Assignment 6

Nicole Cruz, Julia Haaf, Shravan Vasishth

Are subject relatives easier to process than object relatives (log-normal likelihood)?

This is 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.[1]

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

Grodner and Gibson 2005 investigate an old claim in psycholinguistics that object relative clause (ORC) sentences are more difficult to process than subject relative clause (SRC) sentences. One explanation for this predicted difference is that the distance between the relative clause verb (*sent* in the example below) and the head noun phrase of the relative clause (*reporter* in the example below) is longer in ORC vs. SRC. Examples are shown below. The relative clause is shown in square brackets.

(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.

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 subjects read a sentence or a short text word-by-word or phrase-by-phrase, pressing a button to get each word or phrase displayed; the preceding word disappears every time the button is pressed.

For this experiment, we are expecting longer reading times at the relative clause verbs of ORC sentences in comparison to the relative clause verb of SRC sentences.

```
load("data/df_gg05_rc.rda")
head(df_gg05_rc)
```

```
## # A tibble: 6 x 7
##    subj  item condition    RT residRT qcorrect experiment
##   <int> <int> <chr>     <int>   <dbl>    <int> <chr>
## 1     1     1 objgap      320  -21.4         0 tedrg3
## 2     1     2 subjgap     424   74.7         1 tedrg2
## 3     1     3 objgap      309  -40.3         0 tedrg3
## 4     1     4 subjgap     274  -91.2         1 tedrg2
## 5     1     5 objgap      333   -8.39        1 tedrg3
## 6     1     6 subjgap     266  -87.3         1 tedrg2
```

You should use a sum coding for the predictors. Here, object relative clauses (`"objgaps"`) are coded $+1/2$, subject relative clauses $-1/2$.

---

[1]Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentenial complexity. Cognitive Science, 29(2), 261-290.

```
df_gg05_rc <- df_gg05_rc %>%
  mutate(c_cond = if_else(condition == "objgap", 1/2, -1/2))
```

You should be able to now fit a "maximal" model (correlated varying intercept and slopes for subjects and for items) assuming a log-normal likelihood.

(a) Specify the statistical model and define priors for all parameters in the model. *Hint:* To assess reasonable ranges of values for the mean and sigma parameter of a lognormal distribution, you might want to simulate lognormal data for different values using `rlnorm()`

(b) Examine the effect of relative clause attachment site (the predictor `c_cond`) on reading times `RT` (β_1).

(c) Estimate the median difference between relative clause attachment sites in milliseconds, and report the mean and 95% CI.

(d) Inspect individual differences in the subject relative advantage effect. Are there participants with effects in opposite directions?

(e) Do a sensitivity analysis. What is the estimate of the effect (β_1) under different priors? What is the difference in milliseconds between conditions under different priors?