English relative clauses

- First, i read the data into the data frame ***dat***. The data columns are:
  - subj: subject id
  - item: item id
  - condition: condition a (subject relative) or b (object relative)
  - so: sum coded contrasts for the conditions
  - rt: reading time at the critical region (the relative clause verb)

First six rows of the data-set are shown below

```
##   item cond subj   so       rt
## 1    1    a    1 -0.5 275.8858
## 2    2    b    1  0.5 202.7486
## 3    3    a    1 -0.5 355.4981
## 4    4    b    1  0.5 205.0296
## 5    5    a    1 -0.5 341.1425
## 6    6    b    1  0.5 229.0490
```

Defining a hierarchical linear model

I will define a hierarchical linear model with varying intercepts and varying slopes for subject and item, assuming a correlation between the varying intercepts and slopes for both subject and item, and a gaussian likelihood.

- Likelihood:

$$rt_n \sim Normal(\alpha + u_{subj[n],1} + w_{item[n],1} + so_n * (\beta + u_{subj[n],2} + w_{item[n],2}), \sigma)$$

- Priors:

$$\alpha \sim Normal(0{,}1000)$$

$$beta \sim Normal(0{,}1000)$$

$$sigma \sim Normal_{+}(0{,}500)$$

$$\begin{pmatrix} u_{i,1} \\ u_{i,2} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_u\right)$$

$$\begin{pmatrix} w_{i,1} \\ w_{i,2} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_w\right)$$

Where

$$\Sigma_u = \begin{pmatrix} \tau_{u_1}^2 & \rho_u \tau_{u_1} \tau_{u_2} \\ \rho_u \tau_{u_1} \tau_{u_2} & \tau_{u_2}^2 \end{pmatrix}$$

$$\Sigma_w = \begin{pmatrix} \tau_{w_1}^2 & \rho_w \tau_{w_1} \tau_{w_2} \\ \rho_w \tau_{w_1} \tau_{w_2} & \tau_{w_2}^2 \end{pmatrix}$$

$\tau_{u_1} \sim Normal_+(0{,}500)$

$\tau_{u_2} \sim Normal_+(0{,}500)$

$\rho_u \sim LKJcorr(2)$

$\tau_{w_1} \sim Normal_+(0{,}500)$

$\tau_{w_2} \sim Normal_+(0{,}500)$

$\rho_w \sim LKJcorr(2)$

- In order to simplify the call to brms i will assign the same priors to the by-subject and by-item parameters

```
m_hier_gaussian_brm<-brm(rt~so+(so|subj)+(so|item),
                data=dat,
                family = gaussian(),
                prior =
                  c(prior(normal(0, 1000), class = Intercept),
                    prior(normal(0, 1000), class = b),
                    prior(normal(0, 500), class = sigma),
                    prior(normal(0, 500), class = sd),
                    prior(lkj(2), class = cor)),
                iter = 4000,
                chains = 4)
```

- the summary of the posteriors:

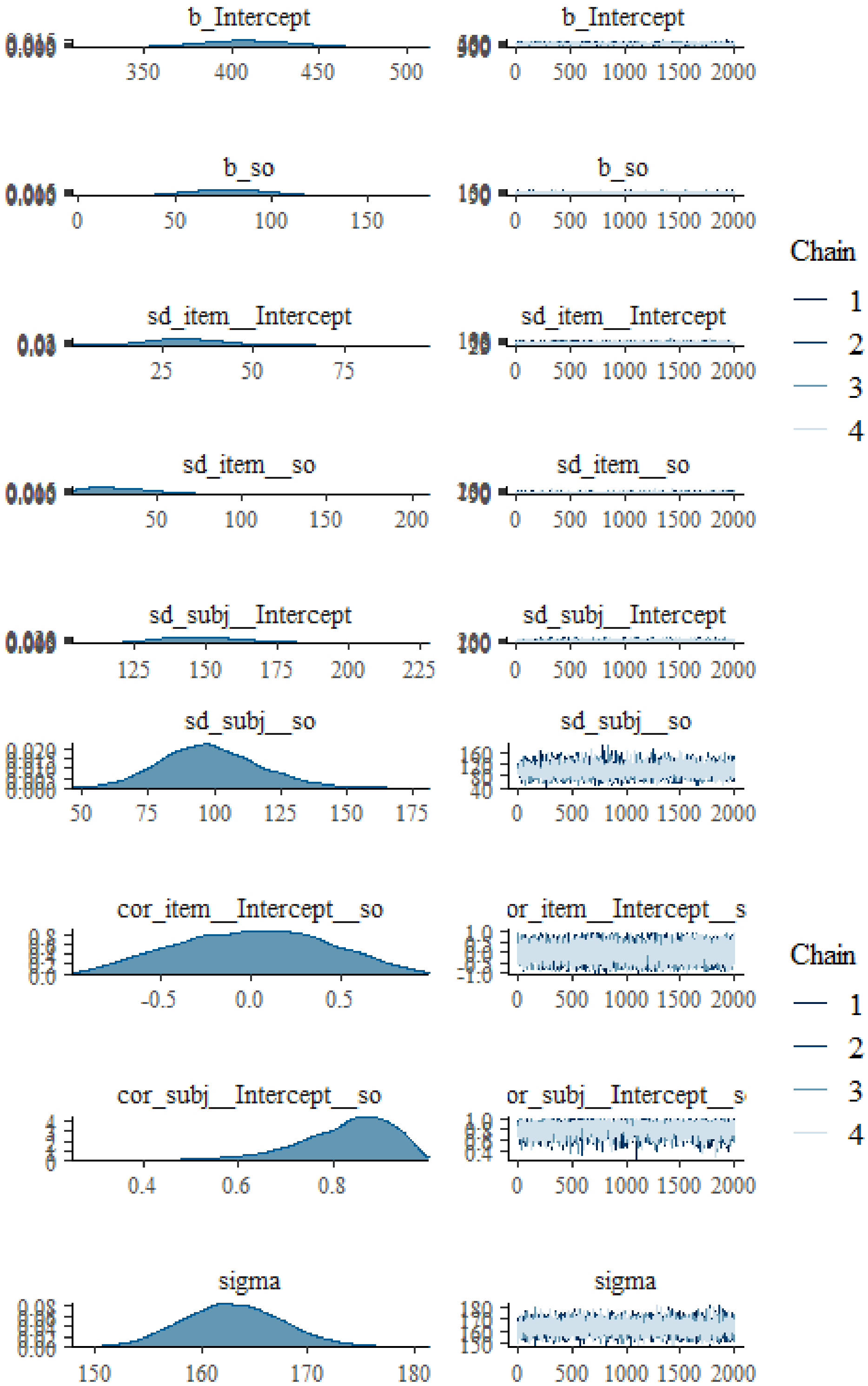
```
##  Family: gaussian
##   Links: mu = identity; sigma = identity
## Formula: rt ~ so + (so | subj) + (so | item)
##    Data: dat (Number of observations: 672)
## Samples: 4 chains, each with iter = 4000; warmup = 2000; thin = 1;
##          total post-warmup samples = 8000
##
## Group-Level Effects:
## ~item (Number of levels: 16)
##                     Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
Tail_ESS
## sd(Intercept)          32.17     11.58    11.45    58.17 1.00     2133
2278
## sd(so)                 29.29     20.18     1.30    74.20 1.00     2028
2407
## cor(Intercept,so)       0.01      0.41    -0.76     0.77 1.00     6833
5467
##
## ~subj (Number of levels: 42)
##                     Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
Tail_ESS
```

```
## sd(Intercept)       150.76     18.05   120.53   191.10 1.00     1407
2899
## sd(so)               98.48     18.23    65.39   137.15 1.00     3049
5119
## cor(Intercept,so)     0.82      0.10     0.58     0.97 1.00     3744
4934
##
## Population-Level Effects:
##           Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept   408.78     26.14   357.24   459.22 1.01      926     1824
## so           77.84     21.80    35.24   121.01 1.00     1750     3770
##
## Family Specific Parameters:
##       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma   163.03      4.81   154.02   172.80 1.00     8030     5941
##
## Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

- summarize the posteriors of all parameters in a table (the fixed effect intercept and slope, the variance components, and the correlations)

