# Informative hypotheses evaluation Bayesian model selection

Rebecca M. Kuiper (credits slides: Herbert Hoijtink and others)

Department of Methodology & Statistics Utrecht University

### **Table of Contents**

### Bayesian Informative Hypotheses Evaluation (bain)

ANOVA and Beyond, Example Analyses with R and JASP

### bain

1. Hoijtink, H., Mulder, J., van Lissa, C., and Gu, X. (2018). A tutorial on testing hypotheses using the Bayes factor. Psychological Methods, 24, 539-556.

#### **Balancing Fit and Complexity**

The Bayes factor quantifies the relative support in the data for two hypotheses, for example,

$$H_i: \mu_1 > \mu_2 > \mu_3$$

$$H_{u}: \mu_{1}, \mu_{2}, \mu_{3}$$

with

$$BF_{iu} = \frac{f_i}{c_i} = \frac{\text{fit } H_i}{\text{complexity } H_i}$$

that is, after observing the data  $H_i$  is  $BF_{iu}$  times as likely as  $H_u$ , for example, .2, 5, 10.

#### Balancing Fit and Complexity

### A (very) loose interpretation of the meaning of fit

$$H_i: \mu_1 > \mu_2 > \mu_3$$

if 
$$\bar{x}_1 = 7 \& \bar{x}_2 = 4 \& \bar{x}_3 = 2$$
 the fit is good

if 
$$\bar{x}_1 = 2 \& \bar{x}_2 = 4 \& \bar{x}_3 = 7$$
 the fit is bad

Balancing Fit and Complexity

### A (very) loose interpretation of the meaning of complexity

$$H_1: \mu_1 = \mu_2 = \mu_3$$

very parsimonious, the means have to be exactly equal.

$$H_1: \mu_1 > \mu_2 > \mu_3$$

one ordering of three means: 1-2-3, thus is parsimonious.

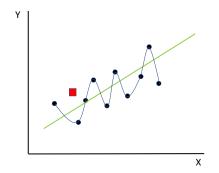
$$H_2: \mu_1 > (\mu_2, \mu_3)$$

2 orderings of three means: 1-2-3 and 1-3-2, less parsimonious.

$$H_{u}: \mu_{1}, \mu_{2}, \mu_{3}$$

contains all six possible orderings of three means, not parsimonious.

#### **Balancing Fit and Complexity**



The straight line results from a linear regression model with 3 parameters (intercept, slope, residual variance).

The other line results from a polynomial regression models with 11 parameters (intercept, nine slopes, residual variance).

The red square is a new observation that is added to the original 10 observations.

What is the predictive value of both models?

#### Balancing Fit and Complexity

#### Three forms of Hypotheses and Bayes factors involving $H_i: \mu_1 > \mu_2 > \mu_3$

 $BF_{iii}$  evaluating  $H_i$  versus  $H_{ii}: \mu_1, \mu_2, \mu_3$ 

 $BF_{ii'}$  evaluating  $H_i$  versus  $H_{i'}$ 

 $BF_{ic}$  evaluating  $H_i$  versus  $H_c$ : not  $H_i$ 

Interpreting (the Size of) the Bayes Factor

- 1. Select the best of a set of hypotheses using  $BF_{iu}$
- 2. Compare two competing hypotheses using  $BF_{ii'}$
- 3. Compare "my theory" with "not my theory" using  $BF_{ic}$

	t <sub>i</sub>	$c_i$	$BF_{iu}$	$BF_{ic}$
H <sub>1</sub> : Sex Match	.0039	.012	.32	.32
$H_2$ : Gender Role Match	.0725	.012	5.85	6.44
$H_3$ : Sex Mismatch	.0007	.012	.06	.06
H <sub>4</sub> : Gender Role Mismatch	.0001	.012	.01	.01

