

Exercises for Bayesian Hierarchical models

Bruno Nicenboim and Shravan Vasishth

Hierarchical model with a lognormal likelihood:

We begin with a classic question from the psycholinguistics literature: are subject relatives easier to process than object relatives? The data come from Experiment 1 in a paper by Grodner and Gibson (2005).

Scientific question: Is there a subject relative advantage in reading?

In two important papers, Gibson (2000) and Grodner and Gibson (2005) suggest that object relative clause sentences are more difficult to process than subject relative clause sentences because the distance between the relative clause verb *sent* and the head noun phrase of the relative clause, *reporter*, is longer in object vs subject relatives. Examples are shown below.

(1a) The *reporter* who the photographer *sent* to the editor was hoping for a good story. (ORC)

(1b) The *reporter* who *sent* the photographer to the editor was hoping for a good story. (SRC)

The underlying explanation has to do with memory processes: shorter linguistic dependencies are easier to process due to either reduced interference or decay, or both. For implemented computational models that spell this point out, see Lewis and Vasishth (2005) and Engelmann, Jäger, and Vasishth (2018).

In the Grodner and Gibson data, the dependent measure is reading time at the relative clause verb, in milliseconds. We are expecting longer reading times in object gap sentences compared to subject gap.

Reading the data and basic pre-processing

```
library(dplyr)
library(readr)
library(ggplot2)
library(brms)
## Parallelize the chains using all the cores:
options(mc.cores = parallel::detectCores())
library(bayesplot)

gg05_data <- read_csv("data/GrodnerGibson2005E1.csv") %>%
  filter(item != 0) %>%
  mutate(word_positionnew = if_else(item != 15 &
                                     word_position > 10,
                                     word_position-1,
                                     word_position))

#there is a mistake in the coding of word position,
#all items but 15 have regions 10 and higher coded
#as words 11 and higher

## get data from relative clause verb:
rc_data <- gg05_data %>%
  filter((condition == "objgap" & word_position == 6 ) |
         ( condition == "subjgap" & word_position == 4 ))
```

You should use a sum coding for the predictors. Here, object gaps are coded +1, subject gaps -1.

```
rc_data <- rc_data %>% mutate(ccond = if_else(condition == "objgap", 1, -1))
```

You should be able to now fit a maximal model (correlated varying intercept and slopes for subjects and items) assuming a lognormal likelihood, and examine the effect of relative clause attachment site (the predictor `ccond`) on reading times `rawRT`.

References

- Engelmann, Felix, Lena A. Jäger, and Shravan Vasishth. 2018. “The Effect of Prominence and Cue Association in Retrieval Processes: A Computational Account.”
- Gibson, Edward. 2000. “Dependency Locality Theory: A Distance-Based Theory of Linguistic Complexity.” In *Image, Language, Brain: Papers from the First Mind Articulation Project Symposium*, edited by Alec Marantz, Yasushi Miyashita, and Wayne O’Neil. Cambridge, MA: MIT Press.
- Grodner, Daniel, and Edward Gibson. 2005. “Consequences of the Serial Nature of Linguistic Input.” *Cognitive Science* 29: 261–90.
- Lewis, Richard L., and Shravan Vasishth. 2005. “An Activation-Based Model of Sentence Processing as Skilled Memory Retrieval.” *Cognitive Science* 29: 1–45.