All for one or some for all

Fayette Klaassen

Summer School Utrecht

31 August 2018

Introduction: Bayes factors

- So far, focus on estimation, but what if you want to test hypotheses?

- Null hypothesis significance testing
  - H0: $\theta_1 = \theta_2 = \theta_3$ **vs** Ha: *not* $\theta_1 = \theta_2 = \theta_3$
  - p-value to reject H0 or not
  - Not possible to compare specific hypotheses

- Bayes factor!

Introduction: Bayes factors

- Bayesian version of hypothesis testing

- BF_{ab} compares any two hypotheses

$$H0: \theta_1 = \theta_2 = \theta_3$$
$$Ha: \text{not } H0$$

Introduction: Bayes factors

- Bayesian version of hypothesis testing
- BF_{ab} compares any two hypotheses

H0: $\theta_1 = \theta_2 = \theta_3$
Ha: not H0
H1: $\theta_1 > \theta_2 > \theta_3$
H2: $\theta_1 > \theta_3 > \theta_2$
H3: not H1 or H2
H4: $\theta_1 = \theta_2 > \theta_3$

Introduction: Bayes factors

- How does it work?

$$\frac{\Pr(D|M_1)}{\Pr(D|M_2)} = \frac{\int \Pr(\theta_1|M_1)\Pr(D|\theta_1, M_1)\,d\theta_1}{\int \Pr(\theta_2|M_2)\Pr(D|\theta_2, M_2)\,d\theta_2}$$

- Balance complexity and fit

- $\mathrm{BF}_{\mathrm{io}} = \dfrac{fit_i}{complexity_i}$

Introduction: Bayes factors

- Bayes factor expresses the support the data for H_a, relative to H_a
  - $BF_{ab} = 5$, H_a is 5 times more likely than H_a
  - $BF_{ab} = .1$, H_b is 10 times more likely than H_a

- Possible to compare multiple hypotheses

All for one or some for all

Fayette Klaassen

Summer School Utrecht

31 August 2018

the Example

participants

the Example

participants

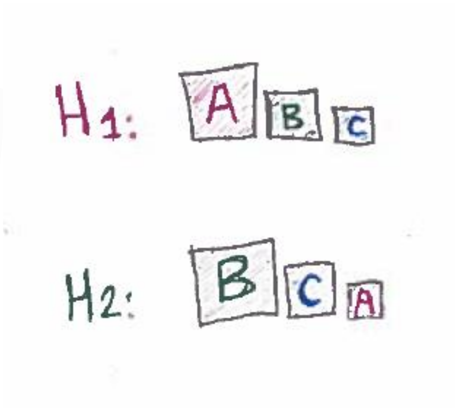
conditions

the Example

participants

conditions

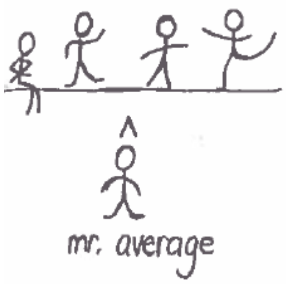
hypotheses



the Question

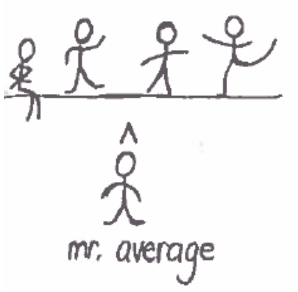
the Question

Group effects



the Method

Is H_1 better than H_2?



- Bayes factor

$$\frac{\Pr(D|M_1)}{\Pr(D|M_2)} = \frac{\int \Pr(\theta_1|M_1)\Pr(D|\theta_1, M_1)\, d\theta_1}{\int \Pr(\theta_2|M_2)\Pr(D|\theta_2, M_2)\, d\theta_2}$$

H_1 or H_2

the Question

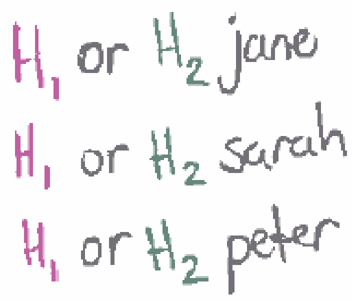
Group effects

Individual effects

the Method

For who is H_1 better than H_2?

