Introduction to Bayes factors

Jeffrey R. Stevens

https://osf.io/h38sx/

9 May 2019

Disclaimers

1. I am not a Bayesian statistics expert.
2. This introduction is not meant to accurately reflect the data analysis process; it is only meant to illustrate general techniques for calculating Bayes factors.
3. Please read the software documentation and Bayes factor literature to learn how to properly implement these techniques.

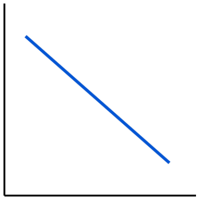
Questions

1. What is a Bayes factor?
2. Why use Bayes factors?
3. When to use Bayes factors?
4. How to present Bayes factors?
5. How to calculate Bayes factors?

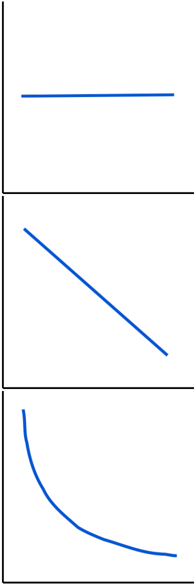
Section 1

What is a Bayes factor?

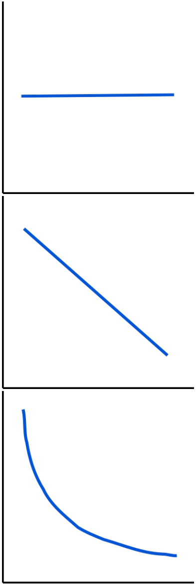
What are we as scientists trying to do?



What are we as scientists trying to do?



What are we as scientists trying to do?



x	y	z
175	62	M
162	48	F
186	82	M
123	71	F
190	60	M
185	51	M
172	68	F
159	55	M
123	51	F
182	88	M
148	45	F

The p-value

What is a p-value?

[1] (Wasserstein & Lazar, 2016, *American Statistician*)

The p-value

What is a p-value?

- The probability under a specified statistical model that a statistical summary of the data would be equal to or more extreme than its observed value[1]

[1] (Wasserstein & Lazar, 2016, *American Statistician*)

The p-value

What is a p-value?

- The probability under a specified statistical model that a statistical summary of the data would be equal to or more extreme than its observed value[1]
- Probability of obtaining results *at least as extreme as* those observed given that the null hypothesis is true

[1] (Wasserstein & Lazar, 2016, *American Statistician*)

The p-value

What is a p-value?

- The probability under a specified statistical model that a statistical summary of the data would be equal to or more extreme than its observed value[1]
- Probability of obtaining results *at least as extreme as* those observed given that the null hypothesis is true
  - Based on hypothetical outcomes more extreme than the ones observed

[1] (Wasserstein & Lazar, 2016, *American Statistician*)

The p-value

What is a p-value?

- The probability under a specified statistical model that a statistical summary of the data would be equal to or more extreme than its observed value[1]
- Probability of obtaining results *at least as extreme as* those observed given that the null hypothesis is true
  - Based on hypothetical outcomes more extreme than the ones observed
  - Gives probability of data given null hypothesis $P(data|hypothesis)$ rather than probability of alternative hypothesis given data $P(hypothesis|data)$

[1] (Wasserstein & Lazar, 2016, *American Statistician*)

The p-value

"You might just as well say," added the March Hare, "that 'I like what I get' is the same thing as 'I get what I like'!"
(Lewis Carroll, Alice in Wonderland)

$$P(data|hypothesis) \neq P(hypothesis|data)$$

The p-value

"You might just as well say," added the March Hare, "that 'I like what I get' is the same thing as 'I get what I like'!"
(Lewis Carroll, Alice in Wonderland)

$$P(data|hypothesis) \neq P(hypothesis|data)$$

The p-value

"You might just as well say," added the March Hare, "that 'I like what I get' is the same thing as 'I get what I like'!"
(Lewis Carroll, Alice in Wonderland)

