Robust Bayes Factor for Independent Two-Sample Comparisons under Imprecise Prior Information

Luisa Ebner[1] - **Patrick Schwaferts**[2] - **Thomas Augustin**[3]

International Symposium on Imprecise Probabilities: Theories and Applications (ISIPTA) 2019 - Ghent, Belgium

[1] *l.t.ebner@student.vu.nl - Master Student of Artificial Intelligence, Vrije Universiteit Amsterdam, Netherlands*

[2] *patrick.schwaferts@stat.uni-muenchen.de - Institut für Statistik, Ludwig-Maximilians Universität München (LMU), Munich, Germany*

[3] *thomas.augustin@stat.uni-muenchen.de - Institut für Statistik, Ludwig-Maximilians Universität München (LMU), Munich, Germany*

Experimental Setup

Data: $\boldsymbol{x} = (\boldsymbol{x_1}, \dots, \boldsymbol{x_i}, \dots, \boldsymbol{x_n}), \boldsymbol{y} = (\boldsymbol{y_1}, \dots, \boldsymbol{y_j}, \dots, \boldsymbol{y_m})$

Model: $X_i \sim N(\mu, \sigma^2), Y_j \sim N(\mu + \alpha, \sigma^2)$

Standardized Effect Size: $\delta \coloneqq \alpha / \sigma$

Paramters: μ, σ^2, δ

The parameters are not independent of each other.
For the depicted context, this is rather umproblematic.

Bayes Factor

Priors: $\mu \sim const, \sigma^2 \sim 1/\sigma^2$

Hypotheses: $H_0: \delta = 0$ vs. $H_1: \delta | \sigma^2 \sim N(\mu_\delta, \sigma_\delta^2)$

Bayes Factor:

$$BF = \frac{\iiint f(\boldsymbol{x}, \boldsymbol{y} | \mu, \sigma^2, \delta)\, \pi(\delta|\sigma^2)\pi(\sigma^2)\pi(\mu)\, d\sigma^2\, d\mu\, d\delta}{\iint f(\boldsymbol{x}, \boldsymbol{y} | \mu, \sigma^2, \delta = 0)\, \pi(\sigma^2)\pi(\mu)\, d\sigma^2\, d\mu}$$

Specification of Hyperparamters: $\mu_\delta, \sigma_\delta^2$

For this context, the Bayes Factor can be calculated by a closed formula. [Gönen et al. 2005]

Interpretation:
The data (x, y) are BF times as much evidence for H_1 than for H_0.

Problem: A precise hyperparameter specification is rarely possible.

Solution: Allow an interval-valued hyperparameter specification.

Robust Bayes Factor

Imprecise Hyperparameters: $\mu_\delta \in [\underline{\mu_\delta}, \overline{\mu_\delta}], \sigma_\delta^2 \in [\underline{\sigma_\delta^2}, \overline{\sigma_\delta^2}]$

Imprecise Hypotheses: $H_0: \delta = 0$ vs. $H_1: \delta | \sigma^2 \sim \mathcal{M}$

$$\mathcal{M} = \left\{ N(\mu_\delta, \sigma_\delta^2) \;\middle|\; \mu_\delta \in [\underline{\mu_\delta}, \overline{\mu_\delta}], \sigma_\delta^2 \in [\underline{\sigma_\delta^2}, \overline{\sigma_\delta^2}] \right\}$$

Robust Bayes Factor: $rBF = [\underline{BF}, \overline{BF}]$

$$\underline{BF} = \min_{N(\mu_\delta, \sigma_\delta^2) \in \mathcal{M}} BF$$

$$\overline{BF} = \max_{N(\mu_\delta, \sigma_\delta^2) \in \mathcal{M}} BF$$

The alternative hypothesis states that δ is distributed in accordance with the (vaguely available) knowledge about δ.

Interpretation:
The data (x, y) are $\underline{BF}$ to $\overline{BF}$ times as much evidence for H_1 than for H_0.

