Bayesian regression models

LOT winter school 2020

Bruno Nicenboim / Shravan Vasishth

2020-01-15

A first linear model: Does attentional load affect pupil size?

Log-normal model: Does trial affect pupil size?

A first linear model: Does attentional load affect pupil size?

Data:
One participant's pupil size of the control experiment of Wahn et al. (2016) averaged by trial

Task:
A participant covertly tracked between zero and five objects among several randomly moving objects on a computer screen; multiple object tracking–MOT– (Pylyshyn and Storm 1988) task

Research question:
How does the number of moving objects being tracked (attentional load) affect pupil size?

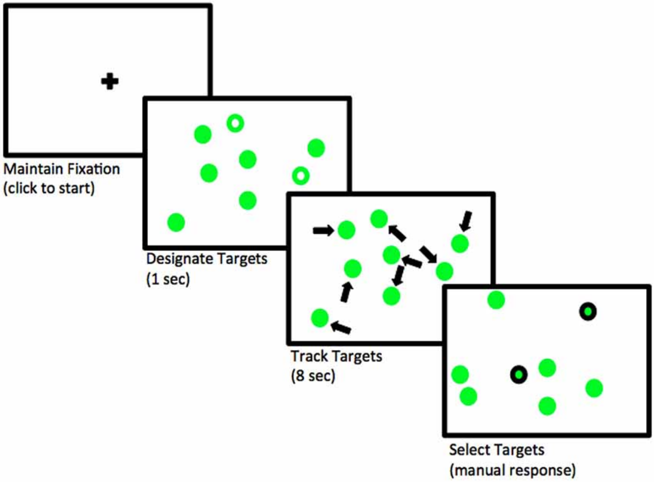


Figure 1: Flow of events in a trial where two objects needs to be tracked. Adapted from Blumberg, Peterson, and Parasuraman (2015); licensed under CC BY 4.0.

Assumptions:

1. There is some average pupil size represented by α.
2. The increase of attentional load has a linear relationship with pupil size, determined by β.
3. There is some noise in this process, that is, variability around the true pupil size i.e., a scale, σ.
4. The noise is normally distributed.

Formal model

Likelihood for each observation n:

$$p\_size_n \sim Normal(\alpha + c\_load_n \cdot \beta, \sigma) \tag{1}$$

where n indicates the observation number with $n = 1 \dots N$

How do we decide on priors?

Priors

- pupil sizes range between 2 and 5 millimeters,
- but the Eyelink-II eyetracker measures the pupils in arbitrary units (Hayes and Petrov 2016)
- we either need estimates from a previous analysis or look at some measures of pupil sizes

Pilot data:
Some measurements of the same participant with no attentional load for the first 100ms, each 10 ms, in `pupil_pilot.csv`:

```
df_pupil_pilot <- read_csv("./data/pupil_pilot.csv")
df_pupil_pilot$p_size %>% summary()
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     852     856     862     861     866     868
```

Prior for α

$$\alpha \sim Normal(1000, 500) \tag{2}$$

Meaning:
We expect that the average pupil size for the average load in the experiment would be in a 95% central interval limited by approximately $1000 \pm 2 \cdot 500 = [20, 2000]$ units:

```
c(qnorm(.025, 1000, 500), qnorm(.975, 1000, 500))
```

```
## [1]   20 1980
```

Prior for σ

$$\sigma \sim Normal_{+}(0, 1000) \tag{3}$$

Meaning:
We expect that the standard deviation of the pupil sizes should be in the following 95% interval.

```
c(
  qtnorm(.025, 0, 1000, a = 0),
  qtnorm(.975, 70, 1000, a = 0)
)
```

```
## [1]   31 2290
```

Prior for β

$$\beta \sim Normal(0, 100) \tag{4}$$

Meaning:
We don't really know if the attentional load will increase or even decrease the pupil size, but we are only saying that one unit of load will potentially change the pupil size consistently with the following 95% interval:

```
c(qnorm(.025, 0, 100), qnorm(.975, 0, 100))
```

```
## [1] -196  196
```