$$P(data|hypothesis) \neq P(hypothesis|data)$$

$\neq$

Bayesian inference

Based on Bayes theorem

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

Bayesian inference

Based on Bayes theorem

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

- Bayesian parameter estimation
- Bayesian hypothesis testing

Bayesian hypothesis testing

Assessing the relative plausibilities of competing hypotheses H_0 and H_1:

$$\underbrace{\frac{P(H_1|D)}{P(H_0|D)}}_{\text{Posterior odds}} = \underbrace{\frac{P(H_1)}{P(H_0)}}_{\text{Prior odds}} \times \underbrace{\frac{P(D|H_1)}{P(D|H_0)}}_{\substack{\text{Likelihood ratio}\\ \text{(Bayes factor)}}}$$

[2] (Wagenmakers, 2007, *Psychonomic Bulletin & Review*)

Bayesian hypothesis testing

Assessing the relative plausibilities of competing hypotheses H_0 and H_1:

$$\underbrace{\frac{P(H_1|D)}{P(H_0|D)}}_{\text{Posterior odds}} = \underbrace{\frac{P(H_1)}{P(H_0)}}_{\text{Prior odds}} \times \underbrace{\frac{P(D|H_1)}{P(D|H_0)}}_{\substack{\text{Likelihood ratio} \\ \text{(Bayes factor)}}}$$

Posterior odds are the relative plausibility of models after observing data. The change from prior to posterior odds brought about by the data is the **Bayes factor**.[2]

[2] (Wagenmakers, 2007, *Psychonomic Bulletin & Review*)

The Bayes factor

- Extent to which the data sway our relative belief from one hypothesis to the other[3]

[3] (Wagenmakers, 2007, *Psychonomic Bulletin & Review*; Etz et al., 2018, *Psychonomic Bulletin & Review*)
[4] (Wagenmakers, 2007, *Psychonomic Bulletin & Review*)
[5] (Wagenmakers, 2007, *Psychonomic Bulletin & Review*; Rouder et al., 2018, *Psychonomic Bulletin & Review*)

The Bayes factor

- Extent to which the data sway our relative belief from one hypothesis to the other[3]
- Strength of evidence from data about the hypotheses[4]

[3] (Wagenmakers, 2007, *Psychonomic Bulletin & Review*; Etz et al., 2018, *Psychonomic Bulletin & Review*)

[4] (Wagenmakers, 2007, *Psychonomic Bulletin & Review*)

[5] (Wagenmakers, 2007, *Psychonomic Bulletin & Review*; Rouder et al., 2018, *Psychonomic Bulletin & Review*)

The Bayes factor

- Extent to which the data sway our relative belief from one hypothesis to the other[3]
- Strength of evidence from data about the hypotheses[4]
- Relative predictive accuracy of one hypothesis over another[5]

[3] (Wagenmakers, 2007, *Psychonomic Bulletin & Review*; Etz et al., 2018, *Psychonomic Bulletin & Review*)

[4] (Wagenmakers, 2007, *Psychonomic Bulletin & Review*)

[5] (Wagenmakers, 2007, *Psychonomic Bulletin & Review*; Rouder et al., 2018, *Psychonomic Bulletin & Review*)

The Bayes factor

- BF_{10} is the Bayes factor giving the evidence for H_1 over H_0

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

The Bayes factor

- BF_{10} is the Bayes factor giving the evidence for H_1 over H_0

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

- Whereas BF_{01} is evidence for H_0 over H_1

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

The Bayes factor

- BF_{10} is the Bayes factor giving the evidence for H_1 over H_0

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

- Whereas BF_{01} is evidence for H_0 over H_1

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

- $BF_{10} = 1/BF_{01}$

The Bayes factor

- BF_{10} is the Bayes factor giving the evidence for H_1 over H_0

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

- Whereas BF_{01} is evidence for H_0 over H_1

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

- $BF_{10} = 1/BF_{01}$
- BF_{10}=8 means that there is 8 times more evidence for H_1 than H_0 (and is equivalent to a $BF_{01} = 1/8 = 0.125$)

