

Bayesian Thinking: Fundamentals, Regression and Multilevel Modeling

Jim Albert and Jingchen (Monika) Hu

November 2020

Webinar 3-1: Multilevel Models for Continuous Data

- 1 Introduction: observations in groups
- 2 A multilevel model with random σ
- 3 MCMC estimation and diagnostics
- 4 Additional Bayesian inferential questions

Section 1

Introduction: observations in groups

Review: the normal model & normal regression

- When you have continuous outcomes, you can use a normal model:

$$Y_i \mid \mu, \sigma \stackrel{i.i.d.}{\sim} \text{Normal}(\mu, \sigma), \quad i = 1, \dots, n.$$

- When you have predictor variables available, $\{x_{i1}, \dots, x_{ip}\}$; you can specify an observation specific mean:

$$Y_i \mid \mu_i, \sigma \stackrel{ind}{\sim} \text{Normal}(\mu_i, \sigma), \quad i = 1, \dots, n,$$

where

$$\mu_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots, \beta_p x_{ip}.$$

- Predictors can be continuous, binary, and categorical (interpretation differs).
- Observations are assumed independent.

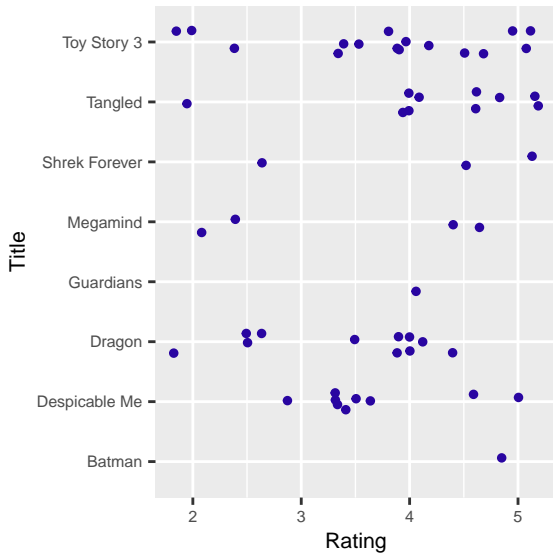
When observations are not necessarily independent

- Observations can be dependent in several ways.
- Observations are nested in groups:
 - Students' test scores from multiple schools;
 - Ratings of movies of different genres;
 - Ratings of dramas of different schedules;
 - Death rates of hospitals.
- We will focus on a movie rating dataset to explore modeling approaches for dependent data.

Example: ratings of animation movies

- MovieLens: personalized movie recommendation for users.
- In one study, a sample on movie ratings for 8 animation movies released in 2010, total 55 movies.
- Each rating is for a movie completed by a user; some movies have many ratings while others have few.
- A natural grouping of these 55 ratings: by movie title.

Example: ratings of animation movies cont'd



Example: ratings of animation movies cont'd

Movie Title	Mean	SD	N
Batman: Under the Red Hood	5.00		1
Despicable Me	3.72	0.62	9
How to Train Your Dragon	3.41	0.86	11
Legend of the Guardians	4.00		1
Megamind	3.38	1.31	4
Shrek Forever After	4.00	1.32	3
Tangled	4.20	0.89	10
Toy Story 3	3.81	0.96	16

Modeling challenges

- Approach 1 - separate estimates for each movie j :

$$Y_{1j}, \dots, Y_{n_jj} \stackrel{i.i.d.}{\sim} \text{Normal}(\mu_j, \sigma_j).$$

- No relation among groups; groups with small sample size might suffer (e.g. $n_j = 1$)
- Approach 2 - combined estimates for all J movies:

$$Y_{ij} \stackrel{i.i.d.}{\sim} \text{Normal}(\mu, \sigma).$$

- Differences in groups are ignored.
- Something in between - multilevel modeling
 - Pooling information across groups.
 - Achieved through a two-stage prior.

Section 2

A multilevel model with random σ

The sampling model

- Without loss of generality, assume a group-specific normal model for movie j :

$$Y_{ij} \overset{i.i.d.}{\sim} \text{Normal}(\mu_j, \sigma), \quad (1)$$

where $i = 1, \dots, n_j$ and n_j is the number of observations in group j .

- Is a commonly shared σ reasonable? If not, σ can be group-specific.
- Model parameters: $\{\mu_1, \dots, \mu_J, \sigma\}$.

A two-stage prior for $\{\mu_1, \dots, \mu_J\}$: stage 1

- All movies are animation movies, we could assume that the mean ratings are similar across movies
- First stage: the same normal prior distribution for each mean μ_j

$$\mu_j \mid \mu, \tau \sim \text{Normal}(\mu, \tau). \quad (2)$$

- This prior allows information pooled across movies (groups).
 - If τ is large, the μ_j 's are very different a priori \rightarrow modest pooling in parameter estimation.
 - If τ is small, the μ_j 's are very similar a priori \rightarrow large pooling in parameter estimation.
- μ and τ : hyperparameters, and treated random.

