intuition, we can define the **Bayes Factor** as the relative evidence in the data. In other words, it means that the evidence in the data favors one hypothesis relative to another.
- How does the odds ratio change in the presence of data?

Bayes Factor

- Suppose we are interested in comparing two hypotheses H_0 and H_1 each of them with probability $P(H_0)$ and $P(H_1)$.
- We can define **prior odds** as the ratio that describes the degree to which we favor one hypothesis over another before we observe the data:

$$O(H_0) = \frac{P(H_0)}{P(H_1)}$$

.

- Now, we can define the **posterior odds** as the ratio that describes the degree to which we favor one hypothesis over another after we observe the data D:

$$O(H_0|D) = \frac{P(H_0|D)}{P(H_1|D)}$$

Bayes Factor

- The posterior odds can be decomposed as the product of the marginal likelihood ratio **(Bayes factor)** and the prior odds:

$$O(H_0|D) = \frac{P(H_0|D)}{P(H_1|D)} = \frac{P(D|H_0)}{P(D|H_1)} \cdot \frac{P(H_0)}{P(H_1)} = \frac{P(D|H_0)}{P(D|H_1)} \cdot O(H_0)$$

- Then, the Bayes factor is given by:

$$\frac{P(D|H_0)}{P(D|H_1)} = BF_{01}$$

Bayes Factor

- The Bayes Factor tells us whether the data provides evidence in favor or against a hypothesis and it can be interpreted as the weight of evidence provided by a set of data.

- The Bayes Factor compares the marginal likelihood of the data under competing models.
  - BF >1: the data provides evidence in favor of hypothesis H_0.
  - BF $= 1$: the data provides no evidence which means the prior and posterior odds are equal.
  - BF <1: the data provides evidence against the hypothesis H_0.

Strength of evidence

- Jeffreys suggested interpreting the BF_{01} in half units on the log_{10} scale. However, in order to have it on the same scale as the familiar deviance and likelihood ratio test statistics, it can be useful to consider $2 * log_{10}(BF_{01})$ and consider the following:

BF	Strength of evidence
1-3	weak evidence
3-20	positive evidence
20-150	strong evidence
$>$150	very strong evidence

Symmetry

- Bayes factors is inherit symmetric.
- If we wanted to assess the weight of evidence in factor of H_1 over H_0, the equation could be adjusted by taking reciprocals.
- Direction is important when talking about the Bayes factor.

Uncertainty in the parameters

- What happens if the hypotheses do not take a specific value, but instead follow a distribution?
- In the politicians example, instead of believing that the true proportion is exactly 60%, Politician B now believes that it is "around" 60%. The same goes for Politician A.
- To compute $P(D|H_B)$, we need to compute the marginal likelihood which is given by the multiplication of the probability that y would occur given θ times the probability of θ ocurring under hypothesis H_B.

$$P(D|H_B) = \int_\theta P(y|\theta) P_B(\theta) d\theta$$

Comparing two models

- Consider a simple model of a coin flipping experiment so that $Y_i \sim Ber(\theta)$ and the hypotheses are:
- H_0:

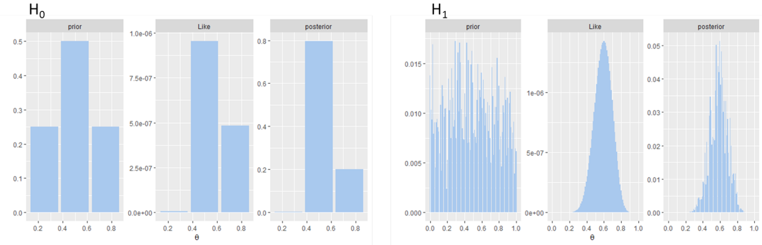
θ	P(θ)
0.25	0.25
0.5	0.5
0.75	0.25

- H_1:

$$H_1 : \theta \neq 0.5 \quad and \quad \theta | H_1 \sim Be(a, b)$$

Comparing two models

- The prior, likelihood and posterior distributions under the two hypotheses can be plotted as following:



Comparing two models

- Then,

$$BF_{01} = \frac{\sum_\theta p(y|\theta)p(\theta)}{\int_\theta p(y|\theta)p(\theta)d\theta} = \frac{0.5^n}{\int_0^1 \theta^{n\bar{y}}(1-\theta)^{n(1-\bar{y})}\frac{\theta^{(\alpha-1)}(1-\theta)^{(\beta-1)}}{Beta(a,b)}d\theta}$$

- Suppose, 100 coins were tossed and the number of heads were as following:

heads	BF_{01}	conclusion
50	3.7771	data supports H_0
20	1.3245	data supports H_0
18	0.6972	data supports H_1

Comparing two models

- How does the Bayes factor change between the simple and complex model?
- The complex model can take more values of θ. Then, if we observe that heads occur 10% of the time, we know that the complex hypothesis will be favored. Why? Because the simple model does not have a value of θ near the observed value of θ.
- However, if we observe that heads occur 50% of the time, we know that the simple hypothesis will be favored. Why? Because both models have the observed value of θ, but the simple model assings a greater probability to that value of θ.

Summary

- Bayes factors represent the weight of evidence in the data for competing hypotheses.
- The Bayes factor is the factor by which the relative beliefs must be multiplied after the data have been observed, providing new (posterior) relative beliefs.
- The Bayes factor can be directly interpreted without recourse to labels. However, some authors provide labels to help interpret evidence.

References

- BayesFactor: Software for Bayesian Inference: What is Bayes Factor? available at http:bayesfactor.blogspot.com201402the-bayesfactor-package-this-blog-is.htmlm=1
- Kass, R., Raftery, A. Bayes Factor, Journal of the American Statistical Association available at http:www.andrew.cmu.eduuserkk3nsimplicity /KassRaftery1995.pdf
- Orloff, J., Bloom. J. Bayesian Updating: Odds, MIT OpenCourseWare available at https:ocw.mit.educoursesmathematics18-05-introduction-to-probability-and-statistics-spring-2014readingsMIT18_05S14_Reading12b.pdf