The Bayes factor

- BF_{10} is the Bayes factor giving the evidence for H_1 over H_0

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

- Whereas BF_{01} is evidence for H_0 over H_1

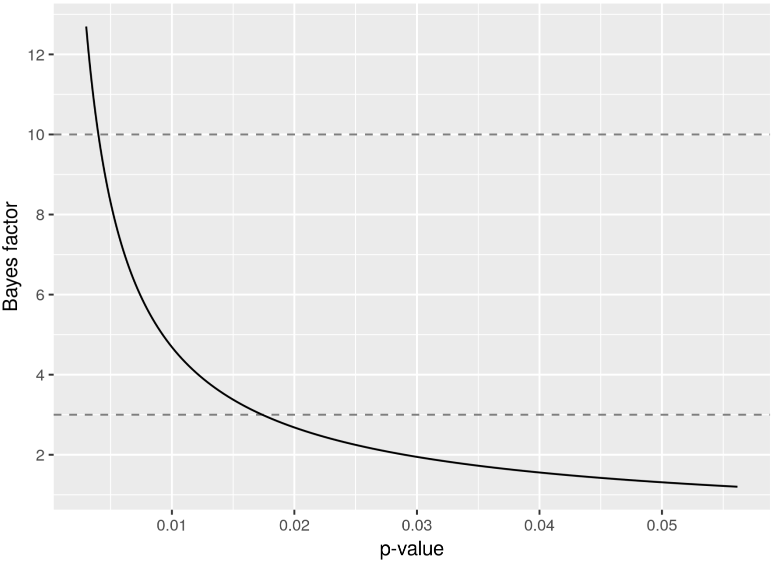
$$\frac{P(D|H_0)}{P(D|H_1)}$$

- $BF_{10} = 1/BF_{01}$
- BF_{10}=8 means that there is 8 times more evidence for H_1 than H_0 (and is equivalent to a $BF_{01} = 1/8 = 0.125$)
- Bayes factors range from 0 to ∞

Bayes factor cutoffs

BF_{10}	Interpretation
> 100	Extreme evidence for H_1
30 - 100	Very strong evidence for H_1
10 - 30	Strong evidence for H_1
3 - 10	Moderate evidence for H_1
1 - 3	Anecdotal evidence for H_1
1	Equal evidence for H_1 and H_0
$1/3$ - 1	Anecdotal evidence for H_0
$1/10$ - $1/3$	Moderate evidence for H_0
$1/30$ - $1/10$	Strong evidence for H_0
$1/100$ - $1/30$	Very strong evidence for H_0
$< 1/100$	Extreme evidence for H_0

Relationship between Bayes factors and p-values



Take home

Bayes factors provide the relative evidence for one model/hypothesis over another

Section 2

Why use Bayes factors?

Advantages

Can quantify evidence for H_1 vs. H_0[6]

- With frequentist statistics, data that are unlikely under H_0 may lead to its rejection, even though these data are just as unlikely under H_1.

[6] (Wagenmakers et al., 2018, *Psychonomic Bulletin & Review*)

Advantages

Can quantify evidence for H_1 vs. H_0[6]

- With frequentist statistics, data that are unlikely under H_0 may lead to its rejection, even though these data are just as unlikely under H_1.

- Bayes factors give relative evidence for H_1.

[6] (Wagenmakers et al., 2018, *Psychonomic Bulletin & Review*)

Advantages

Can quantify evidence in favor of H_0[7]

- With frequentist statistics, you **cannot accept the null hypothesis**.

