04 Hierarchical linear modeling

Shravan Vasishth

27 June 2019

Example 1: Reading time differences in subject vs object relatives in English

We begin with a relatively simple 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).

Example 1: Reading time differences in subject vs object relatives in English

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

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

Example 1: Reading time differences in subject vs object relatives in English

Load data and reformat

```
gg05e1 <- read.table("data/GrodnerGibson2005E1.csv",
                     sep=",", header=T)
gge1 <- gg05e1 %>% filter(item != 0)

gge1 <- gge1 %>%
  mutate(word_positionnew = ifelse(item != 15 &
                                   word_position > 10,
                                   word_position-1,
                                   word_position))
```

Example 1: Reading time differences in subject vs object relatives in English

Load data and reformat

```
#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:
gge1crit <- subset(gge1, ( condition == "objgap" &
                             word_position == 6 ) |
            ( condition == "subjgap" & word_position == 4 ))
```

Example 1: Reading time differences in subject vs object relatives in English

Experiment design: Latin square and crossed subject and items

Two important properties of these data are worth noticing.

- Latin square design
- Crossed subjects and items

Example 1: Reading time differences in subject vs object relatives in English

Latin-square design

First, the design is the classic repeated measure Latin square set-up. To see what this means, first look at the number of subjects and items, and the number of rows in the data frame:

```
length(unique(gge1crit$subject))
```

```
## [1] 42
```

```
length(unique(gge1crit$item))
```

```
## [1] 16
```

```
dim(gge1crit)[1]
```

```
## [1] 672
```

There are 42 subjects and 16 items. There are $42 \times 16 = 672$ rows in the data frame.

Example 1: Reading time differences in subject vs object relatives in English

Latin-square design

Notice also that each subject sees exactly eight object gap and eight subject gap sentences:

```
head(xtabs(~subject+condition,gge1crit),n=4)
```

```
##         condition
## subject objgap subjgap
##       1      8       8
##       2      8       8
##       3      8       8
##       4      8       8
```

Example 1: Reading time differences in subject vs object relatives in English

Latin-square design

Table 1: The Latin-square design in repeated measures experiments.

item id	group 1	group 2
1	objgap	subjgap
2	subjgap	objgap
3	objgap	subjgap
4	subjgap	objgap
⋮	⋮	⋮
16	subjgap	objgap

Example 1: Reading time differences in subject vs object relatives in English

Latin-square design: Ensuring balance

```
sample(rep(c("order1","order2"),11))
```

```
##  [1] "order1" "order1" "order2" "order2" "order2" "order2"
##  [8] "order2" "order2" "order1" "order1" "order1" "order2"
## [15] "order2" "order1" "order2" "order1" "order1" "order2"
## [22] "order1"
```

Latin square designs are used in planned experiments because they are optimal in several ways.

Example 1: Reading time differences in subject vs object relatives in English

Latin-square design: Generating fake data

```
library(MASS)
```

```
##
## Attaching package: 'MASS'
## The following object is masked from 'package:dplyr':
##
##     select
```

```
nitem <- 16
nsubj <- 42
## prepare data frame for two condition in a latin square des
g1<-data.frame(item=1:nitem,
                  cond=rep(c("objgap","subjgap"),nitem/2))
g2<-data.frame(item=1:nitem,
                  cond=rep(c("subjgap","objgap"),nitem/2))
```

Example 1: Reading time differences in subject vs object relatives in English

Latin-square design: Generating fake data

```
## assemble data frame in long format:
gp1<-g1[rep(seq_len(nrow(g1)),
            nsubj/2),]
gp2<-g2[rep(seq_len(nrow(g2)),
            nsubj/2),]


fakedat<-rbind(gp1,gp2)
dim(fakedat) ## sanity check
```

```
## [1] 672   2
```

```
fakedat$subj<-rep(1:nsubj,each=nitem) ## add subjects
fakedat<-fakedat[,c(3,1,2)]
fakedat$so<-ifelse(fakedat$cond=="objgap",1,-1)
```

Example 1: Reading time differences in subject vs object relatives in English

Latin-square design: Generating fake data

For example, subject 1 sees the following conditions and items:

```
head(fakedat,n=16)
```

```
##    subj item    cond so
## 1     1    1  objgap  1
## 2     1    2 subjgap -1
## 3     1    3  objgap  1
## 4     1    4 subjgap -1
## 5     1    5  objgap  1
## 6     1    6 subjgap -1
## 7     1    7  objgap  1
## 8     1    8 subjgap -1
## 9     1    9  objgap  1
## 10    1   10 subjgap -1
## 11    1   11  objgap  1
## 12    1   12 subjgap -1
```

Example 1: Reading time differences in subject vs object relatives in English

Fully crossed subjects and items

In the data, because of the Latin square design, each subject sees exactly one item in one of the two conditions:

```
xtabs(~subject+item,gge1crit)
```

```
##         item
## subject 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
##       1  1 1 1 1 1 1 1 1 1  1  1  1  1  1  1  1
##       2  1 1 1 1 1 1 1 1 1  1  1  1  1  1  1  1
##       3  1 1 1 1 1 1 1 1 1  1  1  1  1  1  1  1
##       4  1 1 1 1 1 1 1 1 1  1  1  1  1  1  1  1
##       5  1 1 1 1 1 1 1 1 1  1  1  1  1  1  1  1
##       6  1 1 1 1 1 1 1 1 1  1  1  1  1  1  1  1
##       7  1 1 1 1 1 1 1 1 1  1  1  1  1  1  1  1
##       8  1 1 1 1 1 1 1 1 1  1  1  1  1  1  1  1
##       9  1 1 1 1 1 1 1 1 1  1  1  1  1  1  1  1
##       10 1 1 1 1 1 1 1 1 1  1  1  1  1  1  1  1
```

Example 1: Reading time differences in subject vs object relatives in English

The implied generative model

The above design implies a particular statistical model that takes us beyond the linear model.
To remind you, a simple linear model of the above data would be:

$$y \sim Normal(\alpha + \beta * so, \sigma) \tag{1}$$

Here, object gaps are coded +1, subject gaps -1. See Schad et al. (2018) for an explanation of contrast coding.

```
gge1crit$so<-ifelse(gge1crit$condition=="objgap",1,-1)
```

Example 1: Reading time differences in subject vs object relatives in English

The implied generative model

As figure 1 shows, a Normal likelihood doesn't seem well motivated, so we will use the log-normal.

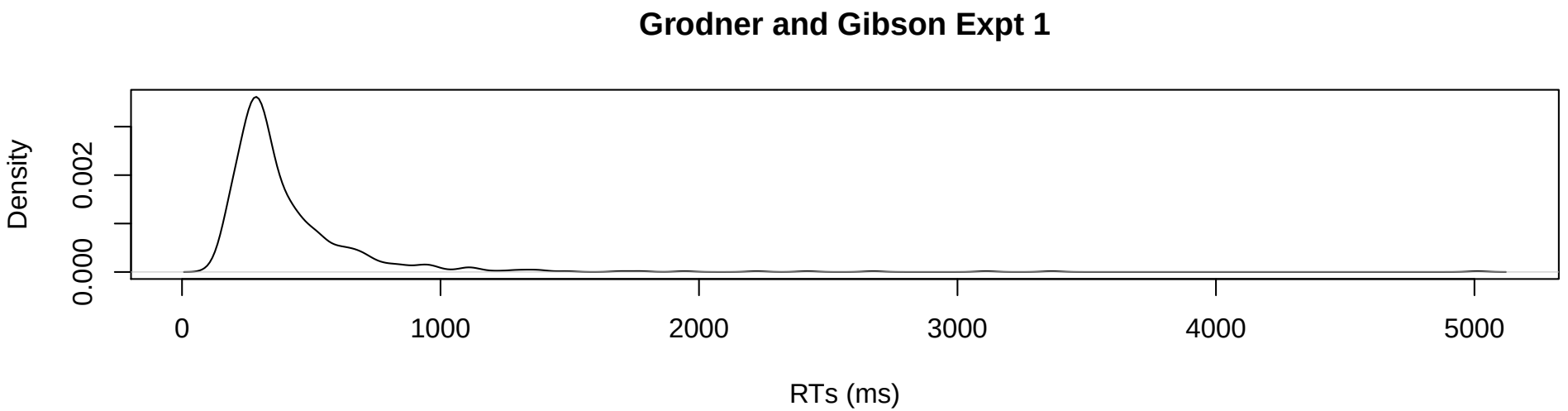


Figure 1: Distribution of reading times in the Grodner and Gibson Experiment 1 data, at the critical region.