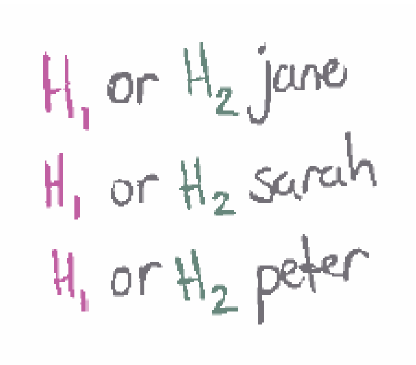
- Individual Bayes factors

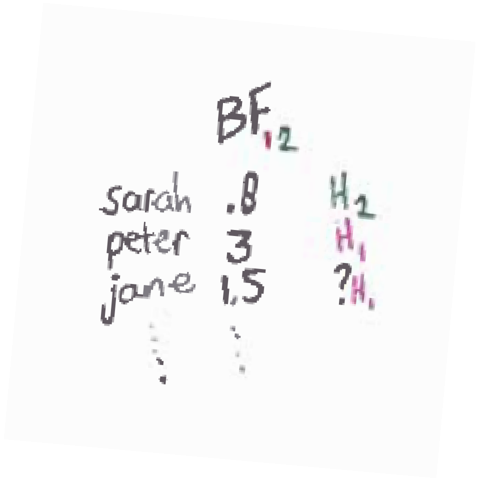


the Method

For who is H_1 better than H_2?

- Individual Bayes factors





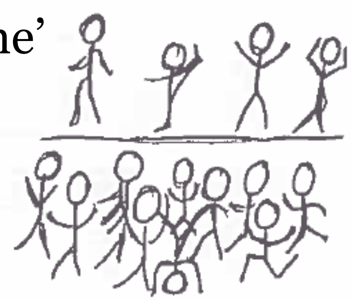
the Question

Group effects

Individual effects

Group of individuals

- 'does it work for everyone'



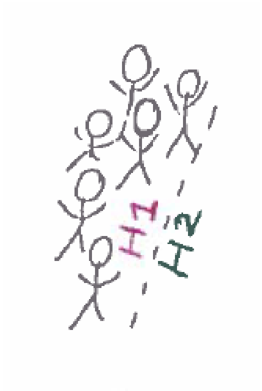
Intermezzo

- Multilevel analysis
  - Random effects
  - Group effect
- Updating
  - Group effect
  - After every new data point
- Synthesis
  - Combine analyses at the individual level
  - To a group level conclusion

the Method

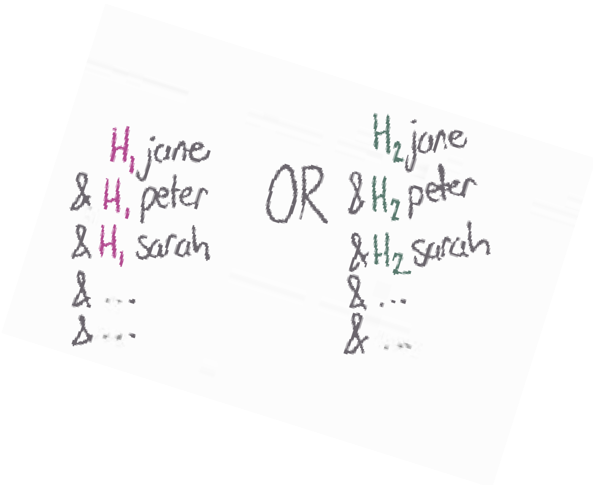
Is H1 better than H2 for everybody?

