CGS698C, Assignment 09

Himanshu Yadav

2024-07-04

Part 1: Hypothesis testing

We have reading time data (in milliseconds) from a repeated measures reading study where the predictability of a particular word in a sentence was manipulated: the word was either predictable (coded as +1) or not predictable (coded as -1). A researcher hypothesizes that **the average reading time will be faster when the word is predictable (i.e., when *pred* is $+1$).** Let us call this the **predictive processing hypothesis**.

The data was collected from 100 participants and 36 items (sentence variants).

Load this data-frame using the following code.

```
pred_dat <-
  read.table("Predictibility-effect-data.csv",
             sep=",",header=T)[,-1]

pred_dat$rt <- round(pred_dat$rt,2)
head(pred_dat)
```

```
##   subj item pred     rt
## 1    1    1    1 267.59
## 2    1    3    1 395.74
## 3    1    5    1 338.81
## 4    1    7    1 180.28
## 5    1    9    1 316.59
## 6    1   11    1 345.30
```

0.1 Hypothesis testing using Bayes factor

The researcher wants to test the *predictive processing hypothesis* using Bayes factors. They need to compare the predictive processing model which assumes that the effect of predictability on reading times is negative against a null model which assumes no effect of predictability on reading times. If the Bayes factors are larger than 10, that would mean there is strong evidence for the predictive processing hypothesis.

The predictive processing model and the null model can be written as follows.

Predictive processing model

rt_i indicate reading time for the word i; $pred_i$ indicate predictability (+1 or -1) for the word i.

$$rt_i \sim Lognormal(\mu_i, \sigma)$$

$$\mu_i \sim \alpha + \beta pred_i$$

$$\alpha \sim Normal(6, 1.5)$$

$$\beta \sim Normal(0, 1)$$

$$\sigma \sim Normal(0, 1)$$

Null model

$$rt_i \sim Lognormal(\mu_i, \sigma)$$

$$\mu_i \sim \alpha + \beta pred_i$$

$$\alpha \sim Normal(6, 1.5)$$

$$\beta = 0$$

$$\sigma \sim Normal(0, 1)$$

The following code implements the parameter estimation using brms package.

```
priors_pred <-
  c(set_prior("normal(6, 1.5)",
                      class = "Intercept"),
            set_prior("normal(0, 1)",
                      class = "b",coef = "pred"),
            set_prior("normal(0, 1)",
                      class = "sigma"))

priors_null <-
  c(set_prior("normal(6, 1.5)",
              class = "Intercept"),
    set_prior("normal(0, 1)",
              class = "sigma"))


m_pred <- brm(rt ~ 1+pred,
              data=pred_dat,
              family=lognormal(),
              prior=priors_pred,
              cores=4,
              iter=20000,
              warmup = 2000,
```

```
              save_pars = save_pars(all = TRUE))

m_null <- brm(rt ~ 1,
              data=pred_dat,
              family=lognormal(),
              prior=priors_null,
              cores=4,
              iter=20000,
              warmup = 2000,
              save_pars = save_pars(all = TRUE))

save(m_pred, file="FittedModels/pred.rda")
save(m_null, file="FittedModels/null.rda")
```

The following code estimates the Bayes factor for the predictive processing model compared to null model using the bridgesampling package.

```
library(bridgesampling)

# Marginal log likelihood for pred model
margLogLik_pred <-
  bridge_sampler(m_pred, silent = TRUE)

margLogLik_null <-
  bridge_sampler(m_null, silent = TRUE)

BF <- bays_factor(margLogLik_pred,
                   margLogLik_null)

BF$bf

## [1] 3.459723e+108
```

Calculate Bayes factors using the above method for the following priors on the parameter β, the effect of predictability on (log) reading times.

- $\beta \sim Normal(0, 0.01)$
- $\beta \sim Normal(0, 0.1)$
- $\beta \sim Normal(0, 0.5)$
- $\beta \sim Normal(0, 5)$

Is there evidence for the predictive processing hypothesis given the above prior assumptions about the effect of interest?

0.2 Hypothesis testing using cross-validation

Evaluate evidence for the predictive processing hypothesis using k-fold cross validation. (use k=6)

(Hint: You can use the cross validation code given at the end of Assignment 8 but also estimate the standard error of the elpd difference.)

Part 2: Hierarchical models and Hypothesis testing

Recall our example study where we test the effect of cognitive load on pupil size, but this time we look at all the subjects of the Wahn et al. (2016) study.

```
dat <- read.csv("df_pupil_complete.csv")
colnames(dat)
```

```
## [1] "X"      "subj"   "trial"  "load"   "p_size"
```

```
dat$c_load <- dat$load - mean(dat$load)
```

You should be able to now fit a "maximal" model (correlated varying intercept and slopes for subjects) assuming a normal likelihood. You can use the same priors as we used previously for this load-pupil size study.