Example 1: Reading time differences in subject vs object relatives in English

The implied generative model

$$y \sim LogNormal(\alpha + \beta * so, \sigma) \tag{2}$$

Example 1: Reading time differences in subject vs object relatives in English

Between subject variability in mean reading time

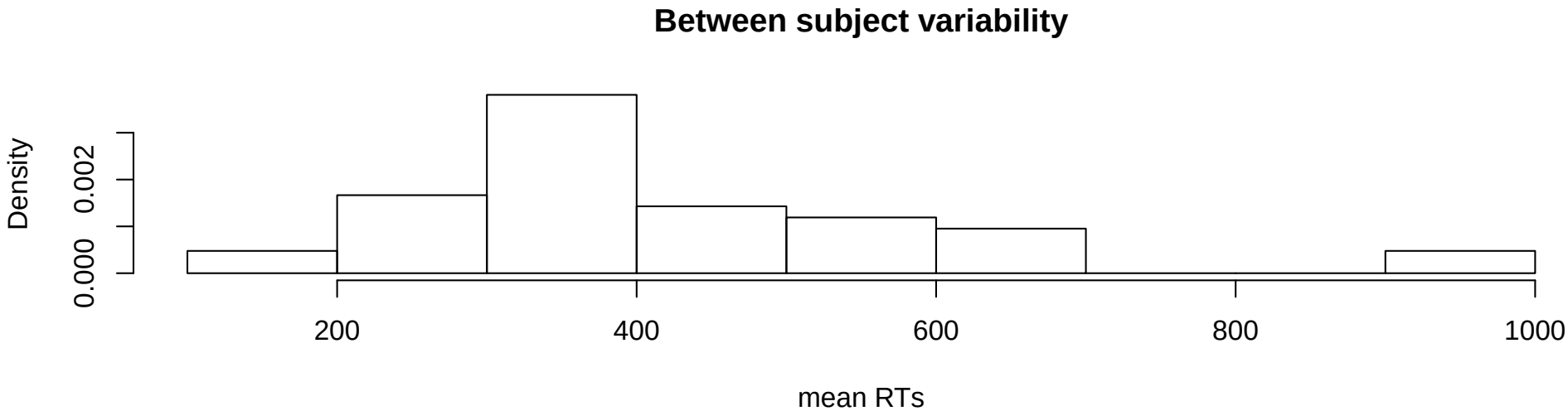


Figure 2: Between subject variability in mean reading times.

Example 1: Reading time differences in subject vs object relatives in English

Between subject variability in mean reading time

In the linear model, we can express the assumption that the grand mean intercept α needs an adjustment by subject, where subjects are indexed from $j = 1, \ldots, J$:

$$y_j \sim LogNormal(\alpha + u_{0j} + \beta * so_j, \sigma) \tag{3}$$

where we now have two sources of variance:

- within subject: σ
- between subject variance in mean reading times: $u_{0j} \sim Normal(0, \sigma_{u0})$

Example 1: Reading time differences in subject vs object relatives in English

Between item variability in mean reading time

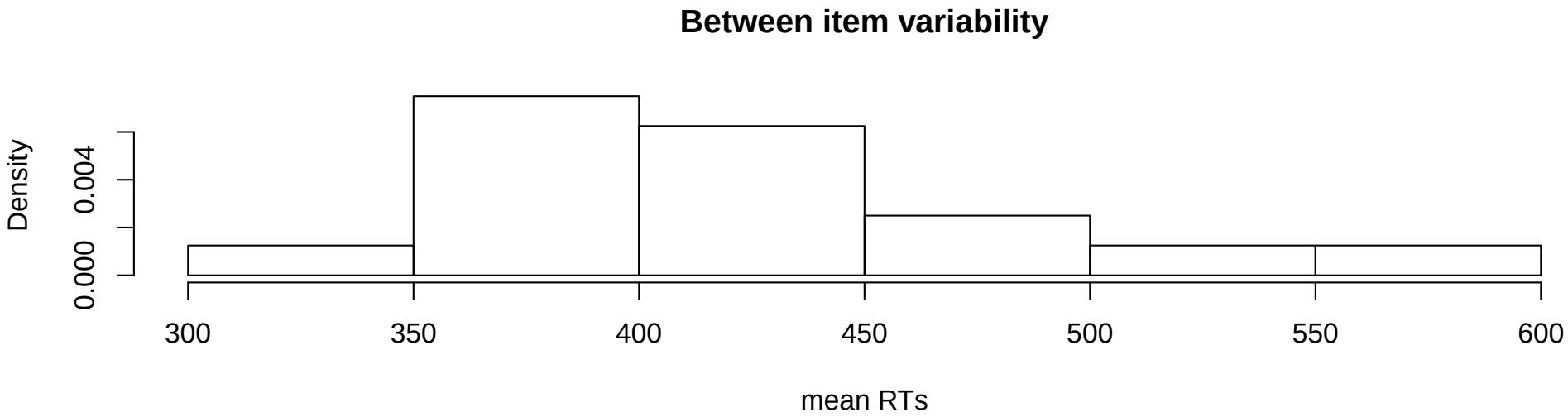


Figure 3: Between item variability in mean reading times.

Example 1: Reading time differences in subject vs object relatives in English

Between item variability in mean reading time

For items ranging from $k = 1, \ldots, K$, we can add this assumption to the model:

$$y_{kj} \sim LogNormal(\alpha + u_{0j} + w_{0k} + \beta * so_{kj}, \sigma) \tag{4}$$