Sometimes the evidence is ambiguous, but then „it seems wisest just to conclude that there is no answer; more evidence is needed to solve the ambiguity.“ [Berger 1990, p. 307]

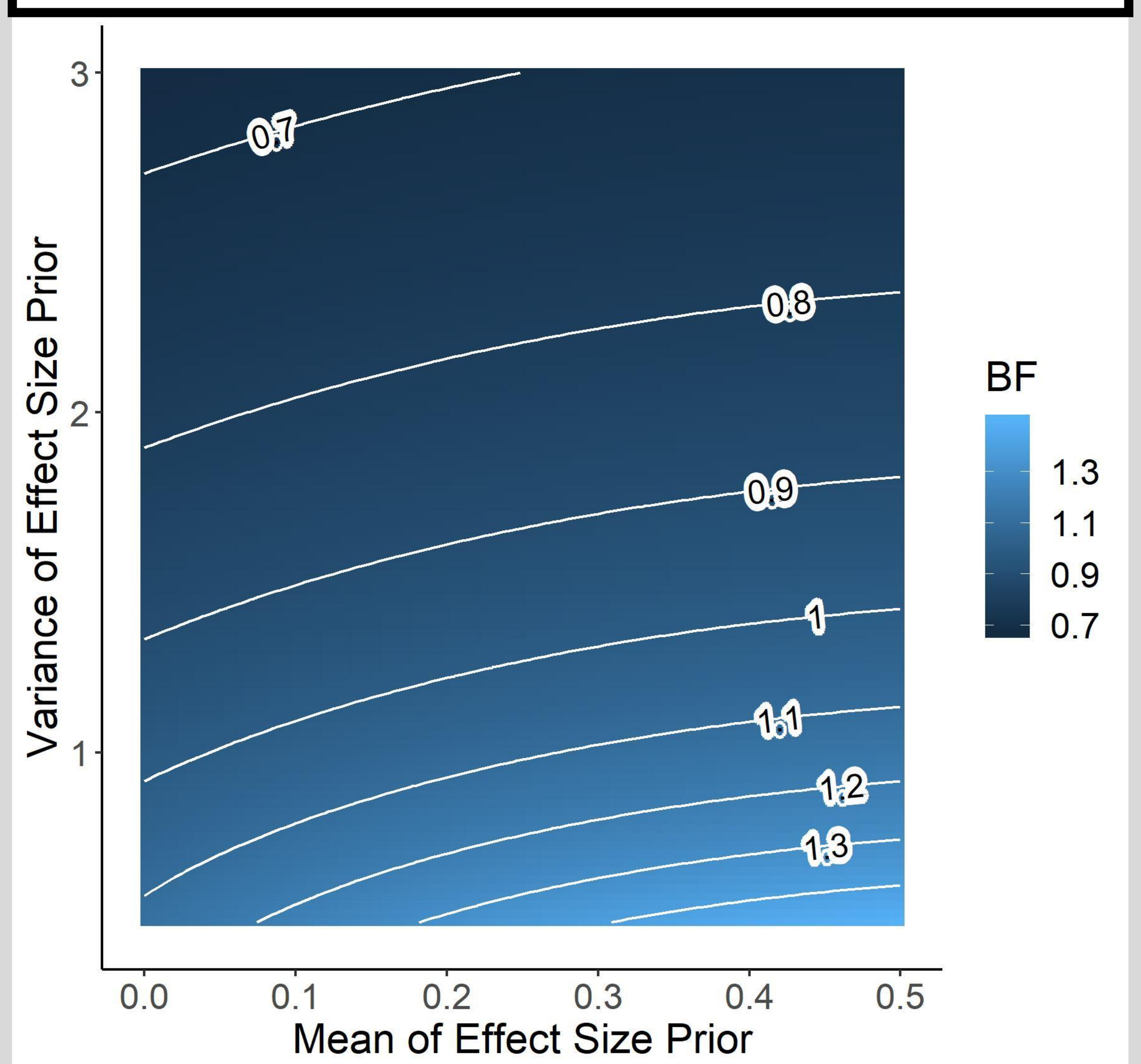
Example (fictitious)

Prior: $\mathcal{M} = \{ N(\mu_\delta, \sigma_\delta^2) \mid \mu_\delta \in [0, 0.5], \sigma_\delta^2 \in [0.5, 3] \}$

Gender differences in recurrence rates of major depression. [van Loo et al. 2017]
The risk of recurrence is expressed by a score for women (Y) and men (X).
$\mathcal{M}$ captures reasonable hyperparameter choices.
H_0 represents no gender difference.
H_1 represents higher recurrence rates for women than for men.

Analysis 1

$n = m = 10, \quad rBF = [0.67, 1.50]$



Analysis 1:
The data are 0.67 to 1.50 times as much evidence for H_1 than for H_0.