2.1 Hierarchical model

Fit a hierarchical model which assumes that the average pupil size and the effect of load on pupil size can vary across subjects and also that average pupil sizes and load-effects can be correlated at the indvidual level. This kind of hierarchical model can be called a **by-subject correlated varying intercept varying slope model**. The model can be specified as follows:

Suppose $psize_i$ is the observed pupil size in the i^{th} observation.

$$psize_i \sim Normal(\mu_i, \sigma)$$

$$\mu_i = \alpha_j + \beta_j load_i$$

where j is the id of the subject who produced the i^{th} observation. The intercedpt and the slope for the j^{th} subject are given by

$$\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim BivariateNormal\left(\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \begin{pmatrix} \tau_a^2 & \rho\tau_a\tau_b \\ \rho\tau_a\tau_b & \tau_b^2 \end{pmatrix} \right)$$

where α and β are population-level mean intercept and slope respectively; τ_a^2 and τ_b^2 are population-level variance in intercept and slope respectively; ρ is correlation between subject-level intercept and slope.

We choose the following priors on parameters in the above likelihood.

$$\alpha \sim Normal(1000, 500)$$

$$\beta \sim Normal(0, 100)$$

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

$$\tau_a \sim Normal_{+}(0, 250)$$

$$\tau_b \sim Normal_{+}(0, 250)$$

$$\rho \sim LKJ(2)$$

The above model can be implemented in brms using the following code.

```
m1.full <- brm(p_size ~ 1+c_load + (1+c_load|subj),
               dat=dat,
               prior = c(prior(normal(1000,500),class=Intercept),
                         prior(normal(0,500),class=sigma),
                         prior(normal(0,100,class=b)),
                         prior(normal(0,250),class=sd),
                         prior(lkj(2),class=corr)),
               chains=4,cores=4,
               iter=4000,
               warmup = 2000)
```

Fit the above model and estimate the effect of cognitive load on pupil size. What is the estimate of the effect size? Is it different from the estimate obtained from a non-hierarchical model? (Report the 95% credible interval of the effect size.)

2.2 *Hypothesis testing*

The parameter estimation does not evaluate our hypothesis. From the 95% credible interval, we cannot conclude whether our hypothesis is correct or not given that data. For hypothesis testing and in general for quantifying evidence for any assumption, we need to do model comparison.

Here, to test the hypothesis that "attentional load affects pupil size", we can do a Bayes factor analysis where a model assuming non-zero effect of cognitive load is compared against a model assuming zero effect of load.

It is important to use a "maximal" hierarchical model for hypothesis testing because we want to use a likelihood function that provides best approximation of how the data were generated. Since, data in reality were generated from individual subjects, it is unreasonable to assume that all individuals have the same average pupil size and are equally sensitive to cognitive load.

You can use the following code to estimate Bayes factors for model assuming non-zero effect of load. Note that we need to fit the model under different prior assumptions about our parameter of interest. What do you conclude from the bayes factors estimated under different prior assumptions about the effect of interest?

```
library(bridgesampling)

priors_null <- c(set_prior("normal(1000, 500)",
                          class = "Intercept"),
                 set_prior("normal(0, 500)",
                          class = "sigma"),
                 set_prior("normal(0, 250)",
                          class = "sd"),
                 set_prior("lkj(2)",
                          class = "corr"))

mfit.null <- brm(p_size ~ 1 + (1+c_load|subj),
               dat=dat,
               prior = priors_null,
               chains=4,cores=4,
               iter=20000,
               warmup = 2000,
               save_pars = save_pars(all = TRUE))

save(mfit.null,file="Pupil-null-model.Rda")
margLogLik_null <-
  bridge_sampler(mfit.null, silent = TRUE)


BayesFactors <- rep(NA,6)
prior_sd <- c(5, 25, 50, 100, 200, 500)

for(k in 1:6){
  p <- as.character(prior_sd[k])
  priors_full <- c(set_prior("normal(1000, 500)",
                          class = "Intercept"),
                 set_prior( paste0("normal(0, ",p,")"),
                           class = "b"),
                 set_prior("normal(0, 500)",
                          class = "sigma"),
                 set_prior("normal(0, 250)",
                          class = "sd"),
                 set_prior("lkj(2)",
                          class = "corr"))


  mfit.full <- brm(p_size ~ 1+c_load + (1+c_load|subj),
               dat=dat,
               prior = priors_full,
```

```
              chains=4,cores=4,
              iter=20000,
              warmup = 2000,
              save_pars = save_pars(all = TRUE))

  save(mfit.full,file=paste0("Pupil-full-model-prior-sd-",p,".Rda"))

  margLogLik_full <-
    bridge_sampler(mfit.full, silent = TRUE)

  BF <- bayes_factor(margLogLik_full,
                     margLogLik_null)

  print(BF$bf)
  BayesFactors[k] <- BF$bf
}
```