Fitting the model

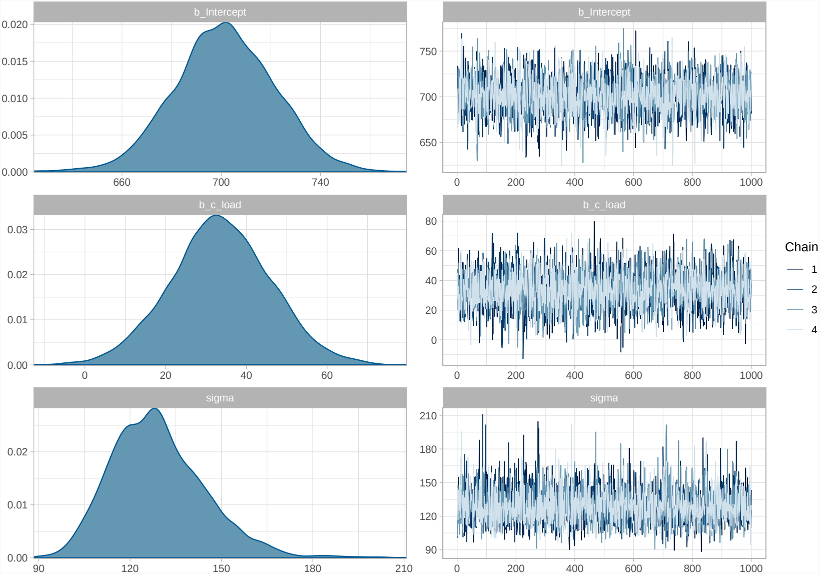
```
df_pupil_data <- read_csv("data/pupil.csv")
df_pupil_data <- df_pupil_data %>%
  mutate(c_load = load - mean(load))
df_pupil_data
```

```
## # A tibble: 41 x 4
##   trial  load p_size c_load
##   <dbl> <dbl>  <dbl>  <dbl>
## 1     1     2  1021. -0.439
## 2     2     1   951. -1.44
## 3     3     5  1064.  2.56
## 4     4     4   913.  1.56
## 5     5     0   603. -2.44
## # ... with 36 more rows
```

Specifying the model in brms

```r
fit_pupil <- brm(p_size ~ 1 + c_load,
  data = df_pupil_data,
  family = gaussian(),
  prior = c(
    prior(normal(1000, 500), class = Intercept),
    prior(normal(0, 1000), class = sigma),
    prior(normal(0, 100), class = b, coef = c_load)
  )
)
```

```r
plot(fit_pupil)
```



```
fit_pupil
```

```
##  Family: gaussian
##   Links: mu = identity; sigma = identity
## Formula: p_size ~ 1 + c_load
##    Data: df_pupil_data (Number of observations: 41)
## Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
##          total post-warmup samples = 4000
##
## Population-Level Effects:
##           Estimate Est.Error l-95% CI u-95% CI Rhat
## Intercept   701.35     20.33   661.86   740.74 1.00
## c_load       33.63     12.44     9.19    58.40 1.00
##           Bulk_ESS Tail_ESS
## Intercept     3748     2801
## c_load        3499     2752
##
## Family Specific Parameters:
##       Estimate Est.Error l-95% CI u-95% CI Rhat
## sigma   129.23     15.55   103.64   163.41 1.00
##       Bulk_ESS Tail_ESS
## sigma     3173     2488
##
## Samples were drawn using sampling(NUTS). For each parameter, Eff.Sample
## is a crude measure of effective sample size, and Rhat is the potential
```

How to communicate the results?

Research question:
"What is the effect of attentional load on the participant's pupil size?"

We'll need to examine what happens with β (`c_load`):

How to communicate the results?

- The most likely values of β will be around the mean of the posterior, 33.63, and we can be 95% certain that the true value of β *given the model and the data* lies between 9.19 and 58.4.
- We see that as the attentional load increases, the pupil size of the participant becomes larger.

