The necessity and formulation of a robust (imprecise) Bayes Factor

Patrick Schwaferts

Ludwig-Maximilians-Universität München

01. August 2018

Introduction

Introduction

Reproducibility crisis in psychological research

Introduction

Reproducibility crisis in psychological research

Rise of popularity of Bayesian Statistics: promoted as being *the* solution

Reproducibility crisis in psychological research

Rise of popularity of Bayesian Statistics: promoted as being *the* solution

Bayes Factor for comparing two hypotheses ("Bayesian t-Test")

What is the Bayes Factor?

What is the Bayes Factor?

Informal:

A generalization of the Likelihood Ratio to include prior information.

From Likelihood Ratio to Bayes Factor I

From Likelihood Ratio to Bayes Factor I

Situation:

Two independent groups with observations x_i and y_j and model

$$X_i \sim N(\mu, \sigma^2), \quad i = 1, ..., n,$$
$$Y_j \sim N(\mu + \alpha, \sigma^2), \quad j = 1, ..., m,$$

with parameters μ, σ^2, $\delta = \alpha/\sigma$.

From Likelihood Ratio to Bayes Factor I

Situation:

Two independent groups with observations x_i and y_j and model

$$X_i \sim N(\mu, \sigma^2), \quad i = 1, ..., n,$$
$$Y_j \sim N(\mu + \alpha, \sigma^2), \quad j = 1, ..., m,$$

with parameters μ, σ^2, $\delta = \alpha/\sigma$.

Research Question:

Is there a difference between both groups?

From Likelihood Ratio to Bayes Factor I

Situation:

Two independent groups with observations x_i and y_j and model

$$X_i \sim N(\mu, \sigma^2), \quad i = 1, ..., n,$$
$$Y_j \sim N(\mu + \alpha, \sigma^2), \quad j = 1, ..., m,$$

with parameters μ, σ^2, $\delta = \alpha/\sigma$.

Research Question:

Is there a difference between both groups?

Hypotheses:

$$H_0 : \delta = 0 \quad \text{vs.} \quad H_1 : \delta \neq 0$$

From Likelihood Ratio to Bayes Factor II

Likelihood Ratio:

$$LR_{10} = \frac{\max\limits_{\mu,\sigma^2,\delta} \; f(data|\mu,\sigma^2,\delta)}{\max\limits_{\mu,\sigma^2} \; f(data|\mu,\sigma^2,\delta=0)}$$

From Likelihood Ratio to Bayes Factor II

Likelihood Ratio:

$$LR_{10} = \frac{\max\limits_{\mu,\sigma^2,\delta} \; f(data|\mu,\sigma^2,\delta)}{\max\limits_{\mu,\sigma^2} \; f(data|\mu,\sigma^2,\delta=0)}$$

Law of Likelihood:

The extent to which the data support one model over another ($:=$ *evidence*) is equal to the ratio of their likelihoods.

From Likelihood Ratio to Bayes Factor II

Likelihood Ratio:

$$LR_{10} = \frac{\max\limits_{\mu,\sigma^2,\delta} \; f(data|\mu,\sigma^2,\delta)}{\max\limits_{\mu,\sigma^2} \; f(data|\mu,\sigma^2,\delta=0)}$$

Law of Likelihood:

The extent to which the data support one model over another ($:=$ *evidence*) is equal to the ratio of their likelihoods.

Interpretation of LR:

The data is LR_{10} times as much evidence for the model chosen$^{(*)}$ under H_1 than for the model chosen$^{(*)}$ under H_0.

$^{(*)}$: *chosen* refers to the *max*-operation

$\Rightarrow$ LR_{10} quantifies the maximum evidence for H_1 (in a comparison with H_0)

From Likelihood Ratio to Bayes Factor III

Introducing Prior Probabilities:

$$P(H_1) \quad \text{and} \quad P(H_0)$$

From Likelihood Ratio to Bayes Factor III

Introducing Prior Probabilities:

$$P(H_1) \quad \text{and} \quad P(H_0)$$

Bayes Rule:

$$\underbrace{\frac{P(H_1|data)}{P(H_0|data)}}_{PosteriorOdds} = LR_{10} \cdot \underbrace{\frac{P(H_1)}{P(H_0)}}_{PriorOdds}$$

The data is used to learn about $P(H_1)$ and $P(H_0)$.