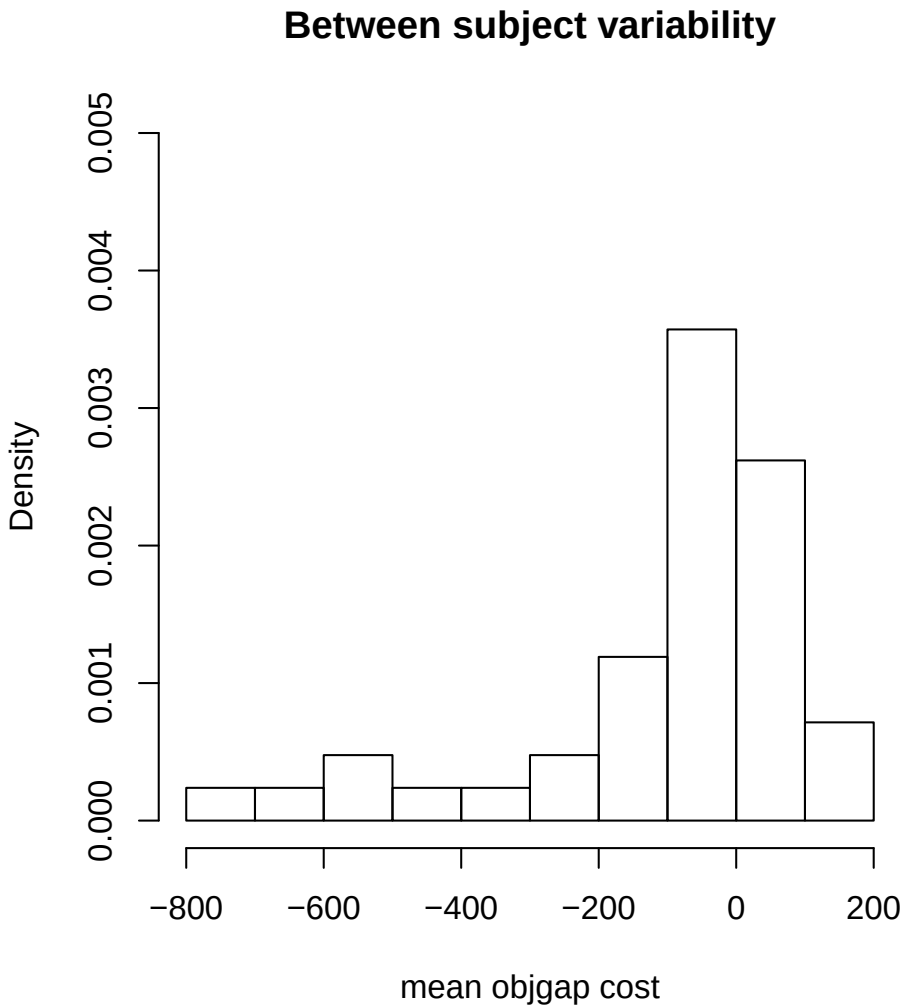
where there are now three variance components:

- σ
- $u_{0j} \sim Normal(0, \sigma_{u0})$
- between item variability in mean reading time, $w_{0k} \sim Normal(0, \sigma_{w0})$

This model is called a *varying intercepts model* with crossed varying intercepts for subjects and for items.

Example 1: Reading time differences in subject vs object relatives in English

Between subject and between item variability in objgap cost



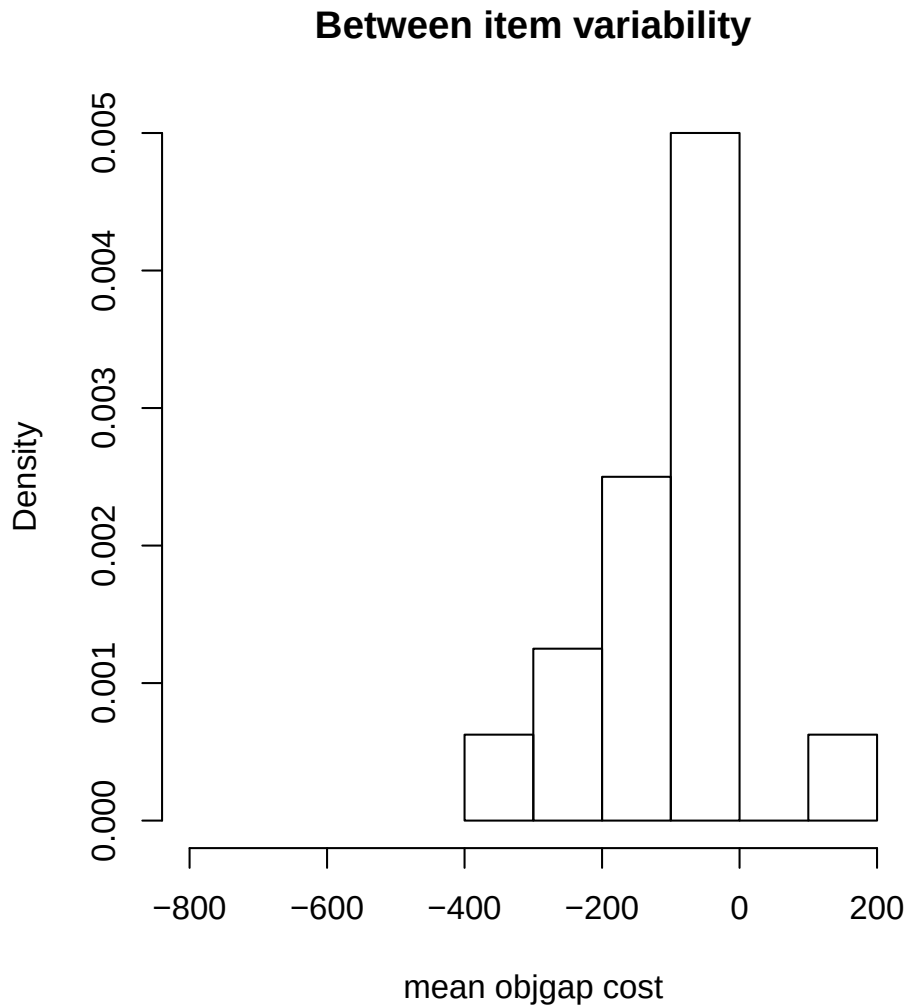


Figure 4: Between subject and item variability in object gap vs subject gap

Example 1: Reading time differences in subject vs object relatives in English

Between subject and between item variability in objgap cost

We can incorporate this assumption into the model by adding adjustments to the β parameter:

$$y_{kj} \sim LogNormal(\alpha + u_{0j} + w_{0k} + (\beta + u_{1j} + w_{1k}) * so_{kj}, \sigma) \quad (5)$$

where

- σ
- $u_{0j} \sim Normal(0, \sigma_{u0})$
- $u_{1j} \sim Normal(0, \sigma_{u1})$
- $w_{0k} \sim Normal(0, \sigma_{w0})$
- $w_{1k} \sim Normal(0, \sigma_{w1})$

This is called the *varying intercepts and slopes* model with *no correlation* between the intercepts and slopes.

Example 1: Reading time differences in subject vs object relatives in English

The "maximal" model

- There is one detail still missing in the model: the adjustments to the intercept and slope are correlated for subjects, and also for items.
- In other words, we have a bivariate distribution for the subject and item random effects:

Example 1: Reading time differences in subject vs object relatives in English

The "maximal" model

$$y_{kj} \sim LogNormal(\alpha + u_{0j} + w_{0k} + (\beta + u_{1j} + w_{1k}) * so_{kj}, \sigma) \quad (6)$$

where we have variance components: σ and

$$\Sigma_u = \begin{pmatrix} \sigma^2_{u0} & \rho_u \sigma_{u0} \sigma_{u1} \\ \rho_u \sigma_{u0} \sigma_{u1} & \sigma^2_{u1} \end{pmatrix} \quad \Sigma_w = \begin{pmatrix} \sigma^2_{w0} & \rho_w \sigma_{w0} \sigma_{w1} \\ \rho_w \sigma_{w0} \sigma_{w1} & \sigma^2_{w1} \end{pmatrix} \quad (7)$$

$$\begin{pmatrix} u_0 \\ u_1 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_u \right), \quad \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_w \right) \quad (8)$$

Example 1: Reading time differences in subject vs object relatives in English

The maximal model

This is a varying intercepts and slopes model with fully specified variance-covariance matrices for the subject and item random effects. It is sometimes called the **maximal model** (Barr et al. 2013).

Example 1: Reading time differences in subject vs object relatives in English

Implementing the model

The above model is simple to implement in the Bayesian framework.