```
##                            Estimate        Q2.5       Q97.5
## b_Intercept             408.78387487 357.2428779 459.2230118
## b_so                     77.83611610  35.2385162 121.0110843
## sd_item__Intercept       32.17420315  11.4465741  58.1708145
## sd_item__so              29.29211761   1.2964634  74.1950454
## sd_subj__Intercept      150.75514816 120.5330783 191.0959313
## sd_subj__so              98.48240850  65.3878596 137.1459913
## cor_item__Intercept__so   0.01018581  -0.7637428   0.7662289
## cor_subj__Intercept__so   0.82477937   0.5815898   0.9686230
## sigma                   163.03430552 154.0199381 172.8028543
```

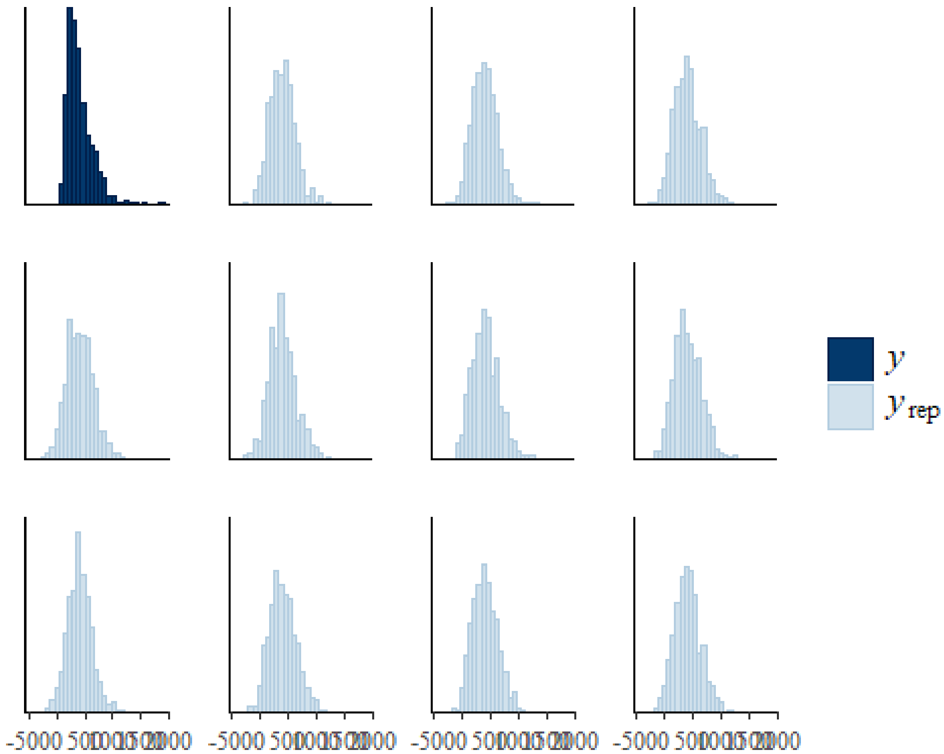
- plot The posterior distributions to obtain a graphical summary of all the parameters in the model:



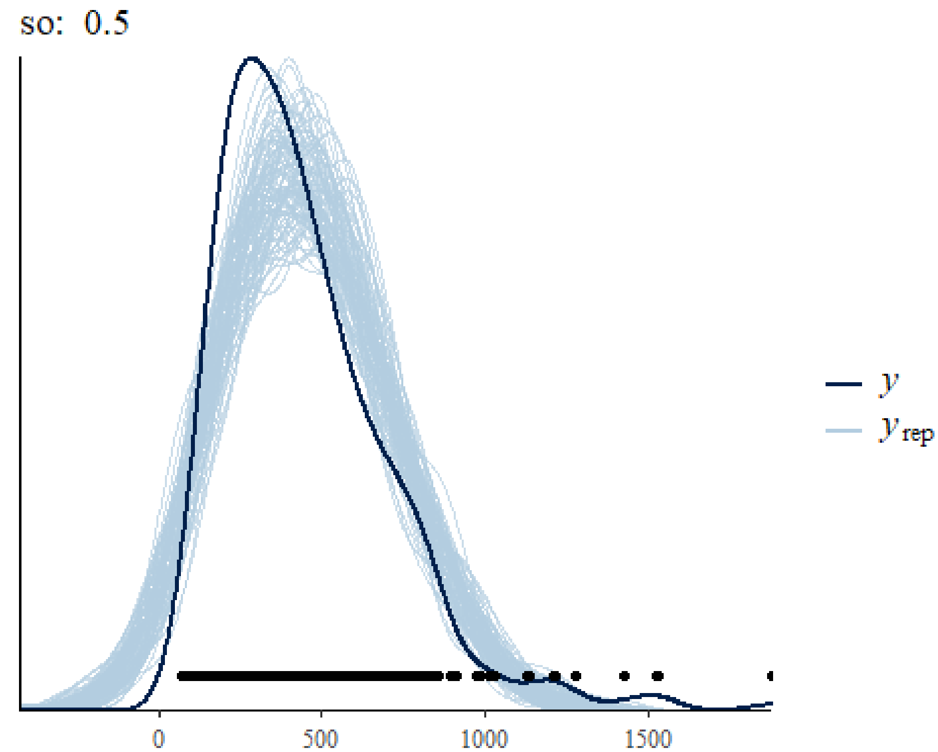
- Histograms of eleven samples from the posterior predictive distribution of the model m_hier_gaussian_brm. the real data is skewed and has no values less than 0 ms while the predictive distributions are centered and symmetrical.

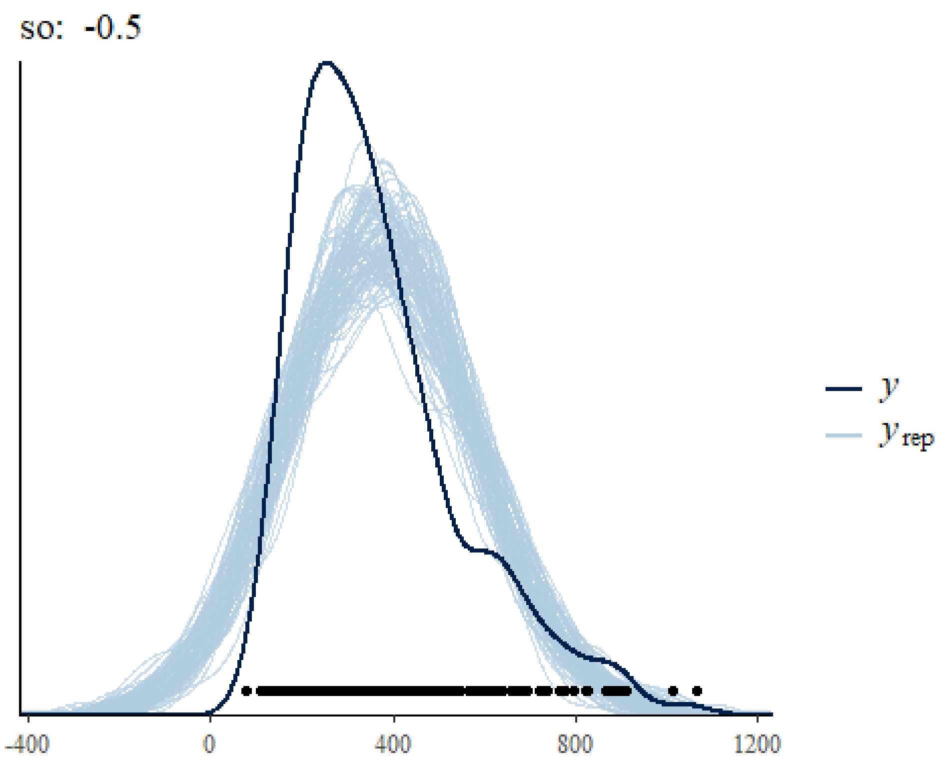
```
pp_check(m_hier_gaussian_brm, nsamples = 11, type = "hist")