From Likelihood Ratio to Bayes Factor III

Introducing Prior Probabilities:

$$P(H_1) \quad \text{and} \quad P(H_0)$$

Bayes Rule:

$$\underbrace{\frac{P(H_1|data)}{P(H_0|data)}}_{PosteriorOdds} = LR_{10} \cdot \underbrace{\frac{P(H_1)}{P(H_0)}}_{PriorOdds}$$

The data is used to learn about $P(H_1)$ and $P(H_0)$.

Interpretation of Posterior Probabilities:

After seeing the data, the maximum belief in H_1 is $P(H_1|data)$.

From Likelihood Ratio to Bayes Factor IV

Introducing Parameter Priors:

$$P_\mu, \quad P_{\sigma^2} \quad \text{and} \quad P_\delta$$

Introducing Parameter Priors:

$$P_\mu, \quad P_{\sigma^2} \quad \text{and} \quad P_\delta$$

Bayesian Hypotheses:

$$H_0^B: \begin{array}{l} \mu \sim P_\mu \\ \sigma^2 \sim P_{\sigma^2} \\ \delta = 0 \end{array} \quad \text{vs.} \quad H_1^B: \begin{array}{l} \mu \sim P_\mu \\ \sigma^2 \sim P_{\sigma^2} \\ \delta \sim P_\delta \end{array}$$

P_δ is called *test-relevant prior*.

From Likelihood Ratio to Bayes Factor IV

Introducing Parameter Priors:

$$P_\mu, \quad P_{\sigma^2} \quad \text{and} \quad P_\delta$$

Bayesian Hypotheses:

$$H_0^B: \begin{array}{l} \mu \sim P_\mu \\ \sigma^2 \sim P_{\sigma^2} \\ \delta = 0 \end{array} \quad \text{vs.} \quad H_1^B: \begin{array}{l} \mu \sim P_\mu \\ \sigma^2 \sim P_{\sigma^2} \\ \delta \sim P_\delta \end{array}$$

P_δ is called *test-relevant prior*.

Marginalized Likelihoods:

$$m(data|H_1^B) = \iiint f(data|\mu, \sigma^2, \delta) P_\mu(\mu) P_{\sigma^2}(\sigma^2) P_\delta(\delta) d\delta d\sigma^2 d\mu$$

$$m(data|H_0^B) = \iint f(data|\mu, \sigma^2, \delta = 0) P_\mu(\mu) P_{\sigma^2}(\sigma^2) d\sigma^2 d\mu$$

From Likelihood Ratio to Bayes Factor V

Bayes Factor:

$$BF_{10} = \frac{m(data|H_1^B)}{m(data|H_0^B)}$$

From Likelihood Ratio to Bayes Factor V

Bayes Factor:

$$BF_{10} = \frac{m(data|H_1^B)}{m(data|H_0^B)}$$

Bayes Rule:

$$\frac{P(H_1^B|data)}{P(H_0^B|data)} = BF_{10} \cdot \frac{P(H_1^B)}{P(H_0^B)}$$

The data is used to learn about $P(H_1^B)$ and $P(H_0^B)$. Nothing can be learned about the parameter priors. $\Rightarrow P_\delta$ is part of the H_1^B-model.

Bayes Factor:

$$BF_{10} = \frac{m(data|H_1^B)}{m(data|H_0^B)}$$

Bayes Rule:

$$\frac{P(H_1^B|data)}{P(H_0^B|data)} = BF_{10} \cdot \frac{P(H_1^B)}{P(H_0^B)}$$

The data is used to learn about $P(H_1^B)$ and $P(H_0^B)$. Nothing can be learned about the parameter priors. $\Rightarrow P_\delta$ is part of the H_1^B-model.

Interpretation of BF:

The data is BF_{10} times as much evidence for the model behind $m(data|H_1^B)$ than for the model behind $m(data|H_0^B)$.

What is the model behind $m(data|H_1^B)$?

What is the model behind $m(data|H_1^B)$?

Again Bayes Rule:

$$P(H_1^B|data) = \frac{m(data|H_1^B) \cdot P(H_1)}{P(data)}$$

What is the model behind $m(data|H_1^B)$?

