Bayesian Methods in Mathematical Cognition – A Conceptual Overview

Tom Faulkenberry

Tarleton State University
Stephenville, TX, USA
https://tomfaulkenberry.github.io
Twitter: @tomfaulkenberry

MCLS Brown Bag Seminar

June 12, 2020

Outline:

- conceptual background of Bayesian inference
- priors on models vs. priors on parameters
- example using JASP, w/ reporting template.

* Faulkenberry, T.J., Ly, A., & Wagenmakers, E.J. (2020). Bayesian inference in numerical cognition: A tutorial using JASP. To appear in Journal of Numerical Cognition. https://psyarxiv.org/vg9pw

* These slides can be downloaded from https://tomfaulkenberry.github.io/talks.html

Suppose we are interested in assessing effectiveness of some mathematics instruction program.

After implementing the program, we measure mathematical ability using the Scale for Advancing Mathematical Ability (SAMA), a national assessment with a known mean score of 50.

Our sample (who had the training) had a mean of 54.4 and a standard deviation of 10.

Did the training work?

Basic statistical framework:

(1) define two competing models ("hypotheses") that could possibly generate our observed data.

- one where the training worked
- one where it didn't

(2) assess the fit of these models against the observed data.

How to proceed?

Define hypotheses about (population) effect size δ "delta"

$\mathcal{H}_0: \delta = 0$, $\mathcal{H}_1: \delta \neq 0$

Collect data

Compute:

$p(\text{data} \mid \mathcal{H}_0)$

"p-value"

Interpretation:

If p is small, data is rare under $\mathcal{H}_0$, so we reject $\mathcal{H}_0$ in favor of $\mathcal{H}_1$.

Compute:

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

"Bayes factor"

Interpretation:

if $BF_{01} > 1$, data more likely under $\mathcal{H}_0$.

if $BF_{01} < 1$, data more likely under $\mathcal{H}_1$.

p-value $= p(\text{data} \mid \mathcal{H}_0)$

1) only considers fit of $\mathcal{H}_0$ as a potential model for data
2) ignores fit of $\mathcal{H}_1$

Thus, "support" for $\mathcal{H}_1$ is only indirect

Bayes factor $= \dfrac{p(\text{data} \mid \mathcal{H}_0)}{p(\text{data} \mid \mathcal{H}_1)}$

1) considers relative adequacy of both models as predictors of data.
2) can directly index support for either $\mathcal{H}_0$ or $\mathcal{H}_1$.

Ex: $BF_{10} = 8 \longrightarrow$ "The observed data are 8 times more likely under $\mathcal{H}_1$ than $\mathcal{H}_0$".

$BF_{10} = \dfrac{1}{BF_{01}}$

Jeffreys (1961):

BF	Evidence*
1 – 3	anecdotal
3 – 10	moderate
10 – 30	strong
30 – 100	very strong
> 100	extreme

* these are only guidelines!

How does Bayes work?

Use data to update the plausibility of model/hypothesis $\mathcal{H}$.

$$\boxed{p(\mathcal{H} \mid \text{data}) = p(\mathcal{H}) \times \frac{p(\text{data} \mid \mathcal{H})}{p(\text{data})}}$$

posterior plausibility = prior plausibility × updating factor

applied to two models:

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

posterior odds = prior odds × Bayes factor

What do we mean by prior?

Two types of "priors":

1) priors on models

2) priors on parameters within a given model

① Priors on models — before observing data, what is relative likelihood of competing models?

- common default: $p(\mathcal{H}_0) = p(\mathcal{H}_1) = 1/2$

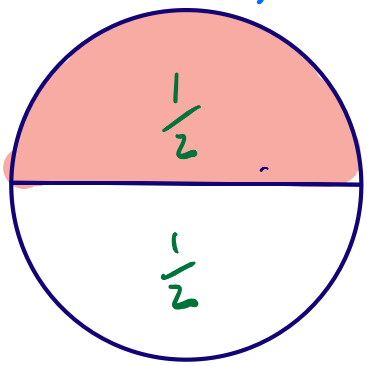
  ↳ i.e., "1-1 prior odds"*

$$\boxed{p(\mathcal{H}_0 \mid \text{data}) = \frac{BF_{01}}{BF_{01} + 1}}$$

* Note: if $p(\mathcal{H}_0) \neq p(\mathcal{H}_1)$, we can still compute the posterior probability. Ask me about this at the end!

Example:

$p(\mathcal{H}_0)$



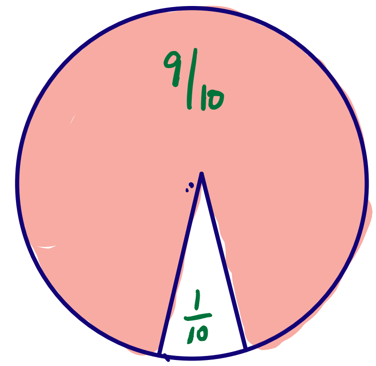
$p(\mathcal{H}_1)$

observe data

$$\xrightarrow{\hspace{3em}}$$

$BF_{01} = 9$

$p(\mathcal{H}_0 \mid data)$



$p(\mathcal{H}_1 \mid data)$

Prior odds = 1:1   Posterior odds = 9:1

4:1 $\longrightarrow$ 36:1

1:5   9:5

$$\boxed{\begin{aligned} * \quad p(\mathcal{H}_0 \mid data) &= \frac{BF_{01}}{BF_{01}+1} \\ &= \frac{9}{9+1} \\ &= 0.9 \end{aligned}}$$

② priors on parameters within a given model.

