

STA 610L: MODULE 2.6

RANDOM EFFECTS ANOVA (PRIORS FOR GROUP-LEVEL VARIANCE)

DR. OLANREWAJU MICHAEL AKANDE

READING

The following slides lean heavily on Andrew Gelman's 2006 paper on priors for variance components in hierarchical models.

I strongly suggest you go through the paper on your own.

HIERARCHICAL NORMAL MODEL

Recall our model (without group-specific variances):

$$\begin{aligned}y_{ij} | \mu_j, \sigma^2 &\sim \mathcal{N}(\mu_j, \sigma^2); \quad i = 1, \dots, n_j \\ \mu_j | \mu, \tau^2 &\sim \mathcal{N}(\mu, \tau^2); \quad j = 1, \dots, J,\end{aligned}$$

also written as

$$\begin{aligned}y_{ij} &= \mu + \alpha_j + \varepsilon_{ij} = \mu_j + \varepsilon_{ij}; \quad i = 1, \dots, n_j \\ \varepsilon_{ij} &\stackrel{iid}{\sim} N(0, \sigma^2); \quad \mu_j | \mu, \tau^2 \sim \mathcal{N}(\mu, \tau^2); \quad j = 1, \dots, J,\end{aligned}$$

with priors

$$\begin{aligned}\pi(\mu) &= \mathcal{N}(\mu_0, \gamma_0^2) \\ \pi(\tau^2) &= \mathcal{IG}\left(\frac{\eta_0}{2}, \frac{\eta_0 \tau_0^2}{2}\right) \\ \pi(\sigma^2) &= \mathcal{IG}\left(\frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2}\right).\end{aligned}$$

RELATIVELY NONINFORMATIVE PRIORS

It is often of interest to specify relatively noninformative priors for parameters.

In general, we have enough data so that we can use any reasonable prior distribution for μ and σ , for example $p(\mu, \sigma) \propto 1$.

Picking a "flattish prior" for τ^2 is trickier because sometimes our data do not contain a lot of information about this parameter -- e.g., we may have relatively few groups, which is the case in our bike data.

This parameter is also problematic in frequentist models -- in particular there is no "always-non-negative" classical unbiased estimator for it.

HIGH VARIANCE PRIORS

One basic idea is to base a prior on a proper density but inflate the variance so that its shape gets pretty flat.

Two commonly considered priors for τ include the uniform(0,A) as $A \rightarrow \infty$ and $\mathcal{IG}(\varepsilon, \varepsilon)$ as $\varepsilon \rightarrow 0$.

LIMITS OF PRIOR DISTRIBUTIONS

Gelman shows

- the uniform(0,A) prior on τ yields a limiting proper posterior density as $A \rightarrow \infty$ as long as we have at least 3 groups. The implication is that if we pick A big enough, our resulting inferences are not sensitive to the choice of A.
- the $\mathcal{IG}(\varepsilon, \varepsilon)$ for τ^2 does *not* have a proper limiting posterior, so that posterior inferences are sensitive to the value of ε chosen, particularly when small values of τ^2 are reasonable.

Unfortunately, these 'small ε ' priors became quite popular due to being used in many illustrative examples in the WinBUGS manuals, though they are no longer recommended.

PARAMETER EXPANSION

Gelman (2006) proposes a parameter expansion that facilitates a family of weakly-informative prior distributions.

$$y_{ij} \sim N(\mu + \xi\eta_j, \sigma^2)$$

$$\eta_j \sim N(0, \sigma_\eta^2)$$

In this case the parameters α_j correspond to $\xi\eta_j$ in this model, and τ corresponds to $|\xi|\sigma_\eta$ here.

PARAMETER EXPANSION

For simplicity, consider independent priors on ξ and σ_η . A conditionally-conjugate choice for ξ is normal -- and given that the scale of ξ cannot be separately identified from that of η_j , a $N(0,1)$ is convenient.

A conditionally-conjugate prior family for σ_η^2 is inverse gamma, and when combined with the $N(0,1)$ we have an implied half-t prior for τ , which is the absolute value of a Student's t distribution centered at zero.

HALF T

The half t prior is a function of a scale parameter A and degrees of freedom ν and can be written

$$p(\tau) \propto \left(1 + \frac{1}{\nu} \left(\frac{\tau}{A}\right)^2\right)^{-(\nu+1)/2}.$$

Gelman proposes a half-Cauchy prior (obtained with a large scale parameter A and $\nu = 1$) as a weakly-informative choice that is desirable at times with a small number of groups.

EXAMPLE: HOSPITAL INCOME

Recall the lab data from the Centers for Medicare and Medicaid Services on hospital costs and profit from the 2014 fiscal year.

Our interest is in examining variability of net income to the hospital across states.

```
load("data/hc2014.RData")  
#library(rstan)  
rstan_options(auto_write = TRUE)  
options(mc.cores = parallel::detectCores())
```

EXAMPLE: HOSPITAL INCOME

We can fit the model

$$y_{ij} = \mu_j + \varepsilon_{ij},$$

where y_{ij} represents the net income of hospital i in state j , and μ_j is the specific income for state j .

```
#library(brms)
m1 <- brm(netincome ~ 0 + (1|state), data=hc2014, family = gaussian(),
  control = list(adapt_delta = .99),
  prior = c(set_prior("student_t(3, 0, 1)", class = "sd"))
) #By default, sigma also has a half student-t prior
summary(m1)
```

EXAMPLE: HOSPITAL INCOME

```
## Family: gaussian
## Links: mu = identity; sigma = identity
## Formula: netincome ~ 0 + (1 | state)
## Data: hc2014 (Number of observations: 6170)
## Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
## total post-warmup draws = 4000
##
## Group-Level Effects:
## ~state (Number of levels: 55)
## Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept) 1.09 1.28 0.04 4.05 1.00 6045 2149
##
## Family Specific Parameters:
## Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma 94845657.69 850103.95 93189555.98 96536915.38 1.00 10650 2703
##
## Draws were sampled 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).
```

WHAT'S NEXT?

MOVE ON TO THE READINGS FOR THE NEXT MODULE!