Again Bayes Rule:

$$P(H_1^B|data) = \frac{m(data|H_1^B) \cdot P(H_1)}{P(data)}$$

In order to apply Bayes Rule, $m(data|H_1^B)$ needs to be a likelihood, which describes the data-generating process.

What is the model behind $m(data|H_1^B)$?

Again Bayes Rule:

$$P(H_1^B|data) = \frac{m(data|H_1^B) \cdot P(H_1)}{P(data)}$$

In order to apply Bayes Rule, $m(data|H_1^B)$ needs to be a likelihood, which describes the data-generating process.

So the model behind $m(data|H_1^B)$ models a data-generating process with likelihood

$$m(data|H_1^B) = \iiint f(data|\mu, \sigma^2, \delta) P_\mu(\mu) P_{\sigma^2}(\sigma^2) P_\delta(\delta) d\delta d\sigma^2 d\mu$$

What is the model behind $m(data|H_1^B)$?

Again Bayes Rule:

$$P(H_1^B|data) = \frac{m(data|H_1^B) \cdot P(H_1)}{P(data)}$$

In order to apply Bayes Rule, $m(data|H_1^B)$ needs to be a likelihood, which describes the data-generating process.

So the model behind $m(data|H_1^B)$ models a data-generating process with likelihood

$$m(data|H_1^B) = \iiint f(data|\mu, \sigma^2, \delta) P_\mu(\mu) P_{\sigma^2}(\sigma^2) P_\delta(\delta) d\delta d\sigma^2 d\mu$$

$\Rightarrow$ A model with subjective components!

What is the model behind $m(data|H_1^B)$?

Again Bayes Rule:

$$P(H_1^B|data) = \frac{m(data|H_1^B) \cdot P(H_1)}{P(data)}$$

In order to apply Bayes Rule, $m(data|H_1^B)$ needs to be a likelihood, which describes the data-generating process.

So the model behind $m(data|H_1^B)$ models a data-generating process with likelihood

$$m(data|H_1^B) = \iiint f(data|\mu, \sigma^2, \delta) P_\mu(\mu) P_{\sigma^2}(\sigma^2) P_\delta(\delta) d\delta d\sigma^2 d\mu$$

$\Rightarrow$ A model with subjective components!

[The Bayes Factor does not directly answer: *Is there an effect?*]

Necessity of properly specifying P_δ

Necessity of properly specifying P_δ

The test-relevant prior P_δ is part of the H_1^B-model.

Necessity of properly specifying P_δ

The test-relevant prior P_δ is part of the H_1^B-model.

P_δ need to be specified properly.

Necessity of properly specifying P_δ

The test-relevant prior P_δ is part of the H_1^B-model.

P_δ need to be specified properly.

If not: H_1^B-model misspecifies the experimental situation. BF results would be worthless.

How to properly specify P_δ? I

How to properly specify P_δ? I

What is P_δ?

P_δ is a probability distribution, which specifies the available knowledge and beliefs about δ prior to data collection. It should reflect the expectations about δ under H_1^B (?).

How to properly specify P_δ? I

What is P_δ?

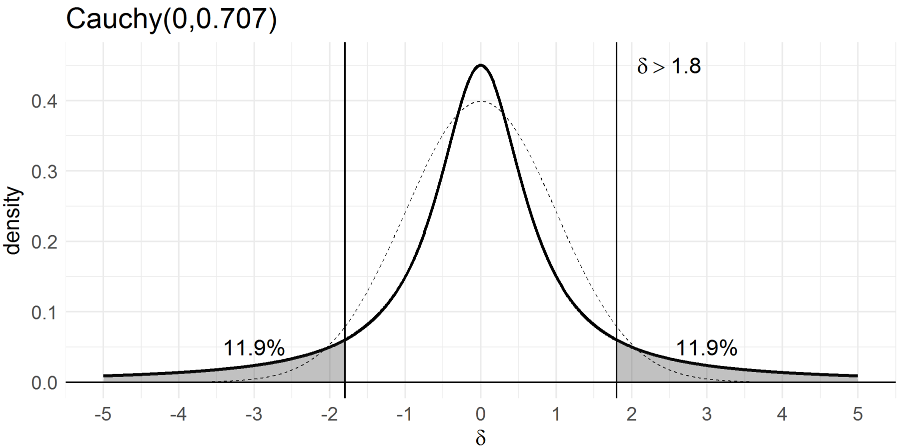
P_δ is a probability distribution, which specifies the available knowledge and beliefs about δ prior to data collection. It should reflect the expectations about δ under H_1^B (?).