- model definitions : $\mathcal{H}_0 : \delta = 0$

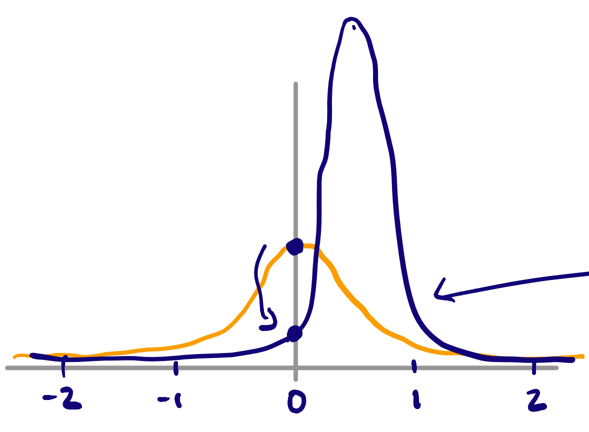
$\mathcal{H}_1 : \delta \neq 0$ ← what exactly do we mean here?

- we quantify our uncertainty about the effect size δ under $\mathcal{H}_1$ by placing a distribution on δ

- suppose we have no idea what to expect. An objective approach is to use a default prior on δ.

  - Jeffreys (1961) recommended a Cauchy distribution with scale $r = 1/\sqrt{2} \approx 0.707$

Prior



observe data

Posterior

ratio of these = BF_{01}

"Savage Dickey Density ratio"

Let's continue our working example. Recall that we tested $N = 65$ participants and observed a sample mean of 54.4 with SD = 10.

- compute $t = \frac{\bar{x} - \mu}{s/\sqrt{N}} = \frac{54.4 - 50}{10/\sqrt{65}} = 3.55$

- use JASP "Summary Statistics" module

Elements to report:

1. report results of hypothesis test

   - define H_0, H_1 and specify prior under H_1.

     "Under the null hypothesis we expect an effect size of 0. Thus, we define $H_0: \delta = 0$. The alternative hypothesis is two-sided, $H_1: \delta \neq 0$, and prior to observing data, we assumed that δ was distributed as a Cauchy distribution with scale $r = 0.707$."

   - report and interpret Bayes factor

     "We found a Bayes factor of $BF_{10} = 34.7$, which means that the observed data are approximately 35 times more likely under H_1 than H_0. This result indicates strong evidence in favor of H_1"

- (optional) calculate and report posterior model probability for preferred model.

  - from earlier,

$$p(\mathcal{H}_1 \mid \text{data}) = \frac{BF_{10}}{BF_{10} + 1}$$

$$= \frac{34.7}{34.7 + 1} = 0.972.$$

- "Assuming prior odds of 1-1 for $\mathcal{H}_1$ and $\mathcal{H}_0$, our observed data updated these odds to 34.7-to-1 in favor of $\mathcal{H}_1$. This is equivalent to a posterior model probability of $p(\mathcal{H}_1 \mid \text{data}) = 0.97$."

2. report results of parameter estimation
   - only if $\mathcal{H}_1$ is the preferred model!
   - specify parameter of interest and remind reader of prior under $\mathcal{H}_1$
     - "of interest is the posterior distribution for δ, the population-level effect size. Under $\mathcal{H}_1$, δ was assigned a Cauchy prior with scale $r = 0.707$."

- report the 95% credible interval.
  - "The posterior distribution for δ had a median of 0.422, with a central 95% credible interval that ranges from 0.171 to 0.675."

For more details, see:

- van Doorn et al. (2019). The JASP guidelines for conducting and reporting a Bayesian analysis. https://psyarxiv.org/yqxfr.
- Faulkenberry, T.J., Ly, A., & Wagenmakers, E.J. (2020). Bayesian inference in numerical cognition: A tutorial using JASP. To appear in Journal of Numerical Cognition. https://psyarxiv.org/vg9pw

Closing thoughts – some pragmatic advantages of Bayesian methods

1) can assess the degree to which data make an empirical claim more (or less) plausible.

   - no need to "accept" or "reject" hypotheses

2) can discriminate between "evidence of absence" and "absence of evidence"

3) can make direct & intuitive statements about the plausibility of models and parameters.