How likely it is that the pupil size increased rather than decreased?

```
mean(posterior_samples(fit_pupil)$b_c_load > 0)
```

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

Take into account that this probability ignores the possibility of the participant not being affected at all by the manipulation, this is because $P(\beta = 0) = 0$.

Descriptive adequacy

```r
# we start from an array of 1000 samples by 41 observations
df_pupil_pred <- posterior_predict(fit_pupil, nsamples = 1000) %>%
  # we convert it to a list of length 1000, with 41 observations in each element
  array_branch(margin = 1) %>%
  # We iterate over the elements (the predicted distributions)
  # and we convert them into a long data frame similar to the data,
  # but with an extra column `iter` indicating from which iteration
  # the sample is coming from.
  map_dfr(function(yrep_iter) {
    df_pupil_data %>%
      mutate(p_size = yrep_iter)
  }, .id = "iter") %>%
  mutate(iter = as.numeric(iter))
```

```r
df_pupil_pred %>% filter(iter < 100) %>%
    ggplot(aes(p_size, group=iter)) +
    geom_line(alpha = .05, stat="density", color = "blue") +
    geom_density(data=df_pupil_data, aes(p_size),
                 inherit.aes = FALSE, size =1)+
    geom_point(data=df_pupil_data, aes(x=p_size, y = -0.001), alpha =.5,
               inherit.aes = FALSE)+
    coord_cartesian(ylim=c(-0.002, .01))+ facet_grid(load ~ .)
```

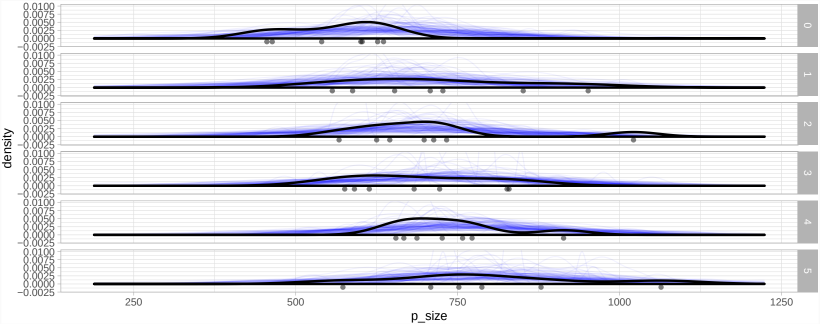


Figure 2: The plot shows 100 predicted distributions in blue density plots, the distribution of pupil size data in black density plots, and the observed pupil sizes in black dots for the five levels of attentional load.

Distribution of statistics

```r
# predicted means:
df_pupil_pred_summary <- df_pupil_pred %>%
  group_by(iter, load) %>%
  summarize(av_p_size = mean(p_size))
# observed means:
(df_pupil_summary <- df_pupil_data %>%
  group_by(load) %>%
  summarize(av_p_size = mean(p_size)))
```

```
## # A tibble: 6 x 2
##    load av_p_size
##   <dbl>     <dbl>
## 1     0      561.
## 2     1      719.
## 3     2      715.
## 4     3      691.
## 5     4      740.
## # ... with 1 more row
```

```r
ggplot(df_pupil_pred_summary, aes(av_p_size)) +
  geom_histogram(alpha = .5) +
  geom_vline(aes(xintercept = av_p_size), data = df_pupil_summary) +
  facet_grid(load ~ .)
```

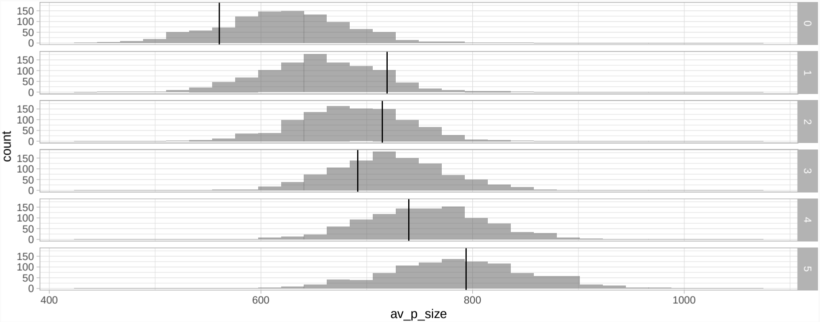


Figure 3: Distribution of posterior predicted means in gray and observed pupil size means in black lines by load.