Example 1: Reading time differences in subject vs object relatives in English

Specify and visualize priors

We define some priors first:

1. $\alpha \sim Normal(0, 10)$
2. $\beta \sim Normal(0, 1)$
3. Residual standard deviation: $\sigma \sim Normal_+(0, 1)$
4. All other standard deviations: $\sigma \sim Normal_+(0, 1)$
5. Correlation matrix: $\rho \sim LKJ(2)$.

Example 1: Reading time differences in subject vs object relatives in English

The LKJ prior on the correlation matrix

- In this model, we assume that the vector $\mathbf{u} = \langle u_0, u_1 \rangle$ comes from a bivariate normal distribution with a variance-covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{u}}$.
- This matrix has the variances of the adjustment to the intercept and to the slope respectively along the diagonal, and the covariance on the off-diagonals.

Example 1: Reading time differences in subject vs object relatives in English

The LKJ prior on the correlation matrix

- Recall that the covariance $Cov(X, Y)$ between two variables X and Y is defined as the product of their correlation ρ and their standard deviations σ_X and σ_Y, such that, $Cov(X, Y) = \rho\sigma_X\sigma_Y$.

$$\boldsymbol{\Sigma_u} = \begin{pmatrix} \sigma^2_{u_0} & \rho_u\sigma_{u_0}\sigma_{u_1} \\ \rho_u\sigma_{u_0}\sigma_{u_1} & \sigma^2_{u_1} \end{pmatrix} \tag{9}$$

Example 1: Reading time differences in subject vs object relatives in English

The LKJ prior on the correlation matrix

The covariance matrix can be decomposed into a vector of standard deviations and a correlation matrix. The correlation matrix looks like this:

$$\begin{pmatrix} 1 & \rho_u \\ \rho_u & 1 \end{pmatrix} \tag{10}$$

Example 1: Reading time differences in subject vs object relatives in English

The LKJ prior on the correlation matrix

In Stan, we write a matrix that has 0's on the off-diagonals as:

$$diag\_matrix(\sigma_{u_0}, \sigma_{u_1}) = \begin{pmatrix} \sigma_{u_0} & 0 \\ 0 & \sigma_{u_1} \end{pmatrix} \tag{11}$$

This means that we can decompose the covariance matrix into three parts:

$$\begin{aligned} \boldsymbol{\Sigma_u} &= diag\_matrix(\sigma_{u_0}, \sigma_{u_1}) \cdot \boldsymbol{\rho_u} \cdot diag\_matrix(\sigma_{u_0}, \sigma_{u_1}) \\ &= \begin{pmatrix} \sigma_{u_0} & 0 \\ 0 & \sigma_{u_1} \end{pmatrix} \begin{pmatrix} 1 & \rho_u \\ \rho_u & 1 \end{pmatrix} \begin{pmatrix} \sigma_{u_0} & 0 \\ 0 & \sigma_{u_1} \end{pmatrix} \end{aligned} \tag{12}$$

Example 1: Reading time differences in subject vs object relatives in English

The LKJ prior on the correlation matrix

So we need priors for the σ_u's and for ρ_u:

- The basic idea of the LKJ prior is that its parameter (usually called *eta*, η, here it has value 2) increases, the prior increasingly concentrates around the unit correlation matrix (i.e., favors smaller correlation: ones in the diagonals and values close to zero in the lower and upper triangles).
- At $\eta = 1$, the LKJ correlation distribution is uninformative (similar to $Beta(1, 1)$), at $\eta < 1$, it favors extreme correlations (similar to $Beta(a < 1, b < 1)$).

Example 1: Reading time differences in subject vs object relatives in English

Visualize the priors

As always, it is a good idea to visualize these priors. See Figure 5.

Example 1: Reading time differences in subject vs object relatives in English

Visualize the priors

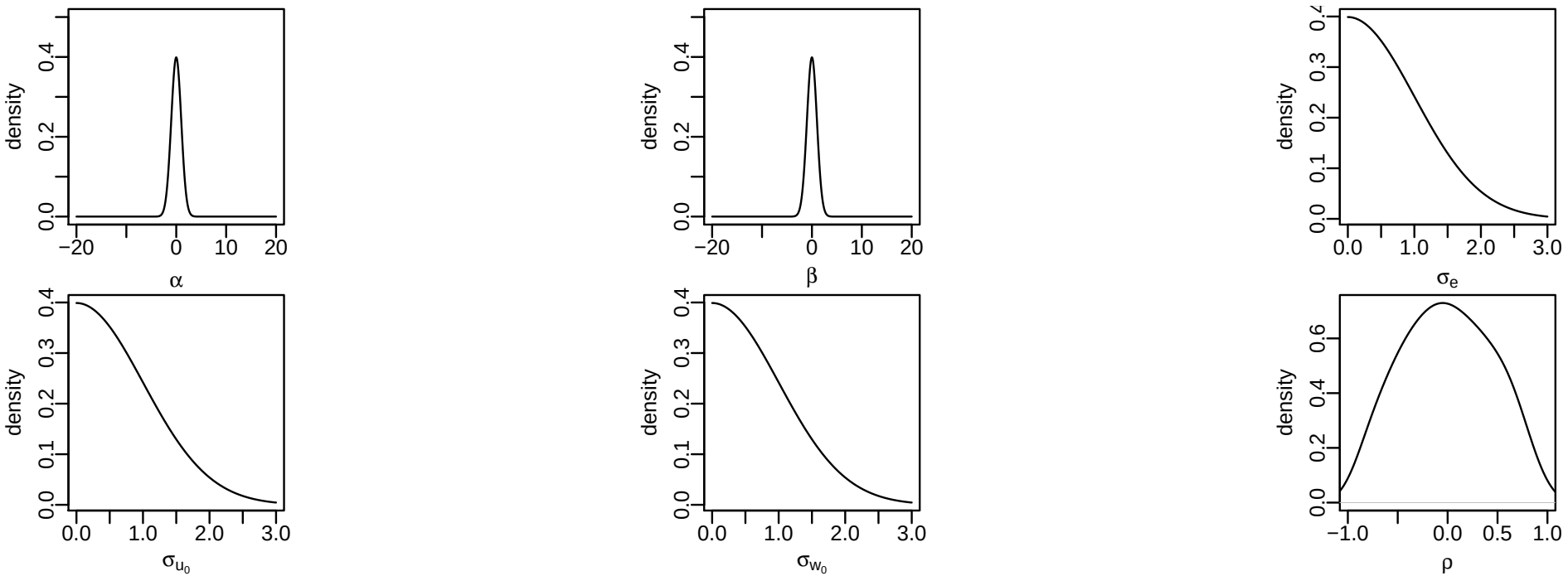


Figure 5: Priors for the Godner and Gibson data.