```

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



- posterior predictive checks with 100 predicted data-sets (in order to evaluate our model)

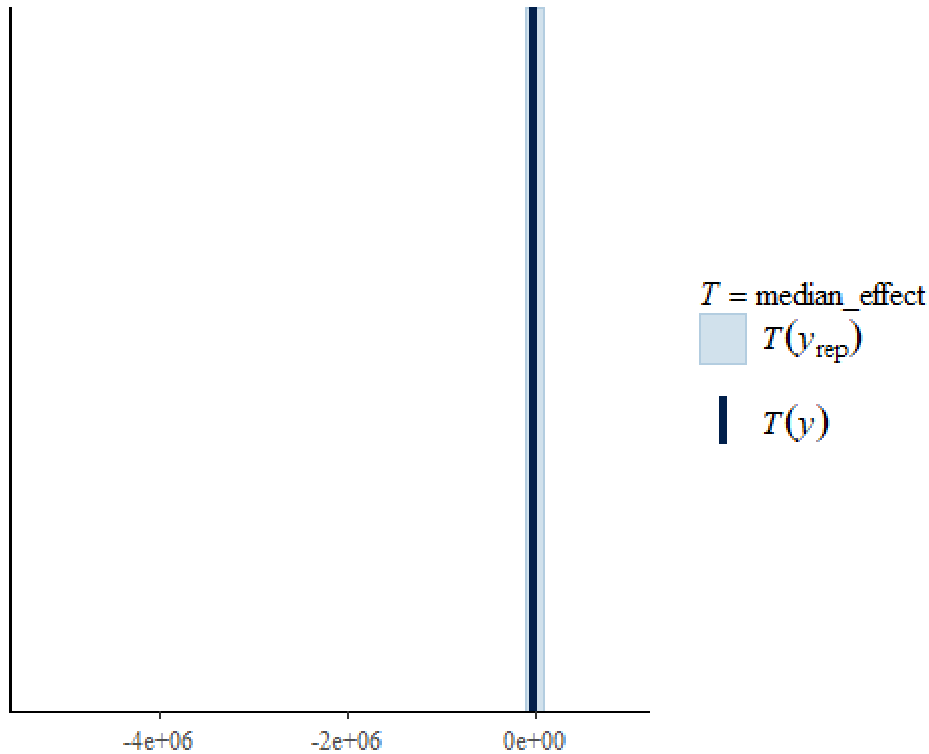




- for different conditions we got the solid line showing us what the data look like and the blue lines are the predicted data from the model. this gives us an idea how badly the model deviates from the observed data. As we see in the two figures above the posterior predicted data are not similar to the real data, *we can't have have less than zero milliseconds reading time*.

I will define a hierarchical linear model with varying intercepts and varying slopes for subject and item, assuming a correlation between the varying intercepts and slopes for both subject and item, and a lognormal likelihood.

- First i do prior predictive check to see whether our priors make sense



- Likelihood:

$rt_n \sim lognormal(\alpha + u_{subj[n],1} + w_{item[n],1} + so_n * (\beta + u_{subj[n],2} + w_{item[n],2}), \sigma)$

- Priors:

$\alpha \sim Normal(6{,}3)$

$beta \sim Normal(0{,}0.1)$

$sigma \sim Normal_{+}(0{,}1)$