- the observed means for no load and for a load of two are falling in the tails of the distributions.
- the data might be indicating that the relevant difference is between (i) no load, (ii) a load between two and three, and then (iii) a load of four, and (iv) of five.
- but beware of overinterpreting noise.

Value of posterior predictive distributions

- If we look hard enough, we'll find failures of descriptive adequacy.[1]
- Posterior predictive accuracy can be used to generate new hypotheses and to compare different models.

[1] all models are wrong

Log-normal model: Does trial affect pupil size?

We revisit the small experiment, where a participant repeatedly pressed the space bar as fast as possible, without paying attention to the stimuli.

New research question:
Does the participant tend to speedup (practice effect) or slowdown (fatigue effect)?

Formal model

Likelihood:

$$rt_n \sim LogNormal(\alpha + c\_trial_n \cdot \beta, \sigma) \tag{5}$$

Priors

$$\begin{aligned} \alpha &\sim Normal(6, 1.5) \\ \sigma &\sim Normal_+(0, 1) \\ \beta &\sim ... \end{aligned} \tag{6}$$

Prior for β

$$\beta \sim Normal(0, 1) \tag{7}$$

We edit our `normal_predictive_distribution_fast` from section and make it log-normal and dependent on trial:

```
lognormal_model_pred <- function(alpha_samples,
                                 beta_samples,
                                 sigma_samples,
                                 N_obs) {
    # pmap extends map2 (and map) for a list of lists:
    pmap_dfr(list(alpha_samples, beta_samples, sigma_samples),
             function(alpha, beta, sigma) {
                 tibble(
                     trialn = seq_len(N_obs),
                     # we center trial:
                     c_trial = trialn - mean(trialn),
                     # we change the likelihood:
                     # Notice rlnorm and the use of alpha and beta
                     rt_pred = rlnorm(N_obs, alpha + c_trial * beta, sigma))
             }, .id = "iter") %>%
    # .id is always a string and needs to be converted to a number
        mutate(iter = as.numeric(iter))}
```

This is our first attempt for a prior predictive distribution:

```
N_obs <- 361
N <- 800
alpha_samples <- rnorm(N, 6, 1.5)
sigma_samples <- rtnorm(N, 0, 1, a = 0)
beta_samples <- rnorm(N, 0, 1)
prior_pred <- lognormal_model_pred(
  alpha_samples = alpha_samples,
  beta_samples = beta_samples,
  sigma_samples = sigma_samples,
  N_obs = N_obs
)
(median_effect <-
  prior_pred %>%
  group_by(iter) %>%
  mutate(diff = rt_pred - lag(rt_pred)) %>%
  summarize(
    median_rt = median(diff, na.rm = TRUE)
  ))
```

```
## # A tibble: 800 x 2
##    iter median_rt
##   <dbl>     <dbl>
## 1     1  1.40e- 5
```

```
median_effect %>%
  ggplot(aes(median_rt)) +
  geom_histogram()
```

