Lecture 3 - Tools for computing Bayes factors

Recall: Bayes' theorem gives us:

$$\frac{P(\mathcal{H}_0 \mid \text{data})}{P(\mathcal{H}_1 \mid \text{data})} = \frac{p(\mathcal{H}_0)}{p(\mathcal{H}_1)} \times \frac{p(\text{data} \mid \mathcal{H}_0)}{p(\text{data} \mid \mathcal{H}_1)}$$

or

Posterior odds = Prior odds × Bayes factor

This gives two conceptual definition for Bayes factor:

1) the factor by which the observed data is more likely under one model / hypothesis compared to the other.

2) the factor by which the prior odds between models is updated after observing data.

An aside – what are odds anyway?

↳ odds = ratio of probabilities

↳ 1:1 $\longrightarrow$ $\dfrac{p(\mathcal{H}_0)}{p(\mathcal{H}_1)} = 1$ (equally likely)

↳ 3:1 $\longrightarrow$ $\dfrac{p(\mathcal{H}_0)}{p(\mathcal{H}_1)} = 3$ ($\mathcal{H}_0$ three times more likely than $\mathcal{H}_1$, a priori)

Ex: Suppose prior odds = 1:1

and observed data gives $BF_{01} = 4$.

Then posterior odds = prior odds × Bayes factor

$= 1 \times 4$

$= 4.$

Thus, after observing data, $\mathcal{H}_0$ is four times as likely as $\mathcal{H}_1$ [4:1]

↳ how do we get from posterior odds to posterior probability?

By above (posterior odds for $\mathcal{H}_0 = 4$)

$$\hookrightarrow \quad \frac{p(\mathcal{H}_0 \mid \text{data})}{p(\mathcal{H}_1 \mid \text{data})} = 4 \qquad (*)$$

Since $\mathcal{H}_0$ & $\mathcal{H}_1$ are the only two possible models, we know that the posterior probabilities must add to 1

$$\hookrightarrow \quad p(\mathcal{H}_0 \mid \text{data}) + p(\mathcal{H}_1 \mid \text{data}) = 1$$

$$\hookrightarrow \quad p(\mathcal{H}_1 \mid \text{data}) = 1 - p(\mathcal{H}_0 \mid \text{data})$$

Sub this into (*)

$$\frac{p(\mathcal{H}_0 \mid \text{data})}{1 - p(\mathcal{H}_0 \mid \text{data})} = 4$$

$$\longrightarrow \quad p(\mathcal{H}_0 \mid \text{data}) = 4 - 4 \cdot p(\mathcal{H}_0 \mid \text{data})$$

$$\longrightarrow \quad p(\mathcal{H}_0 \mid \text{data}) + 4 \cdot p(\mathcal{H}_0 \mid \text{data}) = 4$$

$$\longrightarrow \quad p(\mathcal{H}_0 \mid \text{data}) \left[1 + 4 \right] = 4$$

Thus, we have $p(\mathcal{H}_0 \mid \text{data}) = \frac{4}{1+4} = \frac{4}{5} = 0.80$

Note: in general, the above procedure can be used to prove the following:

$$\boxed{\text{posterior probability} = \frac{\text{posterior odds}}{1 + \text{posterior odds}}}$$

Our goal today - computing Bayes factors

Mathematically, need to compute

$$p(\text{data} \mid \mathcal{H})$$

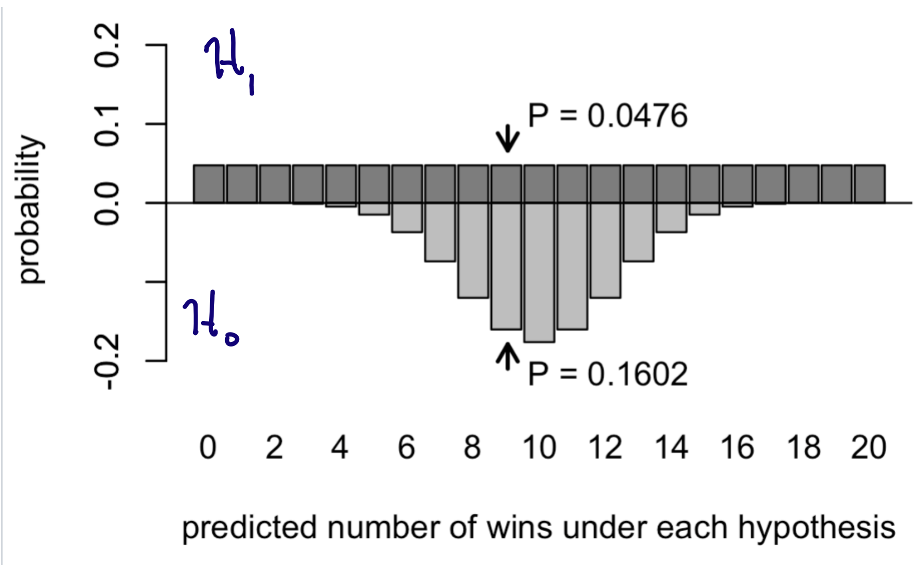
for both models ($\mathcal{H}_0$ and $\mathcal{H}_1$).