#### Descriptives

#### Gender Role Match Effect

$$H_2: (\mu_1, \mu_5) > (\mu_2, \mu_3, \mu_4, \mu_6)$$
 and  $(\mu_7, \mu_{11}) > (\mu_8, \mu_9, \mu_{10}, \mu_{12})$ 

$$H_2: (166, 163) > (158, 154, 155, 164)$$
 and

#### Gender Role Mismatch Effect

$$H_4: (\mu_2, \mu_4) > (\mu_1, \mu_3, \mu_5, \mu_6) \text{ and } (\mu_8, \mu_{10}) > (\mu_7, \mu_9, \mu_{11}, \mu_{12})$$

$$H_4: (158, 155) > (166, 154, 163, 164)$$
 and

Interpreting (the Size of) the Bayes Factor

- 1. The Bayes factor **is** a measure of support (also for the null-hypothesis)
- 2. The Bayes factor can be indecisive. A value around 1 denotes "the data don't tell us which hypothesis to prefer"
- 3. One can update, that is, collect more data and recompute the Bayes factor
- 4. One can compare more than two hypotheses (see extra comments later on)
- "Something is going on and we do know what!"
- 6. The Bayes factor selects the best of the hypotheses **under consideration**. Note that the "true" hypothesis may not be among them, and that all hypotheses may be "wrong"

Interpreting (the Size of) the Bayes Factor

### When is the Bayes factor large enough?

- 1. Guidelines by Jeffreys (1969) and Kass and Raftery (1995), e.g., < 3 is ignorable, > 3 is positive evidence, > 10is strong evidence ...
- 2. Will lead to a return of sloppy science and publication bias (when no pre-registration or pre-registered report)
- 3. Were does the 3 come from?

Interpreting (the Size of) the Bayes Factor

### When is the Bayes factor (BF) large enough?

- 1. Before collecting or accessing the data, formulate informative hypotheses (and decide how large you would like the BF to be).
- 2. Insert this information in pre-registration or pre-registered report.
- 3. Collect data and evaluate hypotheses.
  - Is one good and the best with a "large" Bayes factor: nice!
  - Are the Bayes factors "not large enough": follow up research or updating is needed.
  - Is none good: BIG news, well-constructed hypotheses have been rejected!

### Extra: PMPs

Posterior Model Probabilities, e.g.,  $PMP(H_i \mid data)$  and  $PMP(H_c \mid data)$  quantify the support in the data for each hypothesis.

$$\frac{PMP(H_i|\text{data})}{PMP(H_c|\text{data})} = \mathsf{BF}_{ic} \times \frac{PRI(H_i)}{PRI(H_c)},\tag{1}$$

where  $PRI(H_i)$  and  $PRI(H_c)$  denote the *prior* probabilities, that is, an evaluation of the support for the hypotheses before observing the data.

Usually equal prior model probabilities are used (which means that the PMP's convey the same information as the Bayes factors), but this is not a requirement.

bain 000000000

### **PMPs PMPs**

	$f_i$	$C_i$	$BF_{iu}$	$PMP_i$	$PRI_i$
H <sub>1</sub> : Sex Match	.0039	.012	.32	.04	1/5
H <sub>2</sub> : Gender Role Match	.0725	.012	5.85	.81	1/5
H <sub>3</sub> : Sex Mismatch	.0007	.012	.06	.01	1/5
H <sub>4</sub> : Gender Role Mismatch	.0001	.012	.01	.00	1/5
$H_u$ :				.14	1/5

# Note on hypotheses (1/2)

#### $H_i$ contains 1 ordering of means:

1. 
$$\mu_1 > \mu_2 > \mu_3$$

### $H_c$ contains 5 orderings of means:

- 2.  $\mu_1 > \mu_3 > \mu_2$
- 3.  $\mu_2 > \mu_1 > \mu_3$
- 4.  $\mu_2 > \mu_3 > \mu_1$
- 5.  $\mu_3 > \mu_1 > \mu_2$
- 6.  $\mu_3 > \mu_2 > \mu_1$

 $H_{ii}$  combines  $H_{ii}$  and  $H_{ci}$ .

# Note on hypotheses (2/2)

#### $H_1$ contains 1 ordering of means:

1. 
$$\mu_1 > \mu_2 > \mu_3$$

### $H_2$ contains 1 ordering of means:

2. 
$$\mu_1 > \mu_3 > \mu_2$$

### $H_c$ (of set!) contains 4 orderings of means:

3. 
$$\mu_2 > \mu_1 > \mu_3$$

4. 
$$\mu_2 > \mu_3 > \mu_1$$

5. 
$$\mu_3 > \mu_1 > \mu_2$$

6. 
$$\mu_3 > \mu_2 > \mu_1$$

 $H_{II}$  combines  $H_1$ ,  $H_2$ , and  $H_c$ .