Example 1: Reading time differences in subject vs object relatives in English

Fit the model using brms

```
priors <- c(set_prior("normal(0, 10)",
                      class = "Intercept"),
                      set_prior("normal(0, 1)",
                                class = "b",
                                coef = "so"),
                      set_prior("normal(0, 1)",
                                class = "sd"),
                      set_prior("normal(0, 1)",
                                class = "sigma"),
                      set_prior("lkj(2)",
                                class = "cor"))
```

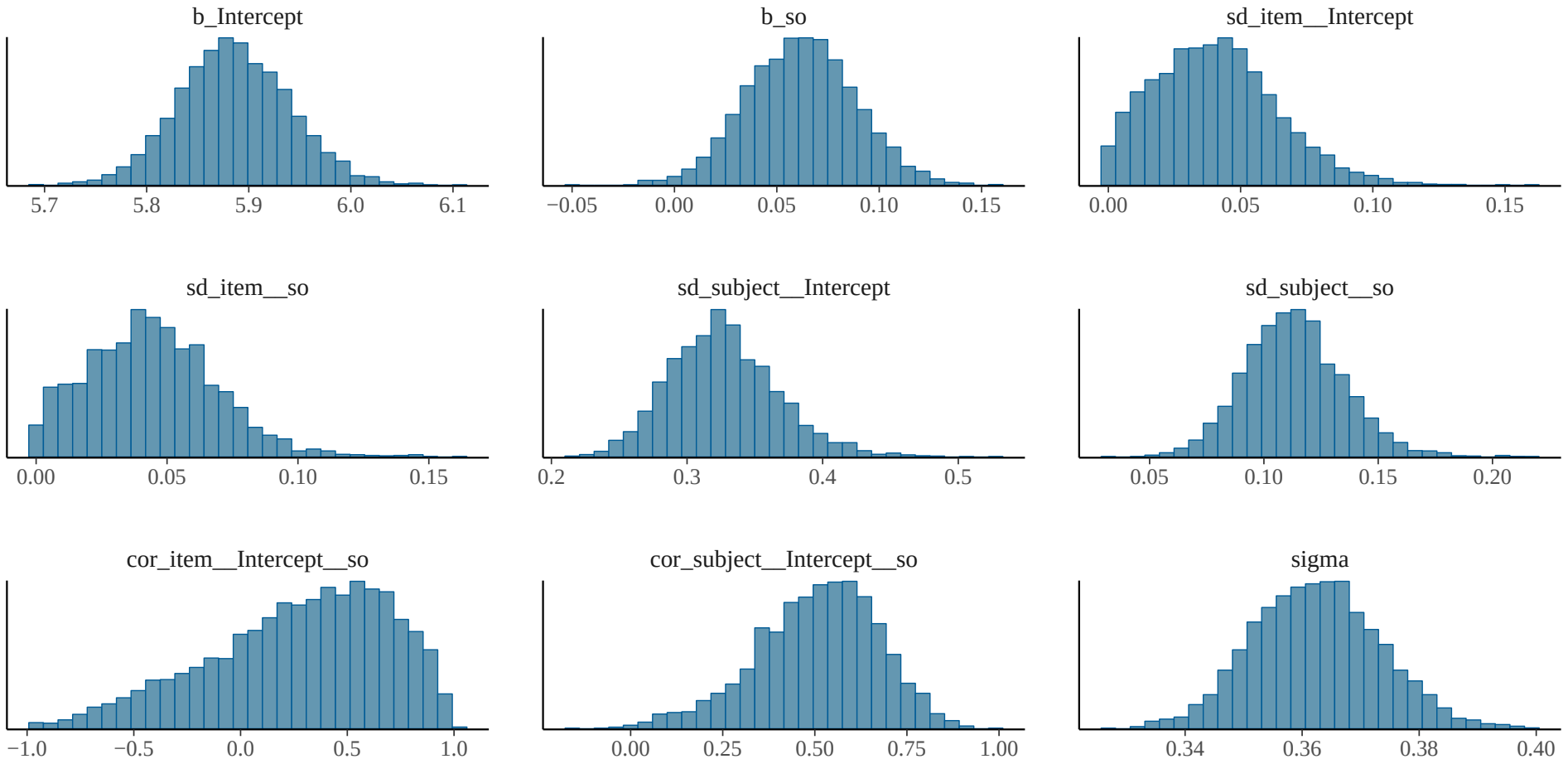
Example 1: Reading time differences in subject vs object relatives in English

Fit the model using brms

```
m_gg<-brm(rawRT~so + (1+so|subject) + (1+so|item),
          gge1crit,family=lognormal(),
    prior=priors)
```

Example 1: Reading time differences in subject vs object relatives in English

Fit the model using brms



Example 1: Reading time differences in subject vs object relatives in English

Fit the model using brms

Look at the posterior distributions of the parameters on the log ms scale (for the coefficients and standard deviations). Notice that

- The object relative takes longer to read than the subject relative, as predicted. We know this because the parameter b_so is positive.
- The largest sources of variance are the subject intercepts, slopes, and the residual standard deviation. Look at the sd_subject parameters, and sigma.
- The by-item variance components are relatively small. Look at the sd_item parameters.
- The correlations have very wide uncertainty—the prior is dominating in determining the posteriors as there isn't that much data to obtain accurate estimates of these parameters. Look at the cor parameters.

Example 1: Reading time differences in subject vs object relatives in English

Examine by subject random effects visually

First, extract the posterior samples of the parameters that we will need to compute individual differences.

```
library(bayesplot)
```

```
postgg<-posterior_samples(m_gg)
## extract variances:
alpha<-postgg$b_Intercept
beta<-postgg$b_so
cor<-posterior_samples(m_gg,"^cor")
sd<-posterior_samples(m_gg,"^sd")
sigma<-posterior_samples(m_gg,"sigma")
```

Example 1: Reading time differences in subject vs object relatives in English

Examine by subject random effects visually

```
## item random effects won't be used below
item_re<-posterior_samples(m_gg,"^r_item")
subj_re<-posterior_samples(m_gg,"^r_subj")
```

Example 1: Reading time differences in subject vs object relatives in English

By subject slope adjustments

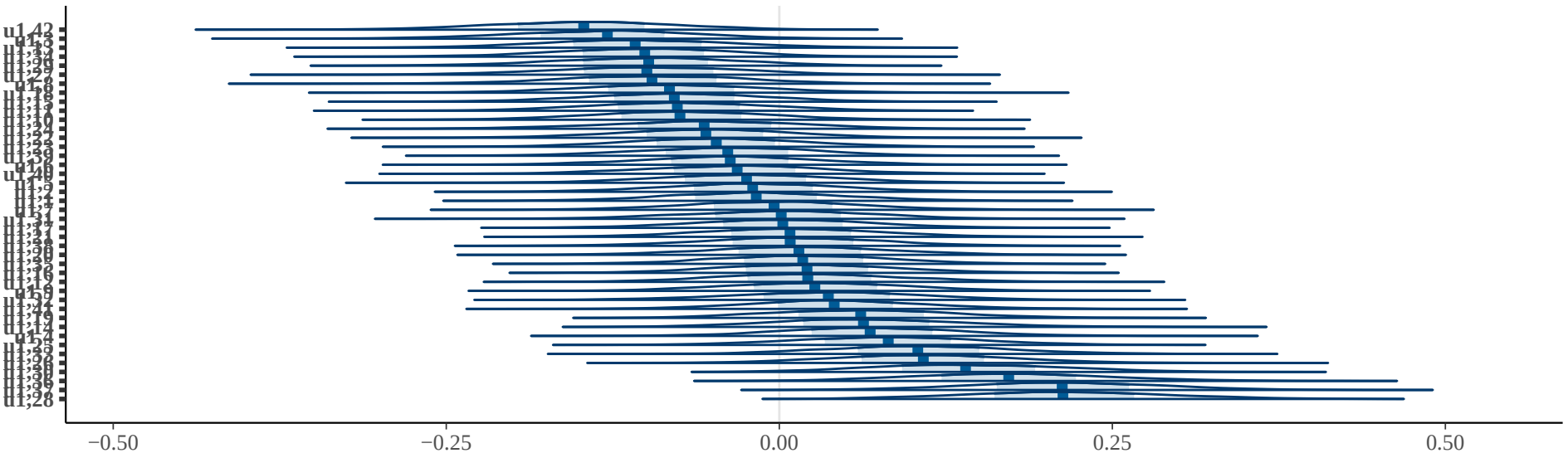


Figure 6: Variability in subject slope adjustments in the Grodner and Gibson data.