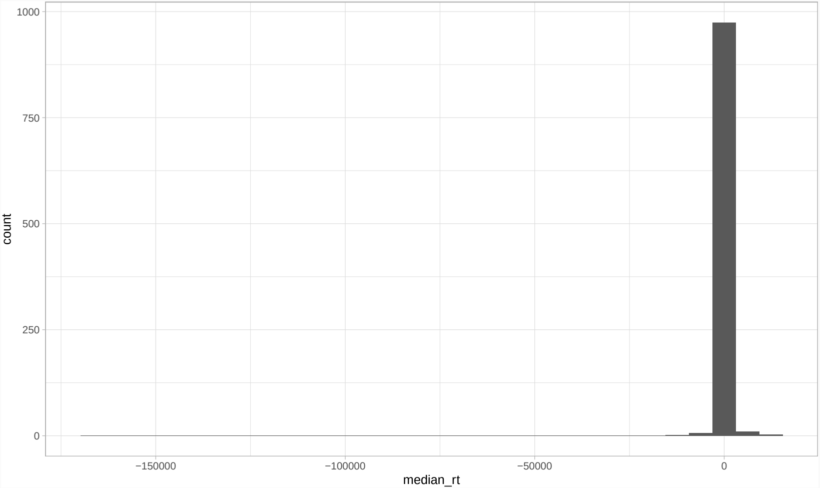


Figure 4: Prior predictive distribution of the median effect of the log-normal model with $\beta \sim Normal(0, 1)$.

Another prior for β

$$\beta \sim Normal(0, .01) \tag{8}$$

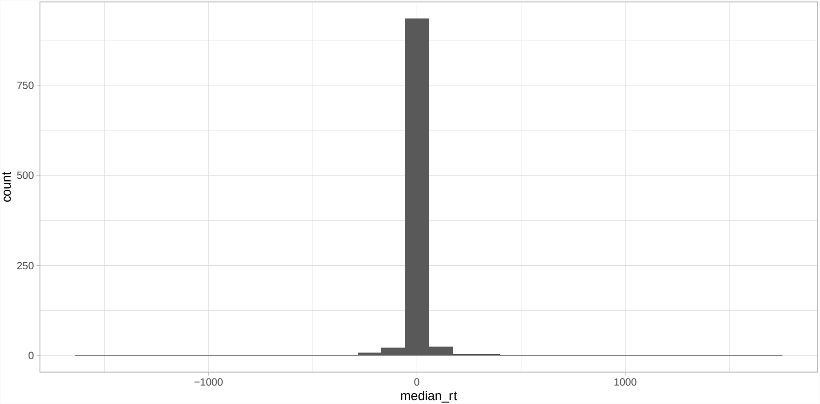


Figure 5: Prior predictive distribution of the median effect of the log-normal model with $\beta \sim Normal(0, .01)$.

Prior selection

Prior selection might look daunting and a lot of work.
However...

- priors can be informed by the estimates from previous experiments;
- this work is usually done only the first time we encounter an experimental paradigm;
- we will generally use very similar (or identical priors) for analyses dealing with the same type of task;
- when in doubt, do a sensitivity analysis.

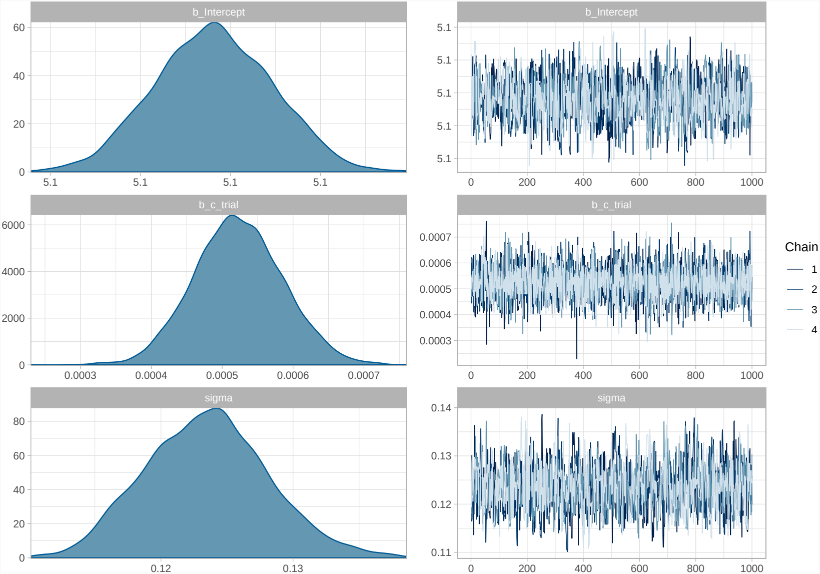
Fitting the model

```
df_noreading_data <- read_csv("./data/button_press.csv")
df_noreading_data <- df_noreading_data %>%
  mutate(c_trial = trialn - mean(trialn))
fit_press_trial <- brm(rt ~ 1 + c_trial,
  data = df_noreading_data,
  family = lognormal(),
  prior = c(
    prior(normal(6, 1.5), class = Intercept),
    prior(normal(0, 1), class = sigma),
    prior(normal(0, .01), class = b, coef = c_trial)
  )
)
```

```r
posterior_summary(fit_press_trial)[, c("Estimate", "Q2.5", "Q97.5")]
```

```
##                Estimate       Q2.5       Q97.5
## b_Intercept     5.11823     5.1059     5.13144
## b_c_trial       0.00052     0.0004     0.00065
## sigma           0.12338     0.1143     0.13333
## lp__        -1603.74209 -1606.9599 -1602.29341
```

```r
plot(fit_press_trial)
```