[7] (Wagenmakers, 2007, *Psychonomic Bulletin & Review*)

Advantages

Can quantify evidence in favor of H_0[7]

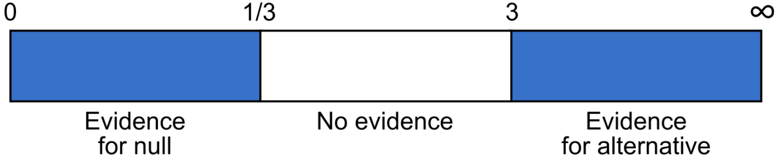
- With frequentist statistics, you **cannot accept the null hypothesis**.
- Bayes factors allow you to directly test the null hypothesis (relative to models under consideration).
  - Sometimes you want to test the null
  - Sometimes your alternative is not supported

[7] (Wagenmakers, 2007, *Psychonomic Bulletin & Review*)

Advantages

Can quantify evidence in favor of H_0[7]

- With frequentist statistics, you **cannot accept the null hypothesis**.
- Bayes factors allow you to directly test the null hypothesis (relative to models under consideration).
  - Sometimes you want to test the null
  - Sometimes your alternative is not supported



[7] (Wagenmakers, 2007, *Psychonomic Bulletin & Review*)

Advantages

Allows evidence to be monitored as data accumulate[8]

[8] (Rouder, 2014, *Psychonomic Bulletin & Review*)

[9] (Simmons et al., 2011, *Psychological Science*)

Advantages

Allows evidence to be monitored as data accumulate[8]

- With frequentist statistics, you must collect data on exactly the sample size established before data collection
  - Cannot stop early or collect more data after analyzing it[9]

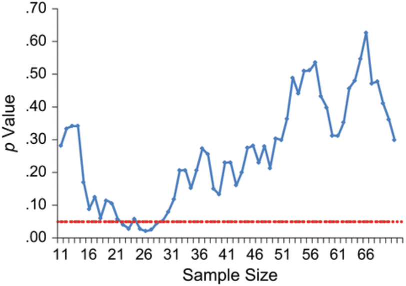
[8] (Rouder, 2014, *Psychonomic Bulletin & Review*)

[9] (Simmons et al., 2011, *Psychological Science*)

Advantages

Allows evidence to be monitored as data accumulate[8]

- With frequentist statistics, you must collect data on exactly the sample size established before data collection
  - Cannot stop early or collect more data after analyzing it[9]



8 (Rouder, 2014, *Psychonomic Bulletin & Review*)

9 (Simmons et al., 2011, *Psychological Science*)

Advantages

Allows evidence to be monitored as data accumulate[10]

- With Bayesian statistics, you can compute sequential Bayes factors[11]
  - Check priors
  - Set pre-determined BF thresholds (e.g., 10 or 1/10)
  - Start with at least 20 samples per group
  - Adjust based on prior knowledge

10 (Rouder, 2014, *Psychonomic Bulletin & Review*)
11 (Schönbrodt et al., 2017, *Psychological Methods*)
12 (Schönbrodt & Wagenmakers, 2018, *Psychonomic Bulletin & Review*)

Advantages

Allows evidence to be monitored as data accumulate[10]

- With Bayesian statistics, you can compute sequential Bayes factors[11]
  - Check priors
  - Set pre-determined BF thresholds (e.g., 10 or 1/10)
  - Start with at least 20 samples per group
  - Adjust based on prior knowledge
- It is possible to conduct something like a Bayesian power analysis[12]

[10] (Rouder, 2014, *Psychonomic Bulletin & Review*)
[11] (Schönbrodt et al., 2017, *Psychological Methods*)
[12] (Schönbrodt & Wagenmakers, 2018, *Psychonomic Bulletin & Review*)

