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

•000000

bain

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

$$BF_{iu} = rac{f_i}{c_i} = rac{ ext{fit } H_i}{ ext{complexity } H_i}$$

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.

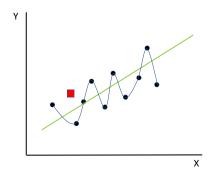
$$H_2: \mu_1 > (\mu_2, \mu_3)$$

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

$$BF_{iu}$$
 evaluating H_i versus H_u : μ_1, μ_2, μ_3

$$BF_{ii'}$$
 evaluating H_i versus $H_{i'}$: $\mu_1 = \mu_2 = \mu_3$

 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 BFiu
- 2. Compare two competing hypotheses using $BF_{ii'}$
- 3. Compare "my theory" with "not my theory" using BF_{ic}

	f _i	C_i	BF_{iu}	BF_{ic}
H ₁ : 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_4 : 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

$$H_4: (\mu_2, \mu_4) > (\mu_1, \mu_3, \mu_5, \mu_6)$$
 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)
- 5. "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?

- Guidelines by Jeffreys (1969) and Kass and Raftery (1995), e.g., < 3 is ignorable, > 3 is positive evidence, > 10 is strong evidence ...
- Will lead to a return of sloppy science and publication bias (when used without pre-registration or a pre-registered report)
- 3. Were does the 3 come from?

Interpreting (the Size of) the Bayes Factor

When is the Bayes factor large enough?

- Before collecting or accessing the data, formulate informative hypotheses (and decide how large you would like the Bayes factor to be).
- 2. Insert this information in a 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.

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 _i	Ci	BF _{iu}	PMP_i	PRI_i
H ₁ : Sex Match	.0039	.012	.32	.04	1/5
H_2 : Gender Role Match	.0725	.012	5.85	.81	1/5
H ₃ : Sex Mismatch	.0007	.012	.06	.01	1/5
H ₄ : Gender Role Mismatch	.0001	.012	.01	.00	1/5
H_u :				.14	1/5

Note on hypotheses

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_u combines H_i and H_c .

Replacing Hu by Hc

	f _i	C_i	BF_{iu}	PMP_i	PRI_i
H ₁ : Sex Match	.0039	.012	.32	.04	1/5
H_2 : Gender Role Match	.0725	.012	5.85	.84	1/5
H ₃ : Sex Mismatch	.0007	.012	.06	.00	1/5
H ₄ : Gender Role Mismatch	.0001	.012	.01	.00	1/5
_ <i>H_c</i> :	.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".



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 _i	Ci	BF _{iu}	PMP_i	PRI_i
H ₁ : Sex Match	.0039	.012	.32	.013	1/13
H_2 : Gender Role Match	.0725	.012	5.85	.270	1/13
H ₃ : Sex Mismatch	.0007	.012	.06	.003	1/13
H ₄ : 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 _i	Ci	BF_{iu}	PMP_i	PRI_i
H ₁ : Sex Match	.0039	.012	.32	.04	1/5
H ₂ : Gender Role Match	.0725	.012	5.85	.81	1/5
H ₃ : Sex Mismatch	.0007	.012	.06	.01	1/5
H ₄ : 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_{ii} :	.0104	.012	1.00	.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

Table of Contents

Bayesian Informative Hypotheses Evaluation (bain)

ANOVA and Beyond, Example Analyses with R and JASP

Enc

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, 2 = advantaged suburban, 3 = 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)
```

R and JASP 00000000

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
      site1 60 29.66667 26.82991 32.50343
      site2 55 38.98182 36.01892 41.94472
3
      site3 64 23.18750 20.44082 25.93418
      site4 43 25.32558 21.97466 28.67650
5
      site5 18 31.72222 26.54303 36.90141
```

Example 1: ANOVA

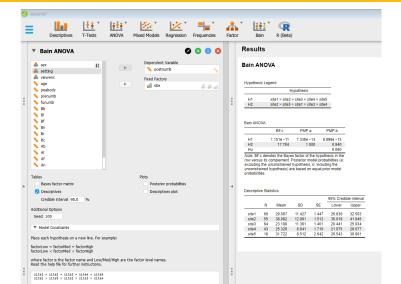
The main output is

```
Fit
         Com
              BF.u BF.c PMPa PMPb
                                        PMPc.
  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
H11
                                   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



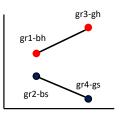
Example 2: ANOVA Interaction Effect

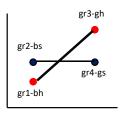
Dependent variable: Knowledge of numbers.

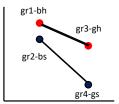
Factors: sex (boy, girl) and setting (watching at home, watching at school).

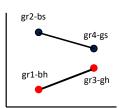
Gr: 1=boyhome, 2= boyschool, 3= girlhome, 4=girlschool.

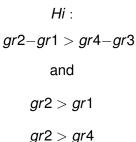
Example 2: ANOVA Interaction Effect











Example 2: ANOVA Interaction Effect

