Exercises for Bayesian Hierarchical models

Bruno Nicenboim and Shravan Vasishth

1. By-participants and by-items N400 model

Everything I said about by-participants group level (or random effects) is also relevant for by-items. Fit a by-participants and by-items maximal model.

2. 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 (ORC) sentences are more difficult to process than subject relative clause (SRC) sentences because the distance between the relative clause verb (e.g., "*sent*" in the example below) and the head noun phrase of the relative clause (e.g., "*reporter*" below) is longer in ORC vs SRC. 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, (e.g., "*sent*") of different sentences with either ORC or SRC. The dependent variable is in milliseconds and was measured in a self-paced reading task. (Self-paced reading is a task where participants read a sentence or a short text word-by-word or phrase-by-phrase, pressing a button to get the next word or phrase displayed while the previous one disappears; Aaronson and Scarborough 1976; Mitchell and Green 1978). For this experiment, we are expecting longer reading times at the relative close verbs of ORC sentences in comparison to the relative close verb of SRC sentences.

```r
library(dplyr)
library(readr)
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(c_cond = 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 `c_cond`) on reading times `rawRT`. For the population-level parameters, try to use the same priors that we elicited in our log-normal regression. Is the prior for β ($\beta \sim Normal(0, .1)$) appropriate? Do a sensitivity analysis. What is the estimate of the effect (β) under different priors? What is the difference in milliseconds between conditions under different priors?

References

Aaronson, Doris, and Hollis S. Scarborough. 1976. "Performance Theories for Sentence Coding: Some Quantitative Evidence." *Journal of Experimental Psychology: Human Perception and Performance* 2 (1): 56–70. https://doi.org/10/bjfzn4.

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.

Mitchell, Don C, and D. W. Green. 1978. "The Effects of Context and Content on Immediate Processing in Reading." *Quarterly Journal of Experimental Psychology* 30 (4): 609–36. https://doi.org/10.1080/14640747808400689.