What is state of the art?

Predominantly $P_\delta \sim Cauchy(0, \sqrt{2}/2)$.
Sometimes $P_\delta \sim N(0, 1)$ or $P_\delta \sim N(\mu_\delta, \sigma_\delta^2)$.

The Cauchy distribution

The Cauchy distribution



Effect sizes:

$\delta = 0.2$: small; $\delta = 0.5$: medium; $\delta = 0.8$: large

$\delta = 1.8$: association gender - body height

How to properly specify P_δ? II

How to properly specify P_δ? II

About the absurdity of the Cauchy effect size prior:

Before seeing the data, the researcher is about 23.8% confident that $|\delta|$ is larger than one of the largest effect sizes in psychology.

How to properly specify P_δ? II

About the absurdity of the Cauchy effect size prior:

Before seeing the data, the researcher is about 23.8% confident that $|\delta|$ is larger than one of the largest effect sizes in psychology.

I would offer bets :-)

About the absurdity of the Cauchy effect size prior:

Before seeing the data, the researcher is about 23.8% confident that $|\delta|$ is larger than one of the largest effect sizes in psychology.

I would offer bets :-)

Necessity of an imprecise effect size prior:

By default, precise information about δ is lacking. Else, no scientific investigation would be needed.
$\Rightarrow$ A proper specification of P_δ should be imprecise.

A first imprecise Bayes Factor

A first imprecise Bayes Factor

Test-relevant prior:

$$\delta \sim N(\mu_\delta, \sigma_\delta^2) \quad \text{with} \quad \mu_\delta \in \left[\underline{\mu}_\delta; \overline{\mu}_\delta\right], \quad \sigma_\delta^2 \in \left[\underline{\sigma}_\delta^2; \overline{\sigma}_\delta^2\right]$$

A first imprecise Bayes Factor

Test-relevant prior:

$$\delta \sim N(\mu_\delta, \sigma_\delta^2) \quad \text{with} \quad \mu_\delta \in \left[\underline{\mu}_\delta; \overline{\mu}_\delta\right], \quad \sigma_\delta^2 \in \left[\underline{\sigma}_\delta^2; \overline{\sigma}_\delta^2\right]$$

$$\mathcal{M} := \left\{ P_\delta = N(\mu_\delta, \sigma_\delta^2) \mid \mu_\delta \in \left[\underline{\mu}_\delta; \overline{\mu}_\delta\right], \sigma_\delta^2 \in \left[\underline{\sigma}_\delta^2; \overline{\sigma}_\delta^2\right] \right\}$$

A first imprecise Bayes Factor

Test-relevant prior:

$$\delta \sim N(\mu_\delta, \sigma_\delta^2) \quad \text{with} \quad \mu_\delta \in \left[\underline{\mu}_\delta; \overline{\mu}_\delta\right], \quad \sigma_\delta^2 \in \left[\underline{\sigma}_\delta^2; \overline{\sigma}_\delta^2\right]$$

$$\mathcal{M} := \left\{P_\delta = N(\mu_\delta, \sigma_\delta^2) \mid \mu_\delta \in \left[\underline{\mu}_\delta; \overline{\mu}_\delta\right], \sigma_\delta^2 \in \left[\underline{\sigma}_\delta^2; \overline{\sigma}_\delta^2\right]\right\}$$

Bayesian Hypotheses:

$$H_0^B : \begin{array}{l} \mu \sim P_\mu \\ \sigma^2 \sim P_{\sigma^2} \\ \delta = 0 \end{array} \quad \text{vs.} \quad H_1^B : \begin{array}{l} \mu \sim P_\mu \\ \sigma^2 \sim P_{\sigma^2} \\ \delta \sim \mathcal{M} \end{array}$$