$$\begin{pmatrix} u_{i,1} \\ u_{i,2} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_u\right)$$

$$\begin{pmatrix} w_{i,1} \\ w_{i,2} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_w\right)$$

Where

$$\Sigma_u = \begin{pmatrix} \tau_{u_1}^2 & \rho_u \tau_{u_1} \tau_{u_2} \\ \rho_u \tau_{u_1} \tau_{u_2} & \tau_{u_2}^2 \end{pmatrix}$$

$$\Sigma_w = \begin{bmatrix} \tau_{w_1}^2 & \rho_w \tau_{w_1} \tau_{w_2} \\ \rho_w \tau_{w_1} \tau_{w_2} & \tau_{w_2}^2 \end{bmatrix}$$

$\tau_{u_1} \sim Normal_+(0,1)$

$\tau_{u_2} \sim Normal_+(0,1)$

$p_u \sim LKJcorr(2)$

$\tau_{w_1} \sim Normal_+(0,1)$

$\tau_{w_2} \sim Normal_+(0,1)$

$\rho_w \sim LKJcorr(2)$

- In order to simplify the call to brms i will assign the same priors to the by-subject and by-item parameters
- i fit the model with 4000 iterations rather than with the default of 2000 iterations by chain. The reason is that when I run the model with the default number of iterations, I get the following warning: *Warning: Bulk Effective Samples Size (ESS) is too low, indicating posterior means and media Running the chains for more iterations may help. See http://mc-stan.org/misc/warnings.html#bulk-ess.*

