Exercise 2.1: Fitting Mixed Models Using lme4

July 2019 (updated: 2019-07-24)

For this exercise we will use the data of Skovgaard-Olsen et al. (2016) already used previously. We can load it directly. We also load package `lme4` for fitting some first mixed models.

```
load("ssk16_dat_prepared_ex2.rda")
str(dat2)
```

```
## 'data.frame':    752 obs. of  7 variables:
##  $ p_id         : Factor w/ 94 levels "102_P(if,then)",..: 63 63 63 63 63 63 63 63 64 64 ...
##  $ i_id         : Factor w/ 12 levels "1","2","3","4",..: 1 2 7 8 9 10 11 12 2 4 ...
##  $ B_given_A    : num  52 60 79 0 51 79 80 77 98 98 ...
##  $ B_given_A_c  : num  0.02 0.1 0.29 -0.5 0.01 0.29 0.3 0.27 0.48 0.48 ...
##  $ if_A_then_B  : num  52 1 84 0 51 94 1 81 95 97 ...
##  $ if_A_then_B_c: num  0.02 -0.49 0.34 -0.5 0.01 0.44 -0.49 0.31 0.45 0.47 ...
##  $ rel_cond     : Factor w/ 2 levels "PO","IR": 1 2 1 2 2 1 2 1 1 1 ...
```

```
library("lme4")
```

```
## Loading required package: Matrix
```

Variables in the data:

- `p_id`: participant identifier
- `i_id`: item identifier (i.e., id of vignette)
- `B_given_A`: original $P(B|A)$
- `B_given_A_c`: $P(B|A)$ centered at midpoint of scale (as used in paper)
- `if_A_then_B`: original $P(\text{if } A \text{ then } B)$
- `if_A_then_B_c`: $P(\text{if } A \text{ then } B)$ centered at midpoint of scale (as used in paper)
- `rel_cond`: relevance condition, irrelevant here.

Research question: Does the Equation (i.e., $P(\text{if } A \text{ then } B) = P(B|A)$) hold?

Task 1: Fit 2 Models

Our general goal is to use a partial pooling approach with this data using `lm4`. We want to recreate the first two models from the lecture, add one more model, and inspect the results.

- Dependent variable: `if_A_then_B_c`
- Fixed effect: `B_given_A_c`

1. The first model only has by-participant (i.e., by `p_id`) random-slopes for `B_given_A_c`.
2. The second model has has by-participant random-slopes for `B_given_A_c` and random-intercepts, as well as the correlation between the random-effects parameters.
3. The third (and new) model only has by-participant random intercepts and no random slopes.

```
# m1 <- lmer( ... , dat2)
# m2 <- lmer( ... , dat2)
# m3 <- lmer( ... , dat2)
```

Inspect the results of all models:

- Which model shows a singular fit or other convergence problems? Do you have an idea where the convergence problems come from?

- What do the parameter estimates of the random effects mean?
- Does the choice of random-effects structure affect the fixed-effect estimates and/or inferential statistics?
- Which model do you think would be the best to report?

Task 2: Suppressing the Correlation

The only model we have not yet considered for this data is a model without correlation among the random-effects parameters. Can you set it up?

```
# m4 <- lmer( ... , dat2)
```

Does this model show any benefit compared to one of the previous models?