In general, this is hard!

One method – compute the "predictive distribution" for each model ($\mathcal{H}_0$, $\mathcal{H}_1$) – these show the data we would expect for each model, weighted by the prior probability for all values of w (i.e., the prior on w under $\mathcal{H}_1$)

(see Etz, Haaf, Rouder, & Vandekerckhove, 2018, for details)

Example 1 – uniform prior



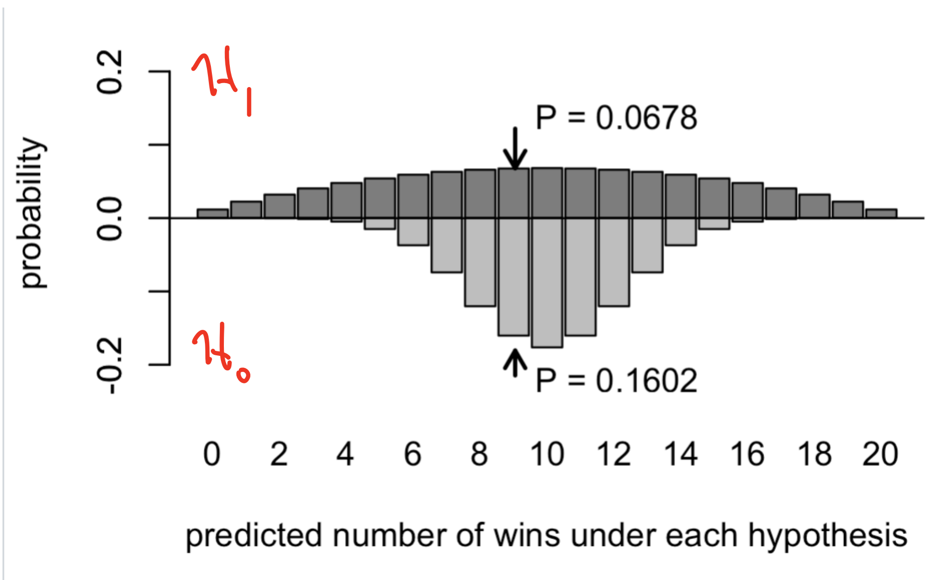
$$BF_{01} = \frac{p(\text{data} \mid \mathcal{H}_0)}{p(\text{data} \mid \mathcal{H}_1)}$$

$$= \frac{0.1602}{0.0476}$$

$$= 3.36$$

Interpretation: "the observed data are 3.36 times more likely under $\mathcal{H}_0$ than under $\mathcal{H}_1$."

Example 2 – Smooth peaked prior [beta (2,2) distribution]



$$BF_{01} = \frac{p(\text{data} \mid \mathcal{H}_0)}{p(\text{data} \mid \mathcal{H}_1)} = \frac{0.1602}{0.0678} = 2.36$$

Interpretation: "the observed data are 2.36 times more likely under $\mathcal{H}_0$ than under $\mathcal{H}_1$."

Downside to this approach:

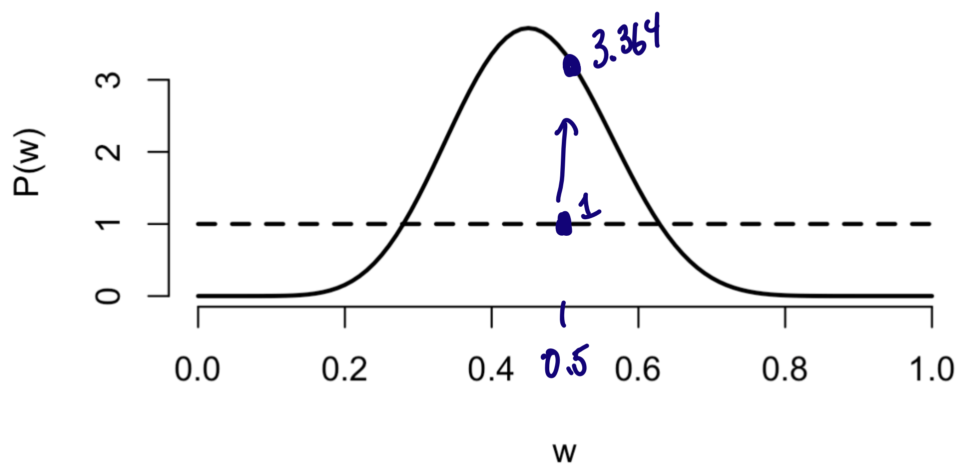
(1) can be computationally difficult

(2) the plot contains no obvious information about the parameter (ω).