How to communicate the results?

We focus on the effect of trial:

- $\hat{\beta} = 0.00052$, 95% CrI $= [0.0004, 0.00065]$.
- But in most cases, the effect is easier to interpret in milliseconds.

We calculate an estimate if we consider the difference between reaction times in a trial at the middle of the experiment (when the centered trial number is zero) and the previous one (when the centered trial number is minus one).

```
alpha_samples <- posterior_samples(fit_press_trial)$b_Intercept
beta_samples <- posterior_samples(fit_press_trial)$b_c_trial
effect_middle_ms <- exp(alpha_samples) - exp(alpha_samples - 1 * beta_samples)
## ms effect in the middle of the expt (mean trial vs. mean trial - 1 )
c(mean = mean(effect_middle_ms), quantile(effect_middle_ms, c(.025, .975)))
```

```
##  mean  2.5%   98%
## 0.088 0.067 0.108
```

Alternatively we consider the difference between the second trial and the first one:

```
first_trial <- min(df_noreading_data$c_trial)
second_trial <- min(df_noreading_data$c_trial) + 1
effect_beginning_ms <- exp(alpha_samples + second_trial * beta_samples) -
  exp(alpha_samples + first_trial * beta_samples)
## ms effect from first to second trial:
c(mean = mean(effect_beginning_ms), quantile(effect_beginning_ms, c(.025, .975)))
```

```
##  mean  2.5%   98%
## 0.080 0.062 0.096
```

There is a slowdown in both cases.

Reporting results

We can

- present the posterior mean and the 95% credible interval;
- assess if the observed estimates are consistent with the prediction from our theory;
- assess the practical relevance of the effect for the research question; (only after 100 button presses do we see a slowdown of 9 ms on average ($0.09 \cdot 100$), with a 95% credible interval ranging from 6.68 to 10.8);
- establish the presence or absence of an effect (Bayes factor)

References

Blumberg, Eric J., Matthew S. Peterson, and Raja Parasuraman. 2015. "Enhancing Multiple Object Tracking Performance with Noninvasive Brain Stimulation: A Causal Role for the Anterior Intraparietal Sulcus." *Frontiers in Systems Neuroscience* 9: 3. https://doi.org/10.3389/fnsys.2015.00003.

Hayes, Taylor R., and Alexander A. Petrov. 2016. "Mapping and Correcting the Influence of Gaze Position on Pupil Size Measurements." *Behavior Research Methods* 48 (2): 510–27. https://doi.org/10.3758/s13428-015-0588-x.

Pylyshyn, Zenon W., and Ron W. Storm. 1988. "Tracking Multiple Independent Targets: Evidence for a Parallel Tracking Mechanism." *Spatial Vision* 3 (3): 179–97.