```
m_hier_log_brm<-brm(rt~so+(so|subj)+(so|item),dat,
                 family = lognormal(),
                 prior =
                   c(prior(normal(6, 3), class = Intercept),
                     prior(normal(0, 0.1), class = b),
                     prior(normal(0, 1), class = sigma),
                     prior(normal(0, 1), class = sd),
                     prior(lkj(2), class = cor)),
                 iter = 4000,
                 chains = 4)
```

- the summary of the posteriors:

```
summary(m_hier_log_brm)

##  Family: lognormal
##   Links: mu = identity; sigma = identity
## Formula: rt ~ so + (so | subj) + (so | item)
##    Data: dat (Number of observations: 672)
## Samples: 4 chains, each with iter = 4000; warmup = 2000; thin = 1;
##          total post-warmup samples = 8000
##
## Group-Level Effects:
## ~item (Number of levels: 16)
##                  Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
Tail_ESS
## sd(Intercept)        0.07      0.03     0.02     0.13 1.00     2294
2209
## sd(so)               0.07      0.05     0.00     0.18 1.00     1993
2344
```

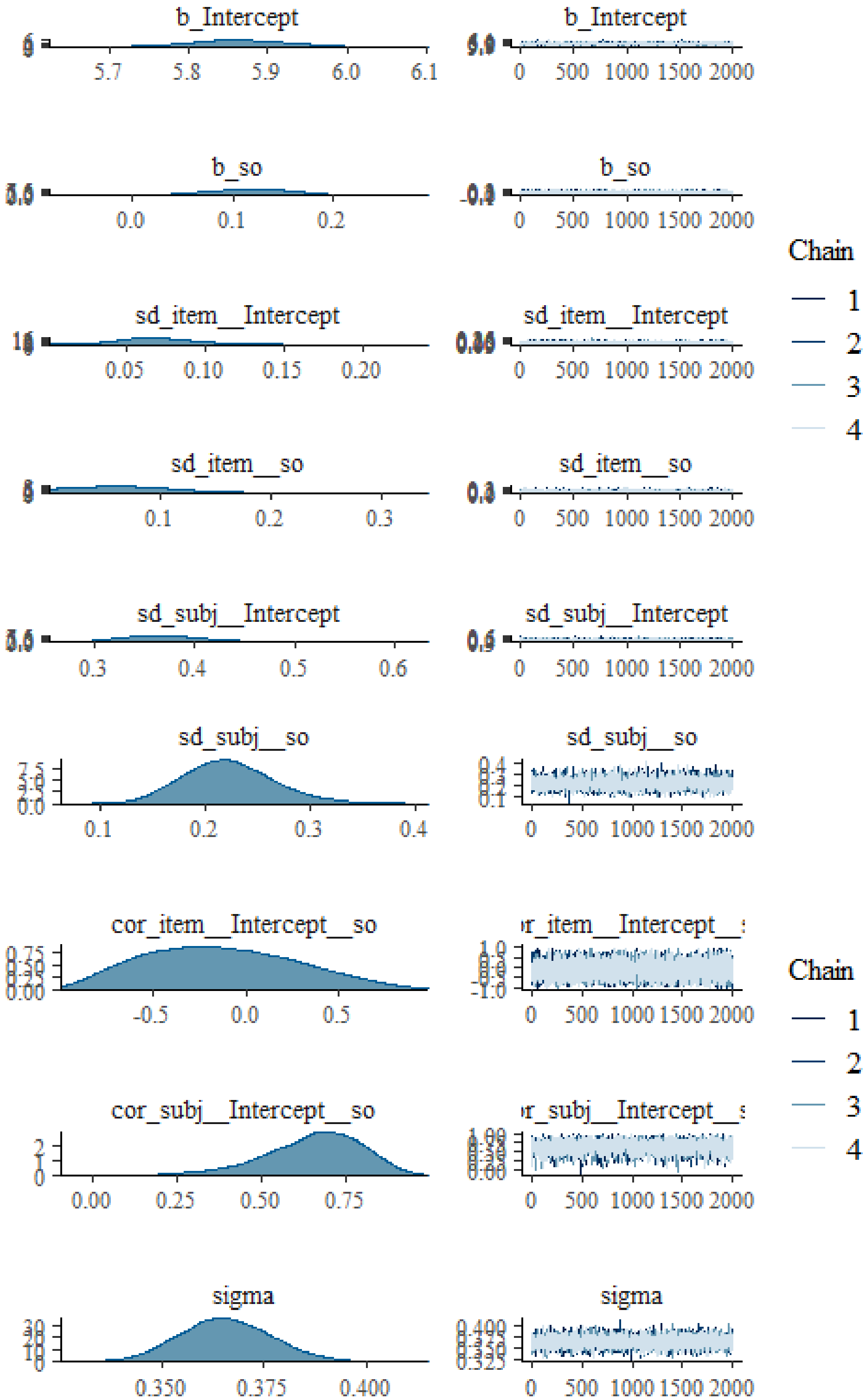
```
## cor(Intercept,so)    -0.13      0.40    -0.82     0.68 1.00     5671
5113
##
## ~subj (Number of levels: 42)
##                   Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
Tail_ESS
## sd(Intercept)         0.37      0.04     0.29     0.47 1.00     1141
2515
## sd(so)                0.22      0.04     0.15     0.31 1.00     2862
4419
## cor(Intercept,so)     0.65      0.15     0.32     0.89 1.00     3543
4292
##
## Population-Level Effects:
##           Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept     5.86      0.06     5.74     5.98 1.00      718     1655
## so            0.12      0.04     0.03     0.20 1.00     2096     3796
##
## Family Specific Parameters:
##       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma     0.37      0.01     0.34     0.39 1.00     6129     5576
##
## Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

- summarize the posteriors of all parameters in a table (the fixed effect intercept and slope, the variance components, and the correlations)

```
##                          Estimate        Q2.5     Q97.5
## b_Intercept            5.86333982  5.740548771 5.9833442
## b_so                   0.11726990  0.026867724 0.2042204
## sd_item__Intercept     0.06984281  0.023109485 0.1281647
## sd_item__so            0.07357435  0.004140397 0.1812257
## sd_subj__Intercept     0.37201840  0.294406284 0.4711587
## sd_subj__so            0.22233959  0.145252658 0.3138087
## cor_item__Intercept__so -0.13491967 -0.824169481 0.6753058
## cor_subj__Intercept__so  0.65244597  0.316920594 0.8896931
## sigma                  0.36563986  0.344353630 0.3882338
```

