

# 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.<sup>1</sup>

*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.

<sup>1</sup>Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentential 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.

- 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()`
- Examine the effect of relative clause attachment site (the predictor `c_cond`) on reading times `RT` ( $\beta_1$ ).
- Estimate the median difference between relative clause attachment sites in milliseconds, and report the mean and 95% CI.
- Inspect individual differences in the subject relative advantage effect. Are there participants with effects in opposite directions?
- Do a sensitivity analysis. What is the estimate of the effect ( $\beta_1$ ) under different priors? What is the difference in milliseconds between conditions under different priors?

## Solution

(a)

We are going to fit the data with the following log-normal likelihood:

$$Y_{ijk} \sim \text{LogNormal}(\beta_1 + \delta_{i,0} + \omega_{k,0} + X_j \cdot (\beta + \delta_{i,1} + \omega_{k,1}), \sigma) \quad (1)$$

We should pay attention to use a not too tight prior for  $\beta_1$ :

- Priors:

$$\begin{aligned} \beta_0 &\sim \text{Normal}(5.5, 0.8) \\ \beta_1 &\sim \text{Normal}(0, 0.1) \\ \sigma &\sim \text{Normal}_+(0, 1) \end{aligned} \quad (2)$$

Here, we will need priors for the group-level parameters. Given that we assume a correlation between by-subject intercept and slope, and by-item intercept and slope, our model has the following structure which requires us to assign priors to  $\Sigma_\delta$  and  $\Sigma_\omega$

$$\begin{aligned} \begin{pmatrix} \delta_{i,0} \\ \delta_{i,1} \end{pmatrix} &\sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_\delta\right) \\ \begin{pmatrix} \omega_{k,0} \\ \omega_{k,1} \end{pmatrix} &\sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_\omega\right) \end{aligned} \quad (3)$$

$$\begin{aligned} \Sigma_\delta &= \begin{pmatrix} \sigma_{\delta_0}^2 & \rho_\delta \sigma_{\delta_0} \sigma_{\delta_1} \\ \rho_\delta \sigma_{\delta_0} \sigma_{\delta_1} & \sigma_{\delta_1}^2 \end{pmatrix} \\ \Sigma_\omega &= \begin{pmatrix} \sigma_{\omega_0}^2 & \rho_\omega \sigma_{\omega_0} \sigma_{\omega_1} \\ \rho_\omega \sigma_{\omega_0} \sigma_{\omega_1} & \sigma_{\omega_1}^2 \end{pmatrix} \end{aligned} \quad (4)$$

In practice this means that we need priors for the by-subject and by-item variances and correlations:

$$\begin{aligned} \sigma_{\delta_0} &\sim \text{Normal}_+(0, 1) \\ \sigma_{\delta_1} &\sim \text{Normal}_+(0, 1) \\ \rho_\delta &\sim \text{LKJcorr}(2) \\ \sigma_{\omega_0} &\sim \text{Normal}_+(0, 1) \\ \sigma_{\omega_1} &\sim \text{Normal}_+(0, 1) \\ \rho_\omega &\sim \text{LKJcorr}(2) \end{aligned} \quad (5)$$

Read the data:

We're ready to fit a model now:

```
fit_df_gg05_rc <- brm(RT ~ c_cond + (c_cond | subj) + (c_cond | item),
  family = lognormal(),
  iter = 10000,
  prior =
    c(
      prior(normal(5.5, 0.7), class = Intercept),
      prior(normal(0, .1), class = b),
      prior(normal(0, 1), class = sigma),
      prior(normal(0, 1), class = sd),
      prior(lkj(2), class = cor)
    ),
  data = df_gg05_rc,
  control=list(adapt_delta=0.99, max_tredepth=15)
)
```

(b) Examine the effect of relative clause attachment site (the predictor `c_cond`) on reading times RT in log-scale.