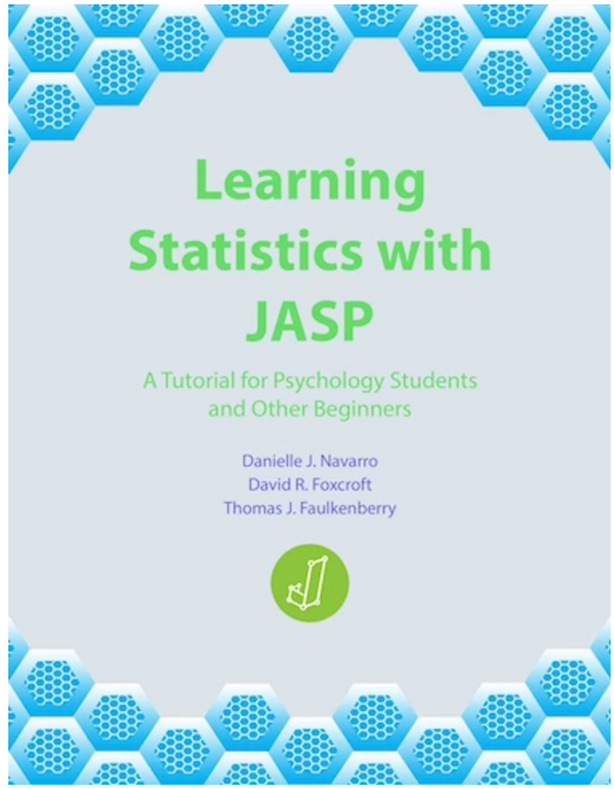
4) can update evidence as data accumulates

5) can incorporate expert knowledge about models & parameters.

Before you go – a couple of shameless plugs!

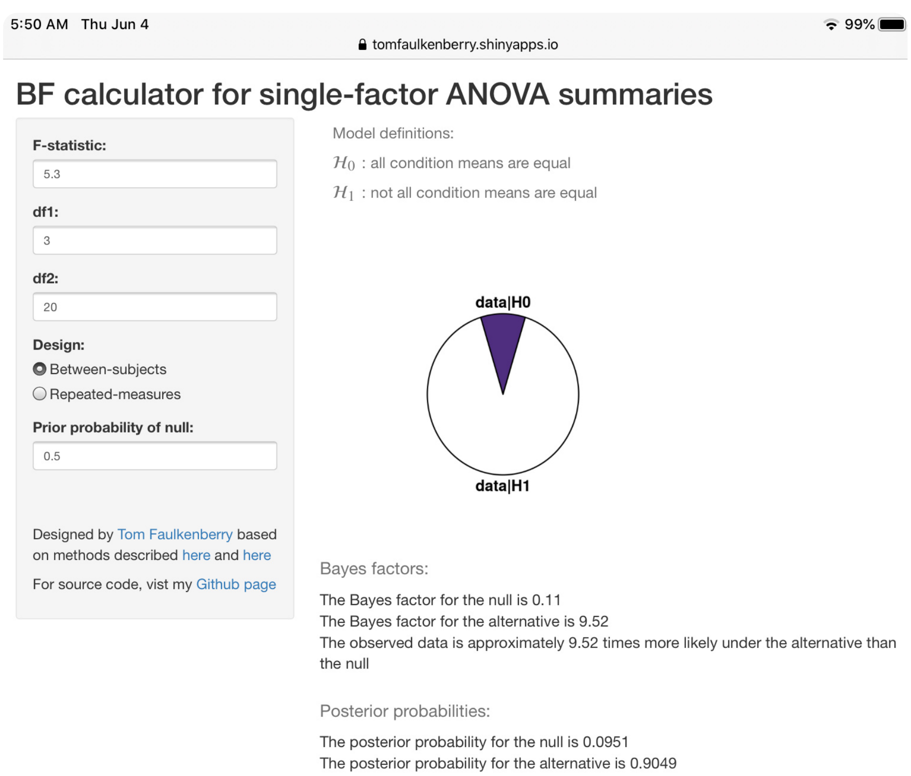
1) our FREE statistics textbook!

www.learnstatswithjasp.com



2) a free online Bayes factor calculator!

https://tomfaulkenberry.shinyapps.io/anovaBFcalc



* for more details, see Faulkenberry (2019), Advances in Methods & Practices in Psychological Science

Take home points:

- Bayes is easy, especially with the right software.
- Bayes answers the questions you thought you were asking
- testing or estimation? No need to choose — Bayes gives you both!

More questions — contact me!

Email: faulkenberry@tarleton.edu

Twitter: @tomfaulkenberry

Web: https://tomfaulkenberry.github.io

Extras:

1. suppose you have different prior odds (e.g., 4-to-1). Then posterior probability can be computed as:

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

2. Informed priors

- experts may express a different prior belief about δ
- if effect exists, what effect sizes do we expect?
    - example: the "Oosterwijk prior" (Gronau et al., 2020)
        - ask expert:

          median = 0.35

          33^{rd} percentile = 0.25

          66^{th} percentile = 0.45
        - put these in MATCH Uncertainty Elicitation Tool

          ↳ result = scaled noncentral t-distribution
            - location = 0.35
            - scale = 0.102
            - df = 3