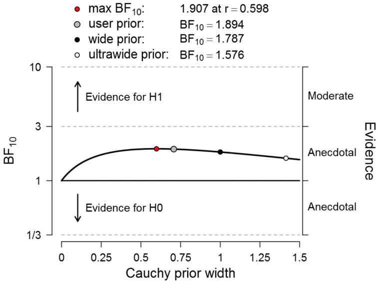
Potential issues

Potential issues

- Rely on same assumptions as frequentist statistics

Potential issues

- Rely on same assumptions as frequentist statistics
- Results can be sensitive to priors



Potential issues

- Rely on same assumptions as frequentist statistics
- Results can be sensitive to priors
- Still use cutoffs

Potential issues

- Rely on same assumptions as frequentist statistics
- Results can be sensitive to priors
- Still use cutoffs
- Not accepted in the field

Potential issues

- Rely on same assumptions as frequentist statistics
- Results can be sensitive to priors
- Still use cutoffs
- Not accepted in the field
- Multiple comparisons

Potential issues

- Rely on same assumptions as frequentist statistics
- Results can be sensitive to priors
- Still use cutoffs
- Not accepted in the field
- Multiple comparisons
- What if they conflict with p-values?

Take home

Bayes factors compare multiple hypotheses, can test null hypotheses, and allow for optional stopping

Section 3

When to use Bayes factors?

Testing hypotheses of no effect

- Binomial tests
- Contingency tables
- T-tests
- ANOVAs
- ANCOVAs
- Correlations
- Linear regressions

Model selection

Let's say you are interested in comparing two models in a regression: H_1 and H_2. First, you will compare them to an intercept-only model H_0 to generate their Bayes factors:

$BF_{10} = \frac{P(D|H_1)}{P(D|H_0)}$ and $BF_{20} = \frac{P(D|H_2)}{P(D|H_0)}$

Model selection

Let's say you are interested in comparing two models in a regression: H_1 and H_2. First, you will compare them to an intercept-only model H_0 to generate their Bayes factors:

$$BF_{10} = \frac{P(D|H_1)}{P(D|H_0)} \text{ and } BF_{20} = \frac{P(D|H_2)}{P(D|H_0)}$$

Then, you can divide the Bayes factors of two models (compared to null) to find the Bayes factor comparing those two models:

$$\frac{BF_{10}}{BF_{20}} = \frac{P(D|H_1)}{P(D|H_2)} = BF_{12}$$

Model selection

Likelihood ratio test

Model	BF
Condition 1	2
Condition 2	1
Condition 1 + Condition 2	14
Condition 1 * Condition 2	16

Model selection

Likelihood ratio test

Model	BF
Condition 1	2
Condition 2	1
Condition 1 + Condition 2	14
Condition 1 * Condition 2	16

$BF_{interaction} = \frac{16}{14} = 1.14$

Approximation from GLMs

- Currently, the `BayesFactor` package does not run on Generalized Linear Models (GLMs).
- But there is an approximation for Bayes factors using the BIC calculated by GLMs[13]

$$BF = e^{\frac{BIC_{null} - BIC_{alternative}}{2}}$$

13 (Wagenmakers, 2007, *Psychonomic Bulletin & Review*)

Estimating from existing test statistics

- You can calculate Bayes factors from existing analyses (e.g., if the full data are not available) with test statistics and sample sizes
  - Binomial tests
  - T-tests
  - One-way ANOVAs
  - Correlations
  - Regressions

Take home

Bayes factors can be calculated directly for standard t-tests, ANOVAs, ANCOVAs, correlations, and linear regressions and can be estimated for other situations

Section 4

How to present Bayes factors?