#### Replacing $H_u$ by $H_c$

	f <sub>i</sub>	Ci	$BF_{iu}$	$PMP_i$	$PRI_i$
H <sub>1</sub> : Sex Match	.0039	.012	.32	.04	1/5
H <sub>2</sub> : Gender Role Match	.0725	.012	5.85	.84	1/5
H <sub>3</sub> : Sex Mismatch	.0007	.012	.06	.00	1/5
H <sub>4</sub> : Gender Role Mismatch	.0001	.012	.01	.00	1/5
$H_c$ :	.9200	.9500	.97	.12	1/5

where  $H_c$  denotes the complement  $H_1$  through  $H_4$ , that is, "not one of these four hypotheses".

#### Bayesian Error Probabilities

PMPs can be interpreted as Bayesian error probabilities, that is, the Bayesian counterparts of the Type I and Type II errors.

	f <sub>i</sub>	Ci	$BF_{iu}$	$PMP_i$	$PRI_i$
H <sub>1</sub> : Sex Match	.0039	.012	.32	.04	1/5
H <sub>2</sub> : Gender Role Match	.0725	.012	5.85	.81	1/5
H <sub>3</sub> : Sex Mismatch	.0007	.012	.06	.01	1/5
H <sub>4</sub> : Gender Role Mismatch	.0001	.012	.01	.00	1/5
$H_{u}$ :				.14	1/5

#### The Number of Hypotheses and PMPs

Look what happens if we compare many hypotheses, the PMPs become smaller and smaller, and thus the Bayesian error probabilities become larger and larger:

	f <sub>i</sub>	Ci	$BF_{iu}$	$PMP_i$	$PRI_i$
H <sub>1</sub> : Sex Match	.0039	.012	.32	.013	1/13
H <sub>2</sub> : Gender Role Match	.0725	.012	5.85	.270	1/13
H <sub>3</sub> : Sex Mismatch	.0007	.012	.06	.003	1/13
H <sub>4</sub> : Gender Role Mismatch	.0001	.012	.01	.000	1/13
$H_5$ : Lets try this one too	.0521	.012	2.61	.180	1/13
$H_{12}$ : Don't miss something	.0164	.012	1.36	.040	1/13
$H_u$ :				.047	1/13

#### The Number of Hypotheses and PMPs

The same results as two slides up are in fact obtained by assigning PMPs of 0 to each hypothesis that is NOT considered:

	f <sub>i</sub>	$C_i$	$BF_{iu}$	$PMP_i$	$PRI_i$
H <sub>1</sub> : Sex Match	.0039	.012	.32	.04	1/5
$H_2$ : Gender Role Match	.0725	.012	5.85	.81	1/5
H <sub>3</sub> : Sex Mismatch	.0007	.012	.06	.01	1/5
H <sub>4</sub> : Gender Role Mismatch	.0001	.012	.01	.00	1/5
$H_5$ : Lets try this one too	.0521	.012	2.61	.18	0
$H_{12}$ : Don't miss something	.0164	.012	1.36	.04	0
$H_U$ :				.14	1/5

# Subjectivity of Bayesian Hypotheses Evaluation

- 1. Which hypotheses to evaluate?
- 2. How to formalize hypotheses? E.g.  $(\mu_1, \mu_2) > (\mu_3, \mu_4)$  or  $\mu_1 = \mu_2 > \mu_3 = \mu_4$
- 3. The (implicit) choice for equal prior model probabilities
- 4. The specification of the prior distribution

### Hands-on/Demo: BMS

#### Let's practice.

- If needed: Start Rstudio again (optional: make project).
- Open 'Hands-on\_1\_BMS\_Unc\_ANOVA\_bain.R' (in 'Hands-on files').
- Install packages and load them.
- Read and inspect data. Use Data\_Lucas.txt.
- Run model (lm()).
- Specify hypotheses (make up your own).
   Note: Use names used in the model.
- Run bain().
- Inspect and interpret output.