We'll focus on  $\beta$ :

```
posterior_summary(fit_df_gg05_rc, variable = "b_c_cond")
```

```
##           Estimate Est.Error      Q2.5      Q97.5
## b_c_cond 0.09776067 0.04589594 0.005170052 0.1867813
```

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

```
beta0 <- as_draws_df(fit_df_gg05_rc)$b_Intercept
beta1 <- as_draws_df(fit_df_gg05_rc)$b_c_cond
# Difference between object RC coded as .5 and subject RC coded as .5
effect <- exp(beta0 + beta1 * .5) - exp(beta0 + beta1 * -.5)
c(mean = mean(effect), quantile(effect, c(.025,.975)))

##      mean      2.5%      97.5%
## 35.016286  1.727591 69.074315
```

(c)

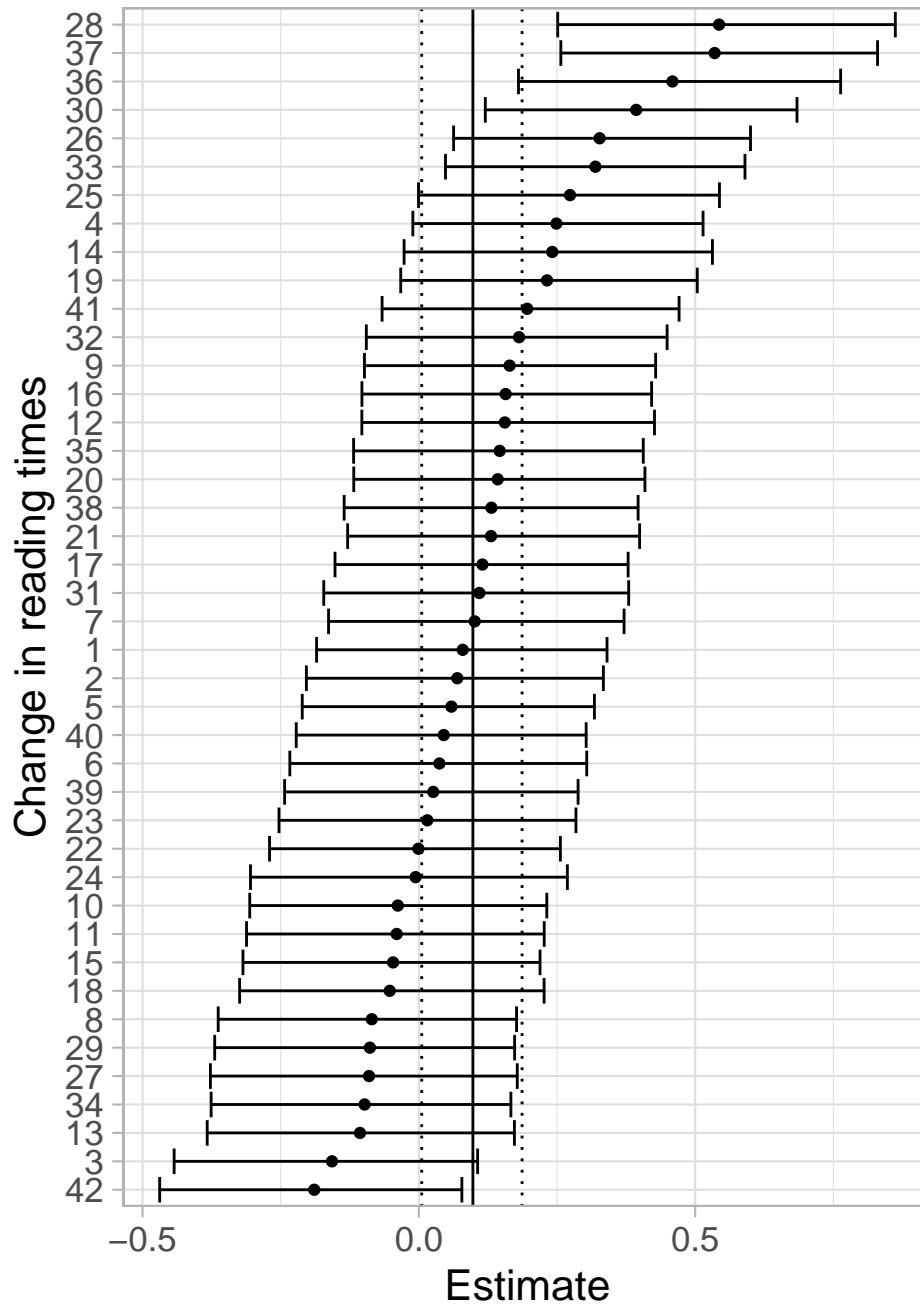
```
# We get the by subject effects in a data frame where each adjustment
# is in each column.
by_subj_effect <- data.frame(coef(fit_df_gg05_rc)$subj[, , "c_cond"]) %>%
  bind_rows() %>%
  # We add a column to identify that the model,
  # and one with the subject labels:
  mutate(
    subj = unique(df_gg05_rc$subj)) %>%
  arrange(Estimate) %>%
  mutate(subj = factor(subj, levels = subj))

ggplot(
```

```

by_subj_effect,
aes(
  ymin = Q2.5, ymax = Q97.5, x = subj, y = Estimate
)
) +
geom_errorbar() +
geom_point() +
# We'll also add the mean and 95% CrI of the overall difference
# to the plot:
geom_hline(
  yintercept =
    posterior_summary(fit_df_gg05_rc)["b_c_cond", "Estimate"]
) +
geom_hline(
  yintercept =
    posterior_summary(fit_df_gg05_rc)["b_c_cond", "Q2.5"],
  linetype = "dotted", linewidth = 0.5
) +
geom_hline(
  yintercept =
    posterior_summary(fit_df_gg05_rc)["b_c_cond", "Q97.5"],
  linetype = "dotted", linewidth = 0.5
) +
xlab("Change in reading times") +
coord_flip() +
theme_light(base_size = 16)