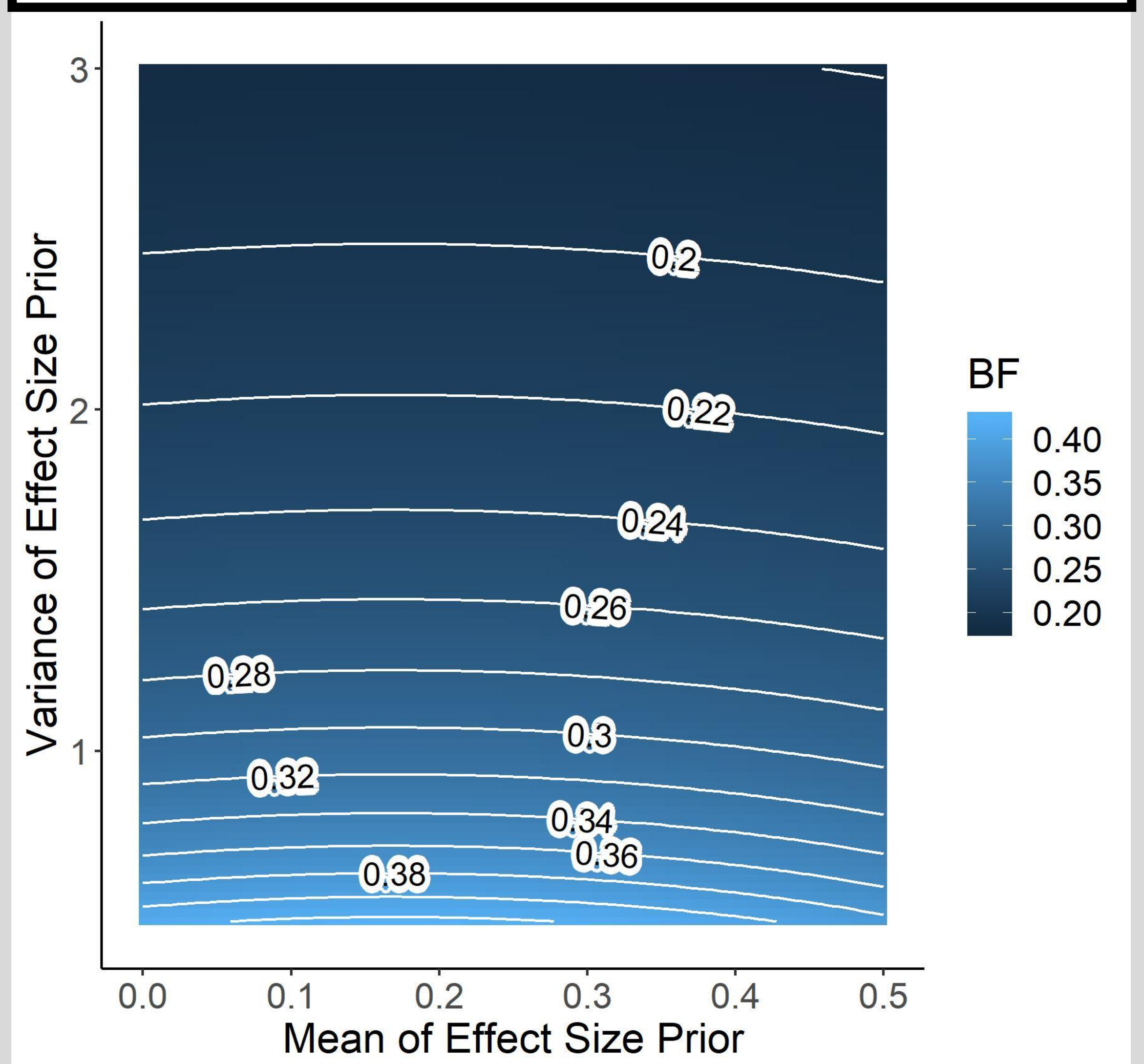
There is no unambiguous evidence and more data are collected.

Analysis 2:
The data are 1/0.42 = 2.4 to 1/0.18 = 5.5 times as much evidence for H_0 than for H_1.

The data are (slightly) favoring the hypothesis of similar recurrance rates.

Analysis 2

$n = m = 30, \quad rBF = [0.18, 0.42]$



Berger, J. O. (1990). Robust Bayesian analysis: sensitivity to the prior. *Journal of statistical planning and inference*, *25*(3), 303-328.

Gönen, M., Johnson, W. O., Lu, Y., & Westfall, P. H. (2005). The Bayesian two-sample t test. *The American Statistician*, *59*(3), 252-257.

van Loo, H. M., Aggen, S. H., Gardner, C. O., & Kendler, K. S. (2018). Sex similarities and differences in risk factors for recurrence of major depression. *Psychological medicine*, *48*(10), 1685-1693.

PS wants to thank the LMU Mentoring Program.