#### **Table of Contents**

Bayesian Informative Hypotheses Evaluation (bain)

ANOVA and Beyond, Example Analyses with R and JASP

End

Extra

Example 1: ANOVA

What is the relation between "knowledge of numbers after watching Sesame Street for a year"

and

site from which the child originates:

- 1. disadvantaged inner city
- advantaged suburban
- advantaged rural
- 4. disadvantaged rural
- 5. disadvantaged Spanish speaking

Example 1: ANOVA

```
library (bain)
sesamesim$site <- as.factor(sesamesim$site)</pre>
anov <- lm(postnumb~site-1, sesamesim)
coef (anov)
set.seed(100)
results <- bain(anov,
                "site1=site2=site3=site4=site5;
                 site2>site5>site1>site3>site4")
print(results)
summary (results, ci = 0.95)
```





Example 1: ANOVA

#### coef(anov) renders

```
site1 site2 site3 site4 site5
29.66667 38.98182 23.18750 25.32558 31.72222
```

#### summary(results) renders

```
Parameter n Estimate lb ub

1 site1 60 29.66667 26.82991 32.50343

2 site2 55 38.98182 36.01892 41.94472

3 site3 64 23.18750 20.44082 25.93418

4 site4 43 25.32558 21.97466 28.67650

5 site5 18 31.72222 26.54303 36.90141
```

Example 1: ANOVA

#### The main output is

```
Fit Com BF.u BF.c PMPa PMPb PMPc
H1 0.000 0.000 0.000 0.000 0.000 0.000 0.000
H2 0.121 0.008 14.559 16.428 1.000 0.936 0.943
Hu 0.064
Hc 0.879 0.992 0.886 0.057
```

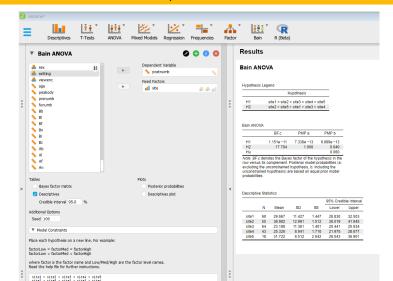
#### Hypotheses:

H1: site1=site2=site3=site4=site5
H2: site2>site5>site1>site3>site4





Example 1: ANOVA



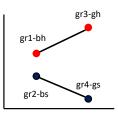
**Example 2: ANOVA Interaction Effect** 

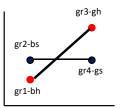
Dependent variable: Knowledge of numbers.

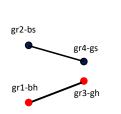
Factors: sex (boy, girl) and setting (watching at home, watching at school).

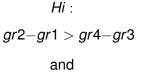
Gr: 1=boyhome, 2= boyschool, 3= girlhome, 4=girlschool.

Example 2: ANOVA Interaction Effect

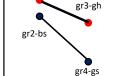








$$gr2 > gr1$$
 $gr2 > gr4$ 



gr1-bh

Example 2: ANOVA Interaction Effect

```
sesamesim$gr <- as.factor(sesamesim$gr)</pre>
anov <- lm(postnumb~gr-1, sesamesim)
results <- bain(anov,
"gr2 - gr1 > gr4 - gr3 & gr2 > gr1 & gr2 > gr4")
```

Example 2: ANOVA Interaction Effect

#### The main output is

```
Fit
      Com BF.u BF.c PMPa PMPb PMPc
H1 0.922 0.283 3.262 29.984 1.000 0.765 0.968
H11
                                 0.235
Hc 0.078 0.717 0.109
                                      0.032
```

#### Hypotheses:

H1: qr2-qr1>qr4-qr3&qr2>qr1&qr2>qr4

How to write down an hypothesis

bain can handle hypotheses build using constraints on (linear combinations) of parameters. Suppose the parameter names are "a". "b". "c".