Example 1: Reading time differences in subject vs object relatives in English

By subject slope adjustments

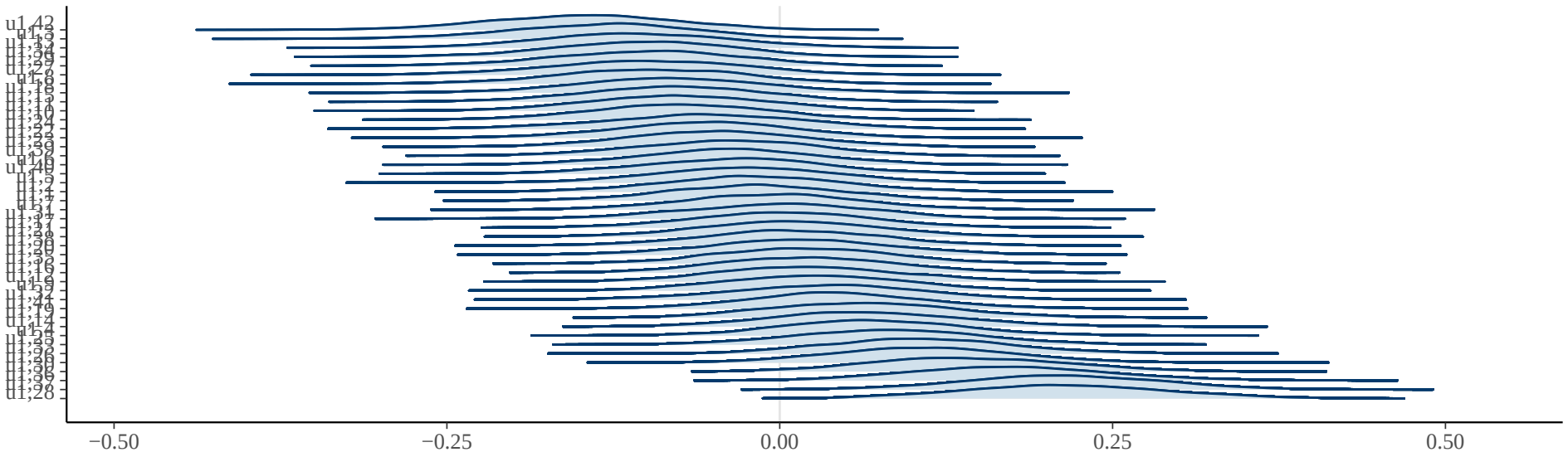


Figure 7: Variability in subject slope adjustments in the Grodner and Gibson data.

Example 1: Reading time differences in subject vs object relatives in English

By subject slope adjustments

```
## `stat_bin()` using `bins = 30`. Pick better value with `bin
```

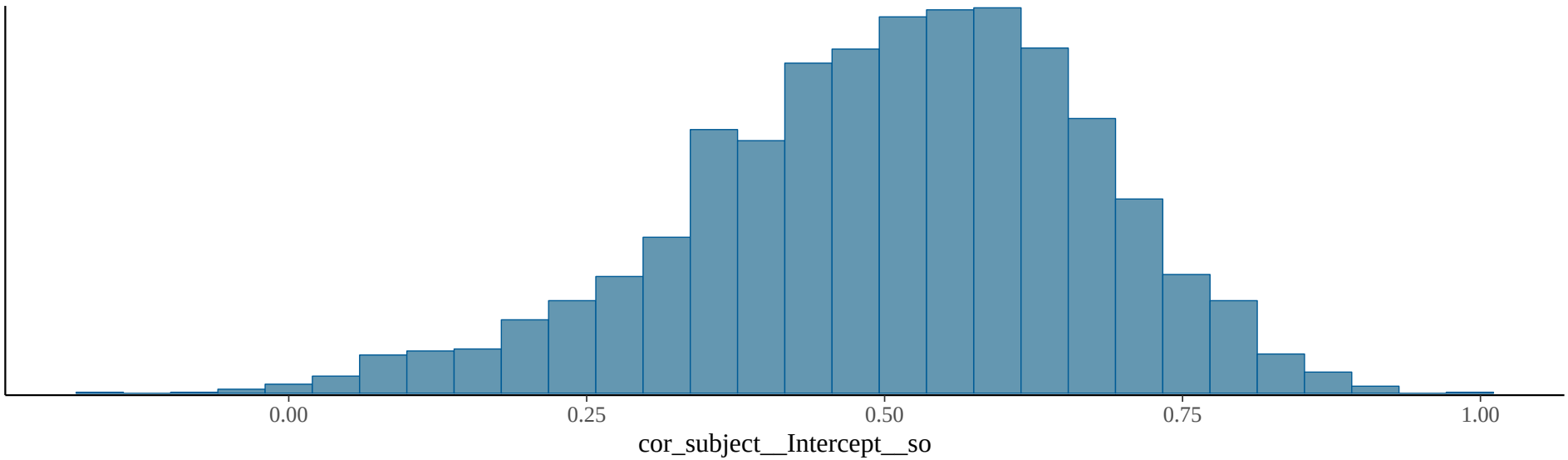


Figure 8: Posterior distributions of subject varying intercept and slope correlation parameter in the Grodner and Gibson data.

Example 1: Reading time differences in subject vs object relatives in English

Examine mean and individual differences on the raw ms scale

It is useful to see the effects on the raw ms scale. The log ms scale is difficult to interpret.

Example 1: Reading time differences in subject vs object relatives in English

Mean difference

```
meandiff<- exp(alpha + beta) - exp(alpha - beta)
mean(meandiff)
```

```
## [1] 44.98135
```

```
round(quantile(meandiff,prob=c(0.025,0.975)),0)
```

```
##  2.5% 97.5%
##     7    85
```

Example 1: Reading time differences in subject vs object relatives in English

Mean difference

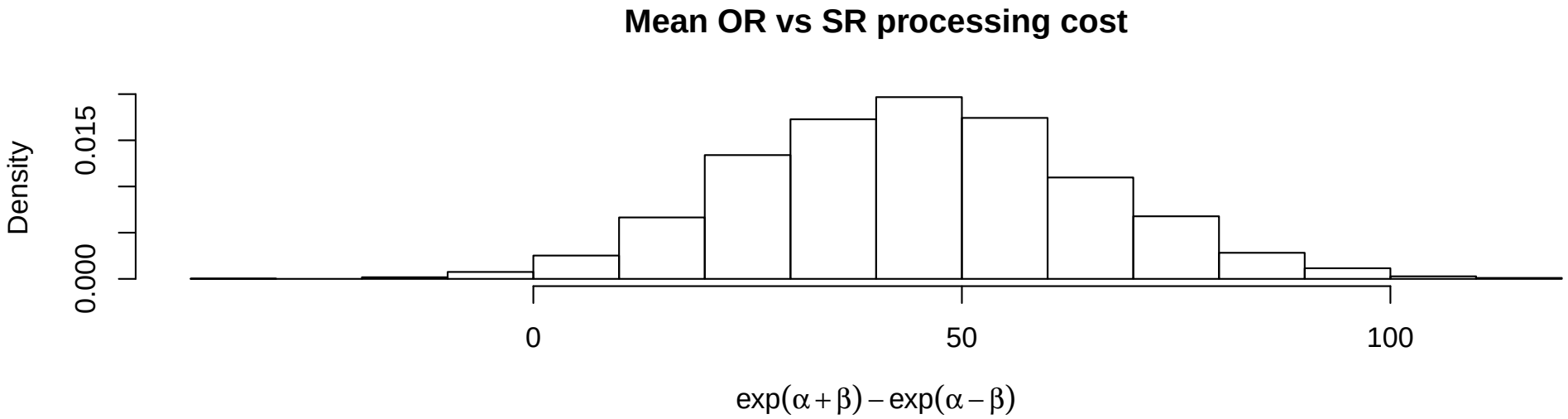


Figure 9: Mean OR processing cost effect in the Grodner and Gibson data.

