Introduction to Bayes Factors with mixed effects logistic regression

Maša Vujović
LiLaC Lab Meeting
1st November 2018

**MODIFIED BAYES' THEOREM:**

\[ P(H|X) = P(H) \times \left( 1 + P(C) \times \left( \frac{P(x|H)}{P(x)} - 1 \right) \right) \]

- **H**: HYPOTHESIS
- **X**: OBSERVATION
- **P(H)**: PRIOR PROBABILITY THAT H IS TRUE
- **P(x)**: PRIOR PROBABILITY OF OBSERVING X
- **P(C)**: PROBABILITY THAT YOU'RE USING BAYESIAN STATISTICS CORRECTLY
Some disclaimers

Work in progress at Language Learning Lab with input from Prof Zoltan Dienes

* Some published work used this method: see http://languagelearninglab-ucl.com for preprints

- Not an introduction to Bayesian statistical modelling
- Combining a Bayesian statistical inference method with mixed-effects logistic regression models

All mistakes today my own!

01/11/2018
Bayes Factor is a measure of strength of evidence

Strength of evidence = amount by which your prior confidence in H1 over H0 ought to change having seen the data
Bayes Factor is a measure of strength of evidence

\[ B = \frac{P(D \mid H1)}{P(D \mid H0)} \]

- If \( B = \) about 1, experiment was not sensitive
- If \( B > 1 \) then the data supported your theory over the null
- If \( B < 1 \), then the data supported the null over your theory

Jeffreys (1961):
- \( B < 0.1 \) – strong evidence for \( H0 \)
- \( B < 0.33 \) – substantial evidence for \( H0 \)
- \( B > 10 \) – strong evidence for \( H1 \)
- \( B > 3 \) suggest substantial evidence for \( H1 \)
- between 0.33 and 3 – inconclusive evidence
Why choose BF over p-values?

• Unlike frequentist hypothesis testing, can give support for the null

• Unlike p-values, BF are not sensitive to optional stopping (Rouder, 2014)

• Differences between Bayes Factors are meaningful – easier to interpret than p-values

• For me personally, BF gets me to engage more with my effects of interest:
  • What is my prior belief?
  • Where does it come from?
  • What data would convince me otherwise?
Computing Bayes Factors

Needs two kinds of information:

1. Model of the data (mean difference between conditions and the standard error) – observed values
2. Model of the H1 (your prediction)

There is an R function/Bf Calculator which does this for you!
Bayes Factor Calculator in R

• Bf R function equivalent to the Dienes (2008) calculator which can be found here: http://www.lifesci.sussex.ac.uk/home/Zoltan_Dienes/inference/Bayes.htm

• The code was provided by Baguely and Kayne (2010) and can be found here: http://www.academia.edu/427288/Review_of_Understanding_psychology_as_a_science_An_introduction_to_scientific_and_statistical_inference
BF with Mixed Models in R

```r
## Prediction 1: Affix by type frequency interaction

gen_1b_lme = glmer(correct ~ affix.ct*type_frequency.ct + (type_frequency.ct|participant),
control=glmerControl(optimizer = "bobyqa"), data = gen_1b, family = binomial)
round(summary(gen_1b_lme)$coefficients, 3)
```

|              | Estimate | Std. Error | z value | Pr(>|z|) |
|--------------|----------|------------|---------|----------|
| (Intercept)  | 0.479    | 0.095      | 5.043   | 0.000    |
| affix.ct     | 0.033    | 0.188      | 0.177   | 0.860    |
| type_frequency.ct | 0.381    | 0.152      | 2.503   | 0.012    |
| affix.ct:type_frequency.ct | -0.646   | 0.299      | -2.159  | 0.031    |
BF with Mixed Models in R

```
{r}
h1 = summary(gen_1_lme)$coefficients["affix.ct:type_frequency.ct", "Estimate"]*1
se = summary(gen_1b_lme)$coefficients["affix.ct:type_frequency.ct", "Std. Error"]
mean = summary(gen_1b_lme)$coefficients["affix.ct:type_frequency.ct", "Estimate"]*1
Bf(sd = se, obtained = mean, uniform = 0, sdtheory = h1, meanoftheory = 0, tail=1)
```

sd – standard error from the LME [se]
obtained – beta estimate from the LME [mean]
sd theory – predicted effect size (here: beta estimate from a corresponding LME with pilot data) [h1]

mean of theory – 0
uniform – 0 (or 1 if using a uniform prior)
tail – 1 or 2 depending on whether one- or two-tailed

Refer to Dienes 2014 for theoretical implications of using different distributions
A note on signs...

- Bf calculator does not allow negative H1
- One way of getting round this: multiply H1 and mean by -1:

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Multiply by -1</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>h1 negative, observed value negative</td>
<td>-h1 * -1 = h1</td>
<td>Both values same sign (as found)</td>
</tr>
<tr>
<td></td>
<td>-mean * -1 = mean</td>
<td></td>
</tr>
<tr>
<td>h1 negative, observed value positive</td>
<td>-h1 * -1 = h1</td>
<td>Values opposite sign (as found)</td>
</tr>
<tr>
<td></td>
<td>mean * -1 = -mean</td>
<td></td>
</tr>
</tbody>
</table>
And now...

The question you’ve been dying to ask!
Where does the prior come from?

From the literature

From a pilot experiment

You come up with a plausible maximum effect

and more...
Coming up with plausible maxima

• One of my studies:
  • Language learning study – participants are trained on an artificial language and then tested on what they learn
  • DV: accuracy at test
  • IV 1: affix – two levels: whether participants are learning a suffixing or a prefixing language
  • IV 2: type frequency – two levels: whether the words I test them on were high or low frequency in learning input
Coming up with plausible maxima

- I predict an affix-by-type-frequency interaction. Specifically:
  - Suffix condition should be above chance on both high and low type-frequency items
  - Prefix condition should be above chance on high, but chance-level on low type-frequency items

- What is the plausible maximum here?
  - All of the type frequency effect is carried by the prefix condition
  - Maximum corresponds to the main effect of type frequency
Coming up with plausible maxima

\[ a \times b \text{ interaction} = 2 \times \text{main effect of } a \text{ or main effect of } b \text{ (depending on theoretical interest)} \]

<table>
<thead>
<tr>
<th></th>
<th>PERCENT</th>
<th>LOG ODDS</th>
<th>LOGG_ODDS*2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prefix</td>
<td>HF</td>
<td>90</td>
<td>2.197</td>
</tr>
<tr>
<td>Prefix</td>
<td>LF</td>
<td>50</td>
<td>0.000</td>
</tr>
<tr>
<td>Suffix</td>
<td>HF</td>
<td>90</td>
<td>2.197</td>
</tr>
<tr>
<td>Suffix</td>
<td>LF</td>
<td>90</td>
<td>2.197</td>
</tr>
<tr>
<td><strong>Affix</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Type_frequency</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Interaction</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Coming up with plausible maxima

• Let’s say I am interested in a main effect of affix:
  ➢ I predict that the Suffix condition will be better than the Prefix condition

• What is the plausible maximum?
  • All learning happens in suffix condition, no learning happens in prefix condition
  • If so, main effect of affix corresponds to the intercept (*if intercept reflects overall learning rather than one baseline condition)
Coming up with plausible maxima

• In general, the maximum is $2 \times \text{“one level up”}$:
  • main effect $\rightarrow 2 \times \text{intercept}$
  • 2-way interaction $\rightarrow 2 \times \text{main effect OR } 4 \times \text{intercept}$
  • And so on

• You could therefore use these values from previous data OR current data
  • Note: recommended you use independent data wherever possible to model your H1
References