Step 1: Construct the elements of the linear combination. E.g. "a" or "a + 2" or "3 \* a" or "2 \* a + 4"

Step 2: Constrain the resultsing elements. E.g. a > b > c

or 
$$a > b + 2 \& b > c + 2$$

or 
$$2 * a > b + c & b > 0 & c > 0$$

or 
$$a > (b, c) \& b - c > 0$$

Example 3: Repeated Measures

#### Development of depression

	Measurement					
	8 years	12 years 16 years 20 yea				
Men	$\mu_1$	$\mu_2$	$\mu_3$	$\mu_{4}$		
Women	$\mu_5$	$\mu_{6}$	$\mu_7$	$\mu_{8}$		

$$H_1: \mu_5 - \mu_1 > \mu_6 - \mu_2 > \mu_7 - \mu_3 < \mu_8 - \mu_4$$

$$H_2: \mu_6 - \mu_5 < \mu_7 - \mu_6 > \mu_8 - \mu_7$$

Example 4: Multiple Regression

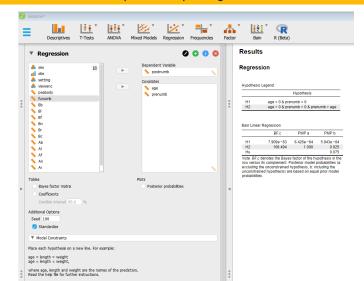
$$\mathsf{postnumb}_i = \beta_0 + \beta_1 \times \mathsf{age}_i + \beta_2 \times \mathsf{prenumb}_i + \epsilon_i$$

$$H_1: \beta_1 > 0, \beta_2 > 0, \beta_1 < \beta_2$$

Note:  $\beta_1$  and  $\beta_2$  are only comparable if age and prenumb are standardized



Example 4: Multiple Regression



**Example 5: About Equality Constraints** 

Is the difference in number knowledge relevantly different between boys and girls?

**Example 5: About Equality Constraints** 

```
sesamesim$sex <- as.factor(sesamesim$sex)</pre>
anov <- lm(postnumb~sex-1, sesamesim)</pre>
results \leftarrow bain(anov, "-2 < sex1 - sex2 < 2")
```

**Example 5: About Equality Constraints** 

```
sex1 sex2
30.09565 28.85600
```

```
Fit Com BF.u BF.c PMPa PMPb PMPc
H1 0.664 0.091 7.304 19.735 1.000 0.880 0.952
Hu 0.120
Hc 0.336 0.909 0.370 0.048
```

#### Hypotheses:

H1: -2 < sex1 - sex2 < 2

Example 6: Structural Equation Modelling

```
library (bain)
library(lavaan)
model <- '
    A = \sim Ab + Al + Af + An + Ar + Ac
    B = Rb + Bl + Bf + Bn + Br + Bc
    A ~ B + age + peabody'
fit <- sem(model, data = sesamesim, std.lv = TRUE)
hypotheses <- "A~B = A~peabody = A~age = 0;
                A \sim B > A \sim peabody > A \sim age = 0"
set.seed(100)
y1 <- bain(fit, hypotheses, standardize = TRUE)</pre>
```



End ●00

ANOVA and Beyond, Example Analyses with R and JASP

End

# Your hypothesis of interest

End

If you have your own data

#### Before:

- What is your research question?
- What is your theory / expectation?
- What is your statistical hypothesis?
- Is there a competing statistical hypothesis?

### Additionally:

- Are you able to specify your statistical hypothesis/-es?
- How will you evaluate it/them? (preference GORIC(A) or BMS?)

End

## Thanks for listening!

## Are there any questions?

#### Websites

https://github.com/rebeccakuiper/Tutorials www.uu.nl/staff/RMKuiper/Extra1 www.uu.nl/staff/RMKuiper/Extra2 informative-hypotheses.sites.uu.nl/software/goric/

#### E-mail

r.m.kuiper@uu.nl



### What's next

End

### Depending on time and wishes:

- Some extra information
- Demo in R
- Demo in JASP

#### We end with:

• Lab







### **Table of Contents**

ANOVA and Beyond, Example Analyses with R and JASP

Extra

comparing two informative hypotheses

The BF quantifies the relative support in the data for two hypotheses.