Example 1: Reading time differences in subject vs object relatives in English

Individual effects of OR processing cost

Example 1: Reading time differences in subject vs object relatives in English

Individual effects of OR processing cost

```
for(i in 1:42){
  hist(subjdiff[,i],xlim=c(min(subjdiff),
                           max(subjdiff)),
       main=paste("subj",colnames(subjdiff)[i],sep=" "))
  Sys.sleep(.2)
}
```

Example 1: Reading time differences in subject vs object relatives in English

Individual effects of OR processing cost

This illustrates a point that Blastland and Spiegelhalter (2014) make: "The average is an abstraction. The reality is variation."

Example 1: Reading time differences in subject vs object relatives in English

To make discovery claims about the mean effect, calibrate the true and false discovery rate

- Suppose that, based on these data and this model, we want to claim that there is a mean OR processing cost in English.
- In order to make a discovery claim, we need to understand the **true discovery rate** of this effect.
- In the frequentist world, this would be the *statistical power*, the probability of detecting an effect if there is in fact one.

Example 1: Reading time differences in subject vs object relatives in English

To make discovery claims, calibrate the true and false discovery rate

First, we write a function to generate fake data.

Example 1: Reading time differences in subject vs object relatives in English

To make discovery claims, calibrate the true and false discovery rate

Load file: gen_fake_lnorm.R

Example 1: Reading time differences in subject vs object relatives in English

To make discovery claims, calibrate the true and false discovery rate

Extract the parameter means from the Bayesian model, and assemble the variance covariance matrices for the subject and item random effects.

```
sds<-colMeans(sd)
cors<-colMeans(cor)
sig<-mean(sigma$sigma)
Sigma_u<-diag(sds[3:4]^2)
Sigma_u[1,2]<-Sigma_u[2,1]<-cors[2]*sds[3]*sds[4]
Sigma_w<-diag(sds[1:2]^2)
Sigma_w[1,2]<-Sigma_w[2,1]<-cors[1]*sds[1]*sds[2]
```

Example 1: Reading time differences in subject vs object relatives in English

To make discovery claims, calibrate the true and false discovery rate

- Then, we run 50 simulations, computing the 95% credible interval of the OR processing cost effect.
- Because this is a very time-consuming calculation, we are going to use previously computed values.

```
nsim<-50
betaquants<-matrix(rep(NA,nsim*2),ncol =2)
betameans<-matrix(rep(NA,nsim),ncol =2)
```

Example 1: Reading time differences in subject vs object relatives in English

To make discovery claims, calibrate the true and false discovery rate

```
for(i in 1:nsim){
gg_fake<-gen_fake_lnorm(alpha=mean(alpha),
                        beta=mean(beta),
                Sigma_u=Sigma_u,Sigma_w=Sigma_w,
                sigma_e=sig)

m_gg_fake<-brm(rt~so + (1+so|subj) + (1+so|item),gg_fake,famil
    prior=priors,
    control = list(adapt_delta = 0.99,max_treedepth=15))
betapost<-posterior_samples(m_gg_fake)$b_so
betaquants[i,]<-quantile(betapost,prob=c(0.025,0.975))
betameans[i]<-mean(betapost)
}
```

Example 1: Reading time differences in subject vs object relatives in English

To make discovery claims, calibrate the true and false discovery rate

- Assume that we are willing to declare an effect just in case 0 is not included in the 95% credible interval of the effect.
- The above simulation shows that we would detect the effect in only half of the repeated experiments.

length(which(betaquants[,1]>0))/50 [1] 0.5

Thus, the true discovery rate is quite low. One would want the true discovery rate to be at least 80%.

Example 1: Reading time differences in subject vs object relatives in English

To make discovery claims, calibrate the true and false discovery rate

- We can also investigate the false discovery rate—the proportion of times we would declare that we found an effect, when there is none.
- In frequentist statistics, this is called Type I error. The only change needed in the above simulation is to set β to 0, to reflect the assumption that there is no effect.

Example 1: Reading time differences in subject vs object relatives in English

Posterior predictive checks

```
pp_check(m_gg, nsamples = 100)+
  theme(text = element_text(size=16),
        legend.text=element_text(size=16))
```

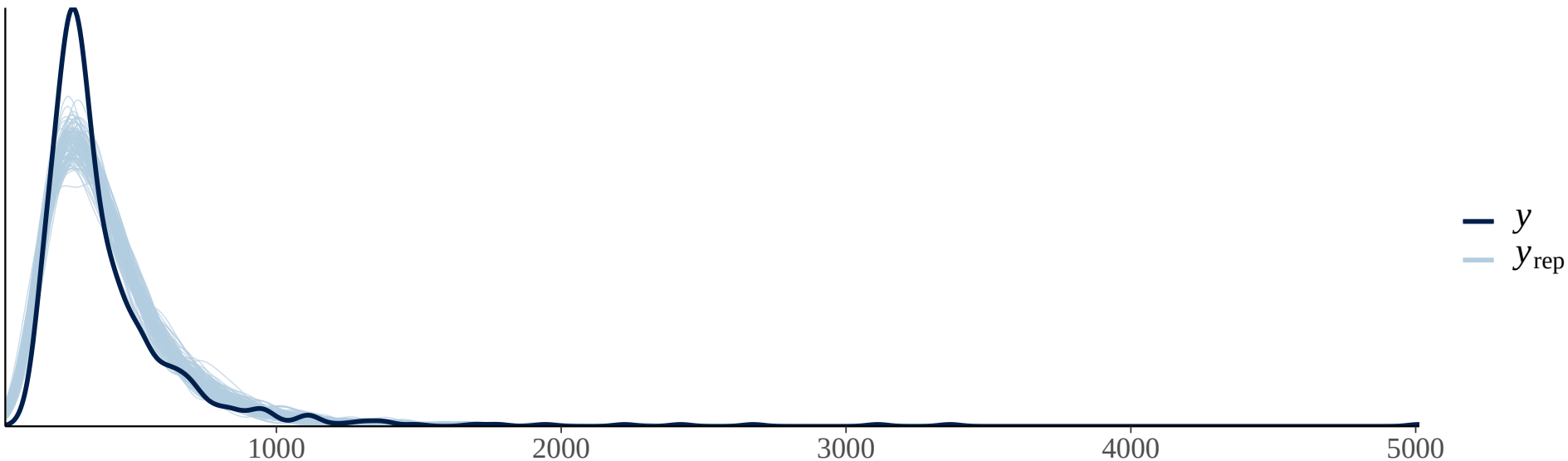


Figure 10: Posterior predictive check for the Grodner and Gibson data.