- Aggregate Bayes factors



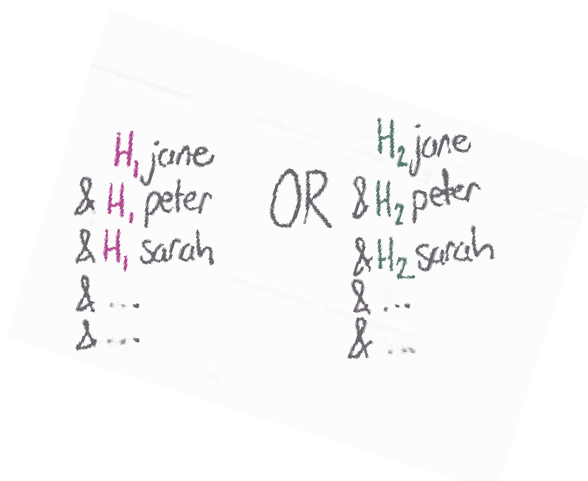
the Method

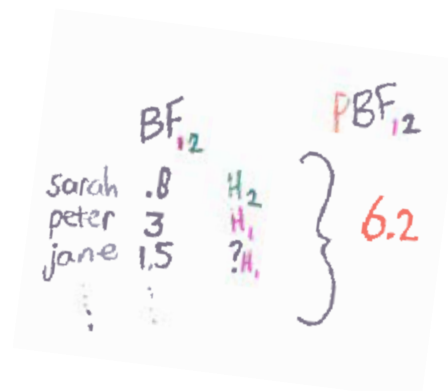
Aggregate Bayes factors



the Method

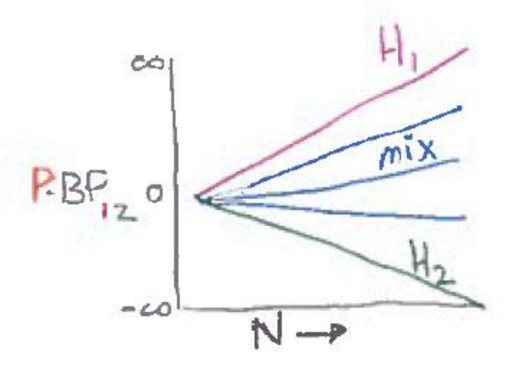
Aggregate Bayes factors





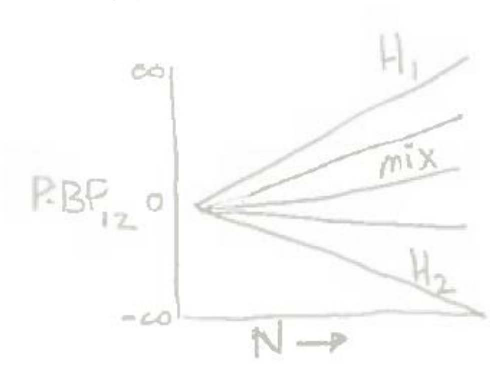
the Research

Product

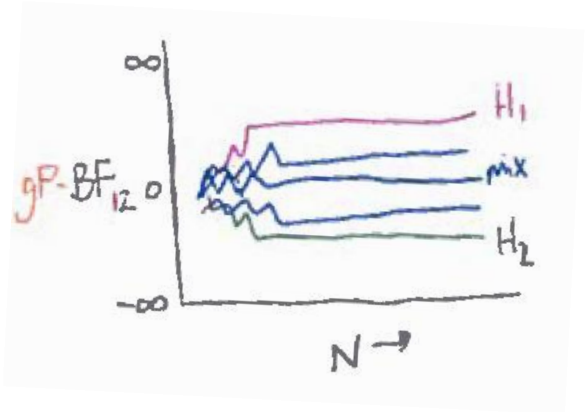