```



(d) Do a sensitivity analysis. What is the estimate of the effect ( $\beta$ ) under different priors? What is the difference in milliseconds between conditions under different priors?

Next, we do a sensitivity analysis using a tighter prior for  $\beta$ ,  $\beta_1 \sim \text{Normal}(0, 0.01)$ :

```
fit_df_gg05_rc2 <- brm(RT ~ c_cond + (c_cond | subj) + (c_cond | item),
  family = lognormal(),
  prior =
    c(
      prior(normal(5.5, 0.7), class = Intercept),
      prior(normal(0, 0.01), class = b),
      prior(normal(0, 1), class = sigma),
      prior(normal(0, 1), class = sd),
```

```

    prior(lkj(2), class = cor)
  ),
  data = df_gg05_rc,
  control=list(adapt_delta=0.99, max_treedepth=15)
)

```

```
posterior_summary(fit_df_gg05_rc2, variable = "b_c_cond")
```

```
##           Estimate  Est.Error      Q2.5      Q97.5
## b_c_cond 0.004034925 0.009752846 -0.01500466 0.02327195
```

And here we use the prior,  $\beta_1 \sim \text{Normal}(0, 1)$ :

```

fit_df_gg05_rc3 <- brm(RT ~ c_cond + (c_cond | subj) + (c_cond | item),
  family = lognormal(),
  prior =
    c(
      prior(normal(5.5, 0.7), class = Intercept),
      prior(normal(0, 1), class = b),
      prior(normal(0, 1), class = sigma),
      prior(normal(0, 1), class = sd),
      prior(lkj(2), class = cor)
    ),
  data = df_gg05_rc,
  control=list(adapt_delta=0.99, max_treedepth=15)
)

```

```
posterior_summary(fit_df_gg05_rc3, variable = "b_c_cond")
```

```
##           Estimate  Est.Error      Q2.5      Q97.5
## b_c_cond 0.1238191 0.05089789 0.02399779 0.2219486
```

We can summarize the estimates of  $\beta$  given different priors in the following way:

Estimate	Q2.5	Q97.5	Prior for $\beta$
0.0977607	0.0051701	0.1867813	Normal(0,.01)
0.0040349	-0.0150047	0.0232720	Normal(0,.1)
0.1238191	0.0239978	0.2219486	Normal(0,1)

We can take a look at the estimate for the difference between conditions in milliseconds:

```

beta0 <- as_draws_df(fit_df_gg05_rc)$b_Intercept
beta1 <- as_draws_df(fit_df_gg05_rc)$b_c_cond
diff1 <- exp(beta0 + beta1/2) - exp(beta0 - beta1/2)
c(mean = mean(diff1), quantile(diff1, c(.025, .975)))

```

```
##      mean      2.5%      97.5%
## 35.016286 1.727591 69.074315
```

If we repeat this with each model, we can take a look at the effect of the prior on the difference between conditions:

Estimate (ms)	Q2.5	Q97.5	Prior for $\beta_1$
35.016286	1.727591	69.074315	Normal(0,.01)
1.390188	-5.205193	7.944704	Normal(0,.1)

Estimate (ms)	Q2.5	Q97.5	Prior for $\beta_1$
44.846232	8.080434	82.749709	Normal(0,1)

This sensitivity shows us that the posterior changes quite a lot under different priors; depending on what we consider to be a reasonable prior (i.e., depending on our prior beliefs), the inference that we derive from the model and data can vary quite a bit when the data are relatively sparse, as in this case.