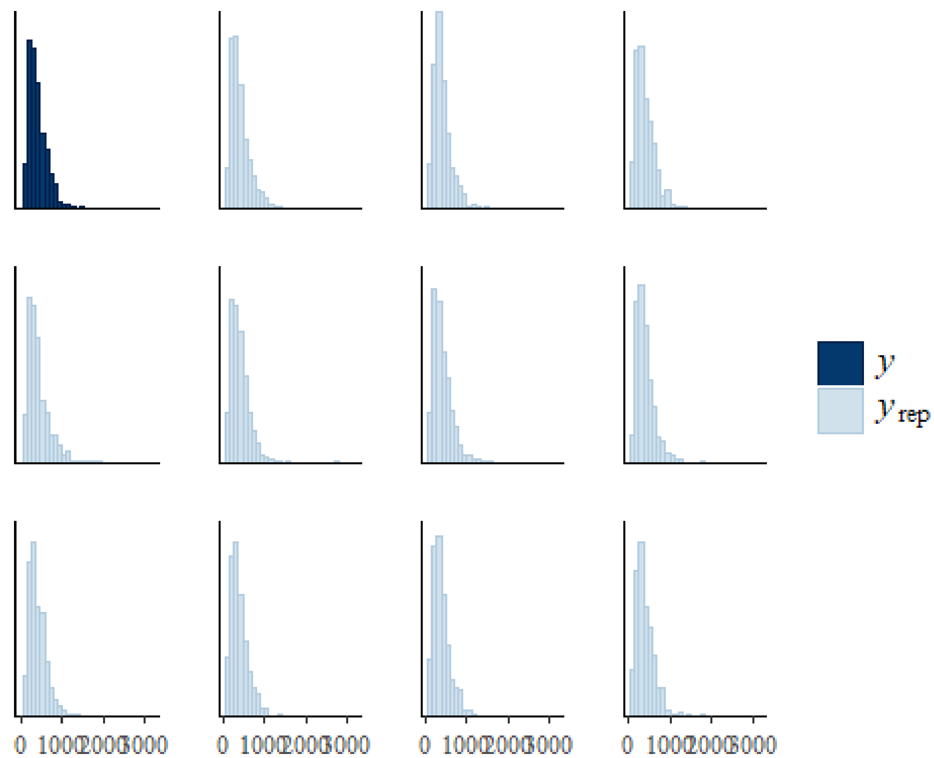
- plot The posterior distributions to obtain a graphical summary of all the parameters in the model

```
plot(m_hier_log_brm)
```



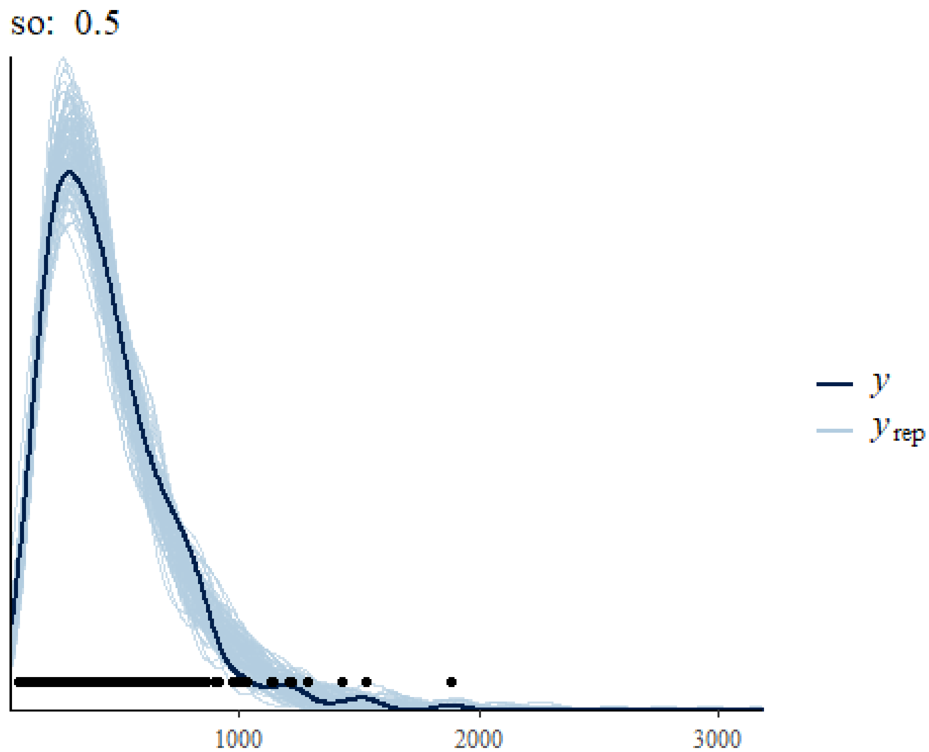
- Histograms of eleven samples from the posterior predictive distribution of the model m_hier_brm1 (the real data is skewed and has no values less than 0 ms)