the Research

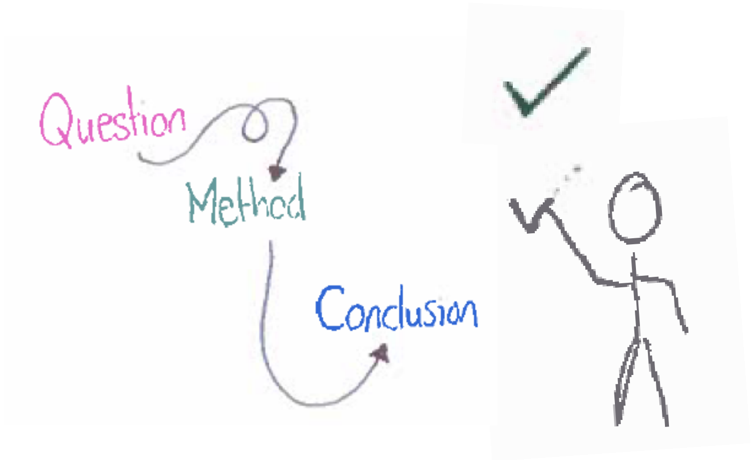
Product



Geometric mean

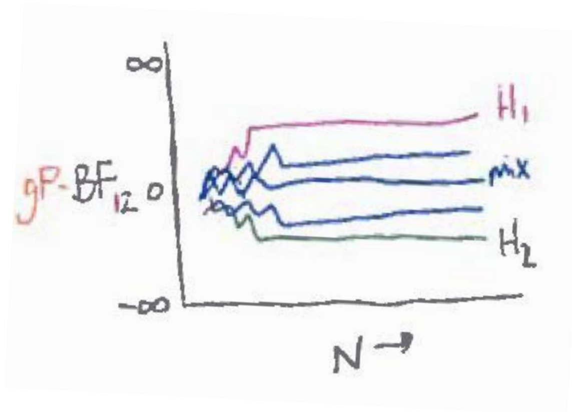


the Conclusion



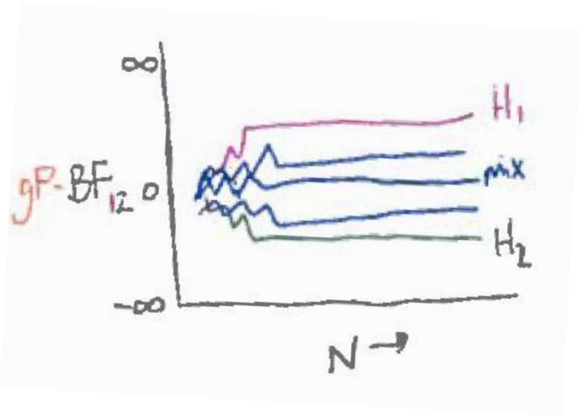
the Conclusion

‘Does everybody…’



the Conclusion

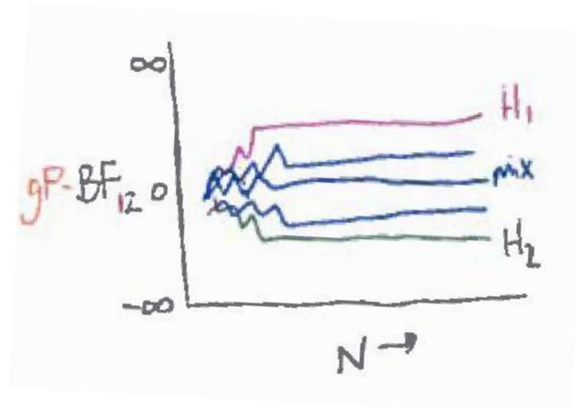
‘Does everybody…’



Evidence Rate

the Conclusion

'Does everybody...'