Presenting Bayes factors

- Methods
  - Define Bayes factor (I cite Wagenmakers, 2007)
  - Describe cutoffs for evidence (I cite Wagenmakers et al., 2018)
  - Describe priors/assumptions (I cite journal articles cited by packages)

[14] (van Doorn et al., 2019, PsyArXiv)

Presenting Bayes factors

- Methods
  - Define Bayes factor (I cite Wagenmakers, 2007)
  - Describe cutoffs for evidence (I cite Wagenmakers et al., 2018)
  - Describe priors/assumptions (I cite journal articles cited by packages)
- Results
  - Clarify direction (alternative/null)
  - Describe effect in "levels of evidence" terms
    - "There is moderate evidence for a difference between . . . "
    - "There is very strong evidence for no difference between . . . "
    - "There is no evidence for a difference between . . . "
  - Give Bayes factors as you would p-values (use 1-2 decimal places for >1; 2-3 decimal for <1; cutoffs for >100, <0.01)

[14] (van Doorn et al., 2019, PsyArXiv)

Presenting Bayes factors

- Methods
  - Define Bayes factor (I cite Wagenmakers, 2007)
  - Describe cutoffs for evidence (I cite Wagenmakers et al., 2018)
  - Describe priors/assumptions (I cite journal articles cited by packages)
- Results
  - Clarify direction (alternative/null)
  - Describe effect in "levels of evidence" terms
    - "There is moderate evidence for a difference between ..."
    - "There is very strong evidence for no difference between ..."
    - "There is no evidence for a difference between ..."
  - Give Bayes factors as you would p-values (use 1-2 decimal places for > 1; 2-3 decimal for < 1; cutoffs for > 100, < 0.01)
- Supplementary materials
  - Assumption checks
  - JASP produces robustness plots
- JASP Guidelines for Conducting and Reporting a Bayesian Analysis[14]

[14] (van Doorn et al., 2019, PsyArXiv)

Take home

Introduce Bayes factors in Methods, provide values in Results, cover your bases in Supplementary Materials

Section 5

How to calculate Bayes factors?

Take home

With both JASP and R, Bayes factors are as easy as frequentist—and you can convert between them!

Section 6

Resources

References

- Etz, A., Gronau, Q. F., Dablander, F., Edelsbrunner, P. A., & Baribault, B. (2018). How to become a Bayesian in eight easy steps: An annotated reading list. *Psychonomic Bulletin & Review*, 25(1), 219–234. https://doi.org/10.3758/s13423-017-1317-5
- Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic Bulletin & Review*, 21(2), 301–308. https://doi.org/10.3758/s13423-014-0595-4
- Rouder, J. N., Haaf, J. M., & Vandekerckhove, J. (2018). Bayesian inference for psychology, part IV: parameter estimation and Bayes factors. *Psychonomic Bulletin & Review*, 25(1), 102–113. https://doi.org/10.3758/s13423-017-1420-7
- Schönbrodt, F. D., & Wagenmakers, E.-J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*, 25(1), 128–142. https://doi.org/10.3758/s13423-017-1230-y
- Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods*, 22(2), 322–339. https://doi.org/10.1037/met0000061
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359–1366. https://doi.org/10.1177/0956797611417632
- van Doorn, J., van den Bergh, D., Bohm, U., Dablander, F., Derks, K., Draws, T., . . . Wagenmakers, E.-J. (2019). The JASP guidelines for conducting and reporting a Bayesian analysis. *PsyArXiv*. https://doi.org/10.31234/osf.io/yqxfr
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, 14(5), 779–804. https://doi.org/10.3758/BF03194105
- Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., . . . Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, 25(1), 35–57. https://doi.org/10.3758/s13423-017-1343-3
- Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose. *American Statistician*, 70(2), 129–133. https://doi.org/10.1080/00031305.2016.1154108

Resources

- *Psychonomic Bulletin & Review* special issue on Bayesian methods for advancing psychological science
- `BayesFactor` package
- JASP statistical software
- Bayesian Spectacles blog
- Understanding Bayes blog
- Anderson, D. R. (2008). Model Based Inference in the Life Sciences. New York: Springer.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). Bayesian Data Analysis (Third Edition). Boca Raton, FL: CRC Press.
- Kruschke, J. K. (2015). Doing Bayesian Data Analysis (Second Edition). New York: Academic Press.