A first imprecise Bayes Factor

Test-relevant prior:

$$\delta \sim N(\mu_\delta, \sigma_\delta^2) \quad \text{with} \quad \mu_\delta \in \left[\underline{\mu}_\delta ; \overline{\mu}_\delta\right], \quad \sigma_\delta^2 \in \left[\underline{\sigma}_\delta^2 ; \overline{\sigma}_\delta^2\right]$$

$$\mathcal{M} := \left\{ P_\delta = N(\mu_\delta, \sigma_\delta^2) \mid \mu_\delta \in \left[\underline{\mu}_\delta ; \overline{\mu}_\delta\right], \sigma_\delta^2 \in \left[\underline{\sigma}_\delta^2 ; \overline{\sigma}_\delta^2\right] \right\}$$

Bayesian Hypotheses:

$$H_0^B : \begin{array}{l} \mu \sim P_\mu \\ \sigma^2 \sim P_{\sigma^2} \\ \delta = 0 \end{array} \quad \text{vs.} \quad H_1^B : \begin{array}{l} \mu \sim P_\mu \\ \sigma^2 \sim P_{\sigma^2} \\ \delta \sim \mathcal{M} \end{array}$$

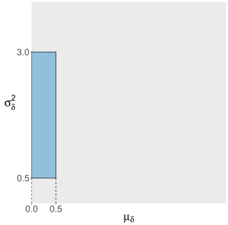
Imprecise Bayes Factor:

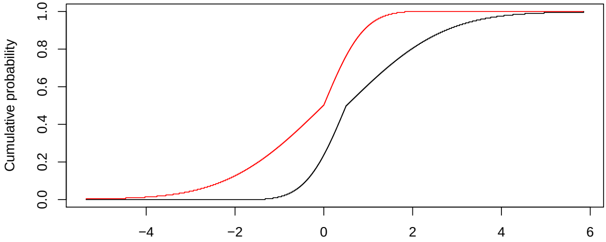
$$IBF_{10} = \left[\min_{P_\delta \in \mathcal{M}} BF_{10} ; \max_{P_\delta \in \mathcal{M}} BF_{10} \right]$$

Example I

Example I

$$\delta \sim N(\mu_\delta, \sigma_\delta^2) \quad \text{with} \quad \mu_\delta \in [0; 0.5], \quad \sigma_\delta^2 \in [0.5; 3]$$

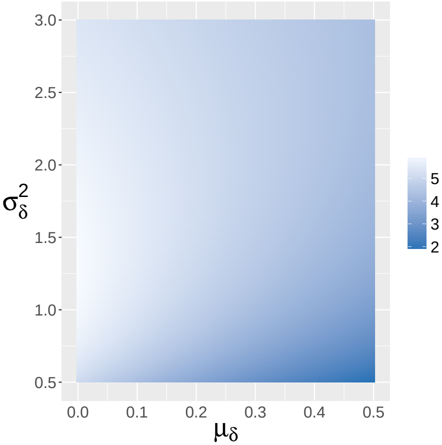




Example II

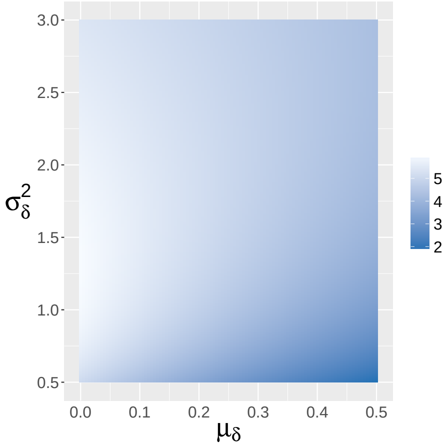
Example II

$$IBF_{10} = [1.84; 5.99]$$



Example II

$$IBF_{10} = [1.84; 5.99]$$



Interpretation:

The data is between 1.84 and 5.99 times as much evidence for H_1^B than for H_0^B, i.e. for an effect with an effect size in accordance with the available knowledge about it than for no effect.
(ignoring P_μ and P_{σ^2})

What is next?

What is next?

This was only a credal set of normal effect size distributions.

What is next?

This was only a credal set of normal effect size distributions.

$\Rightarrow$ p-boxes as effect size priors.

Thank you for your Attention!

Thank you for your Attention!