Example 2: Question-response accuracies (Logistic regression)

The Grodner and Gibson (2005) data also has question-response accuracies: 1 if the response to a question following the sentence was correct, 0 otherwise. We show only the relevant columns below:

```
head(gge1crit[,c(1,2,3,8,11)])
```

```
##    subject item condition qcorrect so
## 6        1    1    objgap        0  1
## 19       1    2   subjgap        1 -1
## 34       1    3    objgap        0  1
## 49       1    4   subjgap        1 -1
## 68       1    5    objgap        1  1
## 80       1    6   subjgap        1 -1
```

Example 2: Question-response accuracies (Logistic regression)

One could aggregate the accuracy by item, and then just fit a hierarchical linear model:

```
meanp<-with(gge1crit,tapply(qcorrect,
                      IND=list(condition,subject),
                      mean))
q_df<-data.frame(subj=rep(c(1:42),2),
            so=rep(c(1,-1),each=42),
            p=c(meanp[1,],meanp[2,]))

head(q_df)
```

```
##   subj so     p
## 1    1  1 0.750
## 2    2  1 0.875
## 3    3  1 1.000
## 4    4  1 1.000
```

Example 2: Question-response accuracies (Logistic regression)

```
mqlmer<-lmer(p~so+(1|subj),q_df)
```

```
## boundary (singular) fit: see ?isSingular
```

```
summary(mqlmer)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: p ~ so + (1 | subj)
##    Data: q_df
##
## REML criterion at convergence: -97.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.03723 -0.77697 -0.07063  0.91823  1.20076
##
## Random effects:
```

Example 2: Question-response accuracies (Logistic regression)

Think about the generative process; a 0,1 response is best seen as generated by a Bernoulli distribution with probability of success p:
response $\sim$ *Bernoulli*(p). This is the same as a Binomial process, with one trial.

One can therefore model each 0,1 response as being generated from a Bernoulli distribution, which is just a Binomial with a single trial. Thus, what is of interest is the probability of correct responses in subject vs object relatives:

```
round(100*with(gge1crit,
               tapply(qcorrect,condition,mean)))
```

```
##  objgap subjgap
##      88      85
```

Example 2: Question-response accuracies (Logistic regression)

We will transform the probability p of a correct response to a log-odds:

$$\log \frac{p}{1-p} \tag{13}$$

and assume that the log-odds of a correct response is affected by the relative clause type:

$$\log \frac{p}{1-p} = \alpha + \beta * so \tag{14}$$

Example 2: Question-response accuracies (Logistic regression)

This model is called a *logistic* regression because it uses the logistic or logit function to transform p to log odds space. Notice that there is no residual term in this model.

Example 2: Question-response accuracies (Logistic regression)

We can fit the above model easily using brms:

```
m_gg_q1<-brm(qcorrect~so,gge1crit,
             family=bernoulli(link="logit"))
```

```
## Compiling the C++ model

## Start sampling
```

```
summary(m_gg_q1)
```

Example 2: Question-response accuracies (Logistic regression)

Obviously, because the question-response data are also repeated measures, we must use a hierarchical linear model, with varying intercepts and slopes for subject and item, as in Example 1:

```
m_gg_q2<-brm(qcorrect~so+(1+so|subject) + (1+so|item),
             gge1crit,family=bernoulli(link="logit"))
```

```
## Compiling the C++ model

## Start sampling
```

```
summary(m_gg_q2)
```

This model is not especially good because many of the response accuracies are at ceiling. However, in principle this kind of model is appropriate for binary responses.

Example 2: Question-response accuracies (Logistic regression)

Convert posteriors back to probability space

What is theoretically important is the posterior distribution of the difference between object and subject relative response accuracy. That is on the probability scale. We can go from log-odds space to probability space by solving this equation for p.
Using simple algebra, we can go from:

$$\log \frac{p}{1-p} = \alpha + \beta * so = \mu \tag{15}$$

to:

$$p = \exp(\mu)/(1 + \exp(\mu)) \tag{16}$$

Example 2: Question-response accuracies (Logistic regression)

Convert posteriors back to probability space

For object gap sentences, the factor *so* is coded as 1, so we have $\mu = \alpha + \beta$.
For subject gap sentences, *so* is coded as -1, so we have $\mu = \alpha - \beta$.
Therefore, we just need to plug in the expression for μ for object and subject relatives.
We can now straightforwardly plot the posterior distribution of the difference between object and subject relatives. We see that there isn't any important difference between the two relative clause types.

Example 2: Question-response accuracies (Logistic regression)

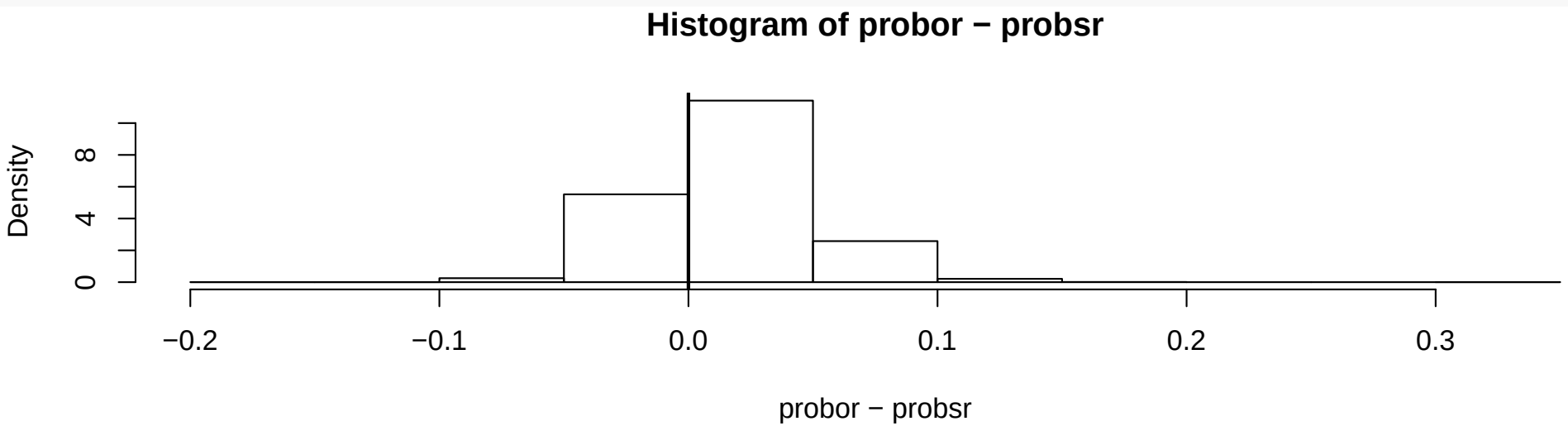
Convert posteriors back to probability space

```
postq<-posterior_samples(m_gg_q2)
alpha<-postq$b_Intercept
beta<-postq$b_so
mu_or<-alpha+beta
probor<-exp(mu_or)/(1+exp(mu_or))
mu_sr<-alpha-beta
probsr<-exp(mu_sr)/(1+exp(mu_sr))
```

Example 2: Question-response accuracies (Logistic regression)

Convert posteriors back to probability space

```
hist(probor-probsr,freq=FALSE)
abline(v=0,lwd=2)
```



References

Barr, Dale J, Roger Levy, Christoph Scheepers, and Harry J Tily. 2013. "Random Effects Structure for Confirmatory Hypothesis Testing: Keep It Maximal." *Journal of Memory and Language* 68 (3). Elsevier: 255–78.

Blastland, Michael, and David Spiegelhalter. 2014. *The Norm Chronicles: Stories and Numbers About Danger and Death*. Basic Books (AZ).

Grodner, Daniel, and Edward Gibson. 2005. "Consequences of the Serial Nature of Linguistic Input." *Cognitive Science* 29: 261–90.

Schad, Daniel J., Sven Hohenstein, Shravan Vasishth, and Reinhold Kliegl. 2018. "How to Capitalize on a Priori Contrasts in Linear (Mixed) Models: A Tutorial."