A two-stage prior for $\{\mu_1, \dots, \mu_J\}$: stage 2

- Second stage: weakly informative hyperpriors for hyperparameters

$$\mu \sim \text{Normal}(3, 1), \quad (3)$$

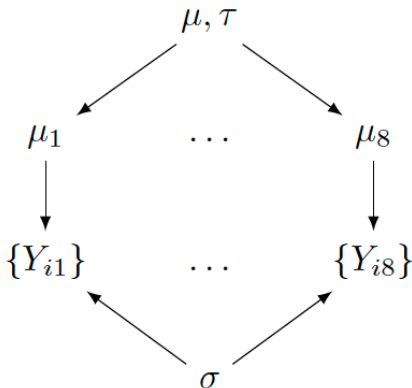
$$\tau \sim \text{Cauchy}(0, 1). \quad (4)$$

- After posterior inference:
 - The posterior of μ is informative about an average mean rating.
 - The posterior of τ is informative about the variation among the μ_j 's.

Prior for σ and graphical representation

- Weakly informative prior for σ :

$$\sigma \sim \text{Cauchy}(0, 1). \quad (5)$$



Section 3

MCMC estimation and diagnostics

Fitting the model

- Use the `brm()` function with `family = gaussian`.
- Use `rating ~ 1 + 1 | Title` expression for model specification.

```
library(brms)
ml_fit <- brm(data = MovieRatings, family = gaussian,
              rating ~ 1 + (1 | Title),
              prior = c(prior(normal(3, 1), class = Intercept),
                        prior(cauchy(0, 1), class = sd),
                        prior(cauchy(0, 1), class = sigma)),
              iter = 20000, warmup = 10000, thin = 10, chains = 2,
              seed = 1234)
```


Saving posterior draws

- Save post as a matrix of simulated posterior draws.
- The model parameters: $\{\mu, \tau, \mu_1, \dots, \mu_8, \sigma\}$

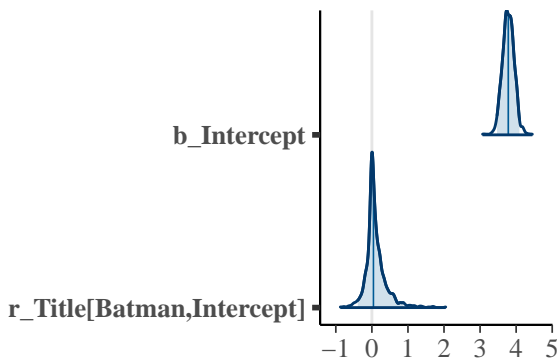
```
post_ml <- posterior_samples(ml_fit)
head(post_ml)
```

```
##      b_Intercept sd_Title__Intercept      sigma r_Title[Batman,Intercept]
## 1      3.942546      0.2808517 0.8786908      0.49813033
## 2      3.705706      0.1590791 0.7761730      0.08335031
## 3      3.796817      0.5044608 0.8182587      0.24449020
## 4      3.839186      0.2093681 0.9983694      0.20169029
## 5      4.024898      0.2159124 1.0201780      0.03042776
## 6      3.748176      0.3035557 0.9314087      -0.09260324
##      r_Title[Despicable.Me,Intercept] r_Title[Dragon,Intercept]
## 1              0.10472786      -0.437729904
## 2             -0.18660562       0.004871502
## 3              0.06337272       0.077854211
## 4             -0.08284498      -0.265128238
## 5             -0.31449212      -0.306096262
## 6             -0.01969473       0.047113685
```

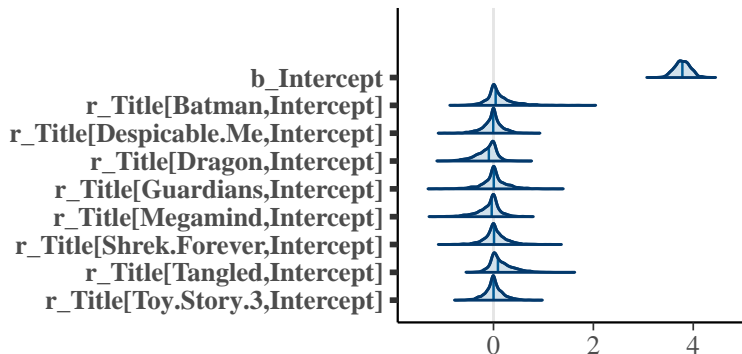
Posterior plots

- Function `mcmc_areas()` displays a density estimate of the simulated posterior draws with a specified credible interval.

```
library(bayesplot)
mcmc_areas(post_ml,
  pars = c("b_Intercept", "r_Title[Batman,Intercept]"),
  prob = 0.95)
```