$$BF_{12} = \frac{BF_{1u}}{BF_{2u}} = \frac{f_1/f_2}{c_1/c_2}$$

using

$$BF_{iu} = \frac{f_i/f_u}{c_i/c_u} = \frac{f_i}{c_i}$$

Three Simple Hypotheses

### Consider the hypotheses:

$$H_1: \mu_1 \approx \mu_2$$
, that is,  $|\mu_1 - \mu_2| < .1$ 

$$H_2: \mu_1 > \mu_2$$

$$H_3: \mu_1, \mu_2$$

Information in the Data about the Two Means

					95% Credible Interval	
	Ν	Mean	SD	SE	Lower	Upper
sex1	115	30.096	13.058	1.175	27.793	32.398
sex2	125	28.856	12.162	1.127	26.647	31.065

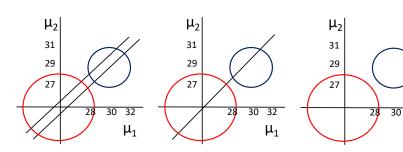
$$g(\mu_1,\mu_2\mid \text{data}) \approx \mathcal{N}\left(\left[\begin{array}{c} \textit{m}_1 \\ \textit{m}_2 \end{array}\right], \left[\begin{array}{cc} \textit{se}_1^2 = \frac{\textit{SD}_1^2}{\textit{N}_1} & \textit{0} \\ \textit{0} & \textit{se}_2^2 = \frac{\textit{SD}_2^2}{\textit{N}_2} \end{array}\right]\right),$$

Posterior Distribution (f), Prior Distribution (c), and Hypotheses

$$H_1$$
:  $\mu_1 \approx \mu_2$ 

$$H_2$$
:  $\mu_1 > \mu_2$ 

$$H_u$$
:  $\mu_1$ ,  $\mu_2$ 



$$BF_{1u} = f_1/c_1 = .25/.05 = 5$$
  $BF_{2u} = f_2/c_2 = .75/.5 = 1.5$   $BF_{12} = 5/1.5 = 3.33$ 

32

 $\mu_1$ 

Fit and Complexity

- The fit of a hypothesis is the proportion of the posterior distribution in agreement with the hypothesis.
   Note: posterior = likelihood × prior.
- The complexity of a hypothesis is the proportion of the prior distribution in agreement with the hypothesis.

The Prior Distribution

$$h(\mu_1, \mu_2 \mid \text{data}) \approx \mathcal{N}\left(\left[\begin{array}{c} m \\ m \end{array}\right], \left[\begin{array}{cc} \frac{SD_1^2}{J} & 0 \\ 0 & \frac{SD_2^2}{J} \end{array}\right]\right),$$

where  $\mu_1$  and  $\mu_2$  have the same prior mean m, and where *J* denotes the size of the training sample.

Possible choices for *J* for the example at hand:

- use the default in bain: J = the number of independent constraints, here, 1. This is a conservative choice. Sensitivity check: e.g., J, 2\*J, and 3\*J (fraction = 1, 2, and 3, resp.).
- or use J = the minimal training sample size, here, 4 because four observations are needed to estimate two means and two variances.
- or use J = Jref, which renders  $BF_{0u} = 19$  if the effect size in the sample equals 0.

Posterior Distribution (f), Prior Distribution (c), and Hypotheses Prior Sensitivity for = Constrained Hypotheses

Here, used bain: J = 1.

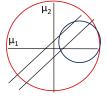
Sensitivity check: J, 2\*J, and 3\*J (i.e., fraction = 1, 2, and 3).

$$H_1$$
:  $\mu_1 \approx \mu_2$ 

$$J = 1$$

$$2*J = 2$$

$$3*J = 3$$







$$BF_{111} = .2/.01 = 20$$
  $BF_{111} = .2/.05 = 4$ 

$$BF_{1.1} = .2/.05 = 4$$

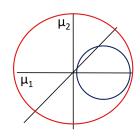
$$BF_{1u} = .2/.2 = 1$$

Posterior Distribution (f), Prior Distribution (c), and Hypotheses Prior In-Sensitivity for > < Constrained Hypotheses

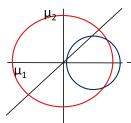
$$H_2$$
:  $\mu_1 > \mu_2$ 

$$2*J = 2$$

$$3*J = 3$$



 $BF_{2u} = .9/.5 = 1.8$ 



$$BF_{2u} = .9/.5 = 1.8$$

$$\mu_1$$
  $\mu_2$ 

$$BF_{2u} = .9/.5 = 1.8$$