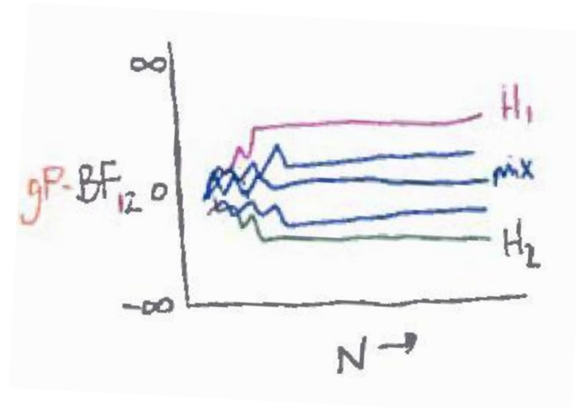


Evidence Rate

Stability Rate

the Conclusion

‘Does everybody…’

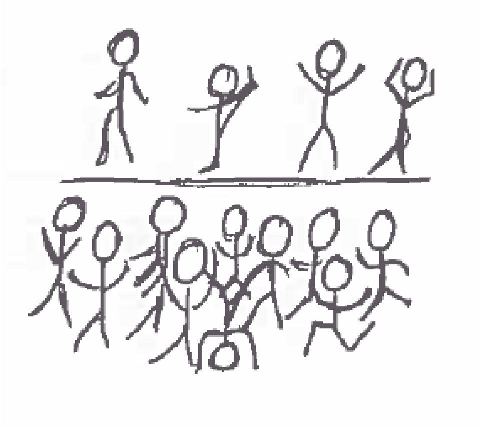


Very specific question

- Works with relatively small sample

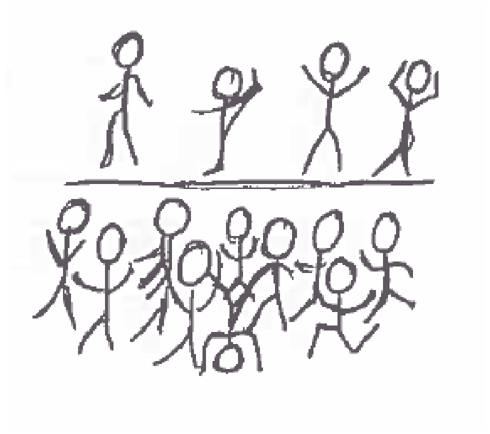
the Illutration

Some for all



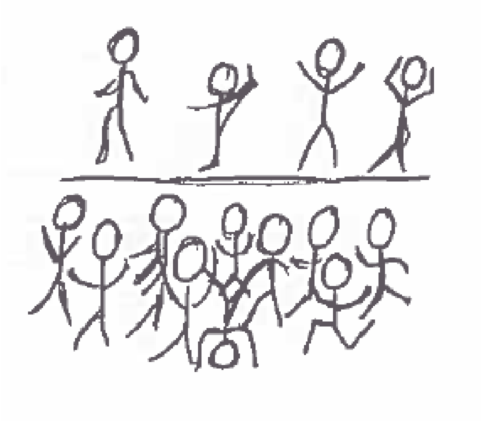
the Discussion

Some for all



the Discussion

Some for all



- Subgroups?
- Indecision?
- Strength of evidence?

Learn more...?

f.klaassen@uu.nl
www.github.com/fayetteklaassen

Behav Res
https://doi.org/10.3758/s13428-017-0992-5

CrossMark

All for one or some for all? Evaluating informative hypotheses using multiple $N = 1$ studies

Fayette Klaassen[1] · Claire M. Zedelius[2] · Harm Veling[3] · Henk Aarts[4] · Herbert Hoijtink[1,5]

Abstract Analyses are mostly executed at the population level, whereas in many applications the interest is on the individual level instead of the population level. In this paper, multiple $N = 1$ experiments are considered, where participants perform multiple trials with a dichotomous outcome in various conditions. Expectations with respect to the performance of participants can be translated into so-called informative hypotheses. These hypotheses can be evaluated for each participant separately using Bayes factors. A Bayes factor expresses the relative evidence for two

individuals. Two additional measures are proposed to support the interpretation of the gP-BF: the evidence rate (ER), the proportion of individual Bayes factors that support the same hypothesis as the gP-BF, and the stability rate (SR), the proportion of individual Bayes factors that express a stronger support than the gP-BF. These three statistics can be used to determine the relative support in the data for the informative hypotheses entertained. Software is available that can be used to execute the approach proposed in this paper and to determine the sensitivity of the outcomes with