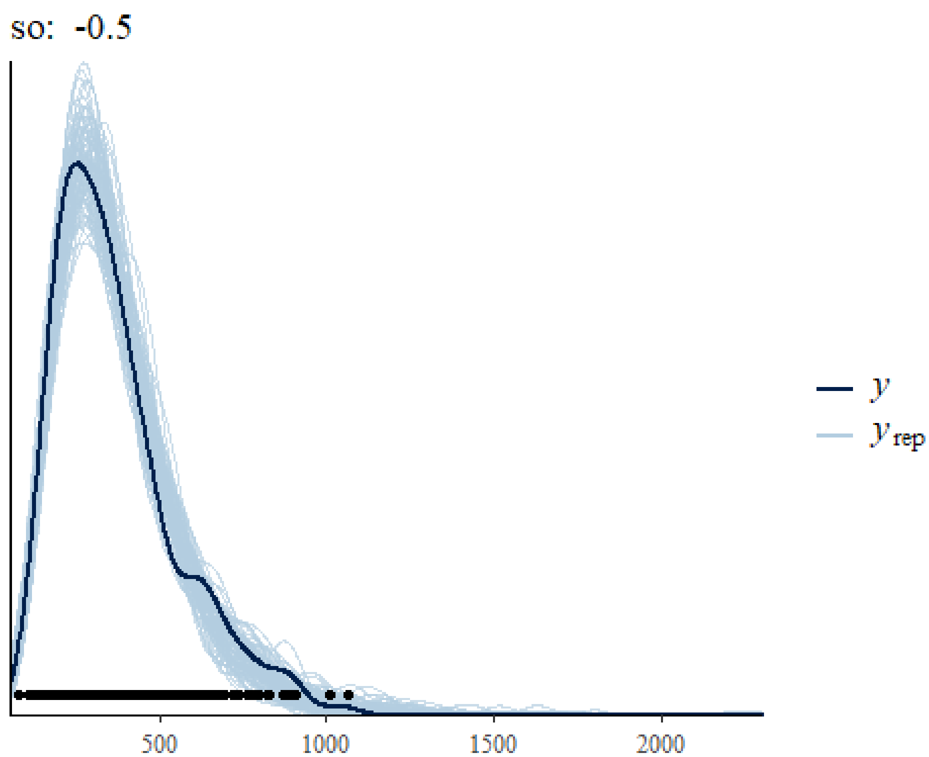
```
pp_check(m_hier_log_brm, nsamples = 11, type = "hist")

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



- posterior predictive checks with 100 predicted data-sets (in order to evaluate our model)

```
for (l in c(0.5,-0.5)){
df_dat <- filter(dat, so == l)
p <- pp_check(m_hier_log_brm, type = "dens_overlay",
nsamples = 100,
newdata = df_dat) +
geom_point(data =df_dat, aes(x = rt, y = 0.0001))+
ggtitle(paste("so: ",l))
print(p)
}
```





- As we see in the two figures above the posterior predicted data are more similar to the real data, compared to the case where we had a Normal likelihood

Conclusion

1. the gaussian model (m_hier_gaussian_brm)

Likelihood: $rt_n \sim Normal(\alpha + u_{subj[n],1} + w_{item[n],1} + so_n * (\beta + u_{subj[n],2} + w_{item[n],2}), \sigma)$

rt = 409.93 + 151.66 + 31.46 + ($\pm$ 0.5)*(78.93+99.24+29.28) + 162.97

condition → subject relative: -0.5

rt = 593.05 + (-0.5)*(207,45) + 162.97 = 593.05 - 103,725 + 162.97 = 652,295 milliseconds

condition → object relative: +0.5

rt = 593.05 + (+0.5)*(207,45) + 162.97 = 593.05 + 103,725 + 162.97 = 859,745 milliseconds

⇒ the condition affects the reading time. subject relative is faster than object relative.

2. the lognormal model (m_hier_log_brm)

Likelihood: $rt_n \sim lognormal(\alpha + u_{subj[n],1} + w_{item[n],1} + so_n * (\beta + u_{subj[n],2} + w_{item[n],2}), \sigma)$

rt = 409.93 + 151.66 + 31.46 + ($\pm$ 0.5)*(78.93+99.24+29.28) + 162.97

condition → subject relative: -0.5

rt = 593.05 + (-0.5)*(207,45) + 162.97 = 593.05 - 103,725 + 162.97 = 652,295 milliseconds

condition → object relative: +0.5

rt = 593.05 + (+0.5)*(207,45) + 162.97 = 593.05 + 103,725 + 162.97 = 859,745 milliseconds

⇒ the condition affects the reading time. subject relative is faster than object relative.

correlation parameters:

Modeling the correlation between varying intercepts and slopes means defining a covariance relationship between by-subject varying intercepts and slopes, and between by-items varying intercepts and slopes. This amounts to adding an assumption that the by-subject slopes $u_{i,2}$ could in principle have some correlation with the by-subject intercepts $u_{i,1}$; and by-item slopes $w_{i,2}$ with by-item intercept $w_{i,1}$.

1. the gaussian model (m_hier_gaussian_brm)

- the correlation ρ_u between by-subject varying intercepts and slopes is positive (0.83). that means the faster a subject's reading time is on average, the faster they read object relatives.
- the correlation ρ_w between by-item varying intercepts and slopes is also positive (0.03). that means the faster a item's reading time is on average, the faster they read object relatives.

2. the lognormal model (m_hier_log_brm)

- the correlation ρ_u between by-subject varying intercepts and slopes is negative (0.65). that means the faster a subject's reading time is on average, the faster they read object relatives.
- the correlation ρ_w between by-item varying intercepts and slopes is also positive (-0.13). that means the slower a subject's reading time is on average, the slower they read object relatives.