```
sesamesim$gr <- as.factor(sesamesim$gr)
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
                                  0.235
H11
Hc 0.078 0.717 0.109
                                        0.032
```

Hypotheses:

```
H1: gr2-gr1>gr4-gr3&gr2>gr1&gr2>gr4
```

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 year					
Men	μ_1	μ_2	μ_3	μ_{4}			
Women	μ_5	μ_{6}	μ_7	μ_{8}			

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

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

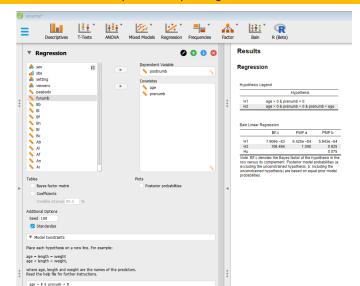
Example 4: Multiple Regression

$$postnumb_i = \beta_0 + \beta_1 \times age_i + \beta_2 \times prenumb_i + \epsilon_i$$

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

Note: β_1 and β_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)
results \leftarrow bain(anov, "-2 < sex1 - sex2 < 2")
```

Informative Hypotheses

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

Informative Hypotheses

Example 6: Structural Equation Modelling

```
library (bain)
library (lavaan)
model <- '
    A = 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>
```

Hands-on/Demo: BMS

Let's practice. Go to

https://github.com/rebeccakuiper/Tutorials:

- 1. Click on green button called Code.
- 2. Download zip (last option in list).
- 3. Unzip it on your machine (that folder is now your working directory).
- Start Rstudio. 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 (Im()).
- Specify hypotheses (make up your own).
 Note: Use names used in the model.
- Run bain().
- Inspect and interpret output.



Table of Contents

End •00

ANOVA and Beyond, Example Analyses with R and JASP

End

The End **BMS**

Thanks for listening!

Are there any questions?

Websites

https://github.com/rebeccakuiper/Tutorials www.uu.nl/staff/RMKuiper/Software www.uu.nl/staff/RMKuiper/Websites%20%2F% 20Shiny%20apps informative-hypotheses.sites.uu.nl/software/ goric/

What's next

End

- Model selection using information criteria
- Possibly: Evidence synthesis / Support aggregation

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

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

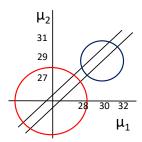
$$g(\mu_1,\mu_2\mid ext{data}) pprox \mathcal{N} \left(\left[egin{array}{c} m_1 \\ m_2 \end{array}
ight], \left[egin{array}{ccc} se_1^2 = rac{SD_1^2}{N_1} & 0 \\ 0 & se_2^2 = rac{SD_2^2}{N_2} \end{array}
ight]
ight),$$

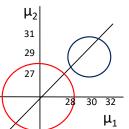
Posterior Distribution, Prior Distribution, and Hypotheses

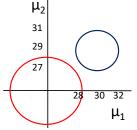
$$H_1: \mu_1 \approx \mu_2$$

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

$$\boldsymbol{H}_{u}\!:\boldsymbol{\mu}_{1}$$
 , $\boldsymbol{\mu}_{2}$







$$BF_{1u} = f_1/c_1 = .25/.05 = 5$$

$$BF_{11} = f_1/c_1 = .25/.05 = 5$$
 $BF_{21} = f_2/c_2 = .75/.5 = 1.5$

$$BF_{12} = 5/1.5 = 3.33$$

A Closer Look at the Bayes Factor Fit and Complexity

- 1. The fit of a hypothesis is the proportion of the posterior distribution in agreement with the hypothesis.
- 2. 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 μ_1 and μ_2 have the same prior mean m, and where J denotes the size of the training sample.

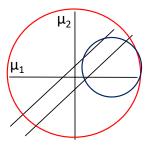
Choices for J for the example at hand:

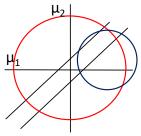
- Default in bain: number of independent constraints, that is, 1, this is a conservative choice (NB. .5 * J)
- Minimal training sample size, that is, 4, because four observations are neede to estimate two means and variances
- *Jref*, which renders $BF_{0u} = 19$ if the effect size in the sample equals 0

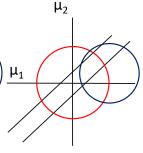
Prior Sensitivity for = Constrained Hypotheses

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

$$J = 3$$







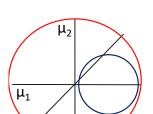
$$BF_{111} = .2/.01 = 20$$
 $BF_{111} = .2/.05 = 4$

$$BF_{1u} = .2/.05 = 4$$

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

Prior In-Sensitivity for > < Constrained Hypotheses

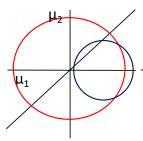
$$H_2$$
: $\mu_1 > \mu_2$



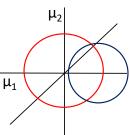
J=1

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

J=2



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



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

Bayes Factor (BF)

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

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