[i.e, missing the posterior distribution]

An alternative approach:

* start by plotting the prior & posterior on same plot.



* find the y-value for $w = 0.5$ ($\mathcal{H}_0$) on both the prior and posterior curves

What do we notice:

(1) our belief in $\mathcal{H}_0$ ($w = 0.5$) increases from prior to posterior

(2) by how much? In R, we can compute them both:

$$\text{posterior} = p(w = 0.5 \mid \text{data}) = 3.364$$

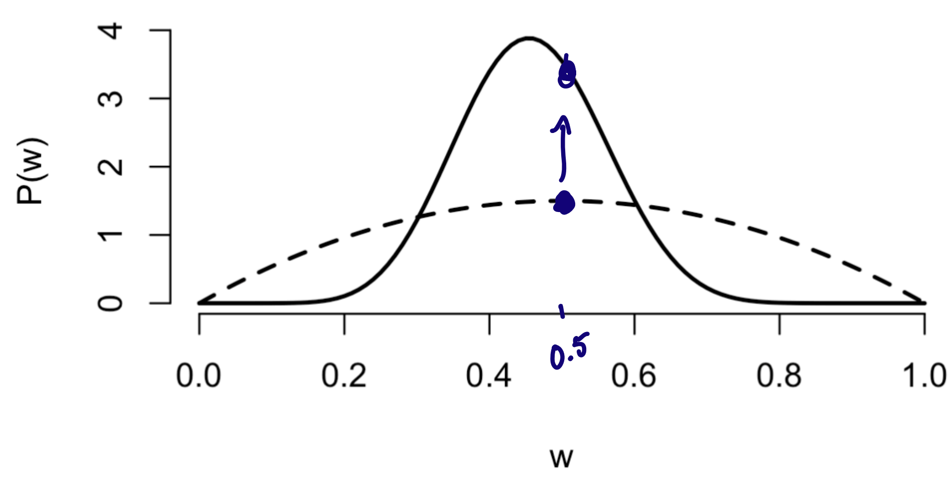
$$\text{prior} = p(w = 0.5) = 1$$

So $$\frac{\text{posterior y-value}}{\text{prior y-value}} = \frac{3.364}{1} = 3.364$$

BF_{01}

Lets try for the "smooth-peaked" prior [beta(2,2)]:

* start by plotting the prior & posterior on same plot.



* find the y-value for $w = 0.5$ (H_0) on both the prior and posterior curves

What do we notice:

(1) our belief in H_0 ($w = 0.5$) increases from prior to posterior

(2) by how much? In R, we can compute them both:

$$\text{posterior} = p(w = 0.5 \mid \text{data}) = 3.546$$

$$\text{prior} = p(w = 0.5) = 1.5$$

So $$\frac{\text{posterior y-value}}{\text{prior y-value}} = \frac{3.546}{1.5} = 2.364$$

BF_{01} again!

This pattern always holds for the types of model comparisons we usually do in the behavioral sciences:

↳ "Savage-Dickey Density Ratio."

Further reading: Wagenmakers et al. (2010)

ELSEVIER

Cognitive Psychology

Volume 60, Issue 3, May 2010, Pages 158-189

Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method

Eric-Jan Wagenmakers [a], Tom Lodewyckx [b], Himanshu Kuriyal [c], Raoul Grasman [a]

https://doi.org/10.1016/j.cogpsych.2009.12.001

Faulkenberry (2019)

https://doi.org/10.29220/CSAM.2019.26.2.217

A tutorial on generalizing the default Bayesian *t*-test via posterior sampling and encompassing priors

Communications for Statistical Applications and Methods 2019;26:217-238

Published online March 31, 2019

Thomas J. Faulkenberry

[a]Department of Psychological Sciences, Tarleton State University, USA

Correspondence to: [1]Department of Psychological Sciences, Tarleton State University, Stephenville, Texas, USA.
faulkenberry@tarleton.edu

Received December 10, 2018; **Revised** January 12, 2019; **Accepted** February 6, 2019.

Why do we care?

↳ the Savage-Dickey method is one of the primary visualization methods used in Jasp!

↳ let's try it.

* open Jasp, click the + symbol to open the "Summary Stats Module"

* select "Bayesian binomial test"

* in our example, we observed 9 successes out of 20 trials.

  ↳ Successes = 9

  Failures = 11

* since data are evidential for $\mathcal{H}_0$, click BF_{01}

* under "Plots", check "Prior and Posterior"

* try changing the prior to beta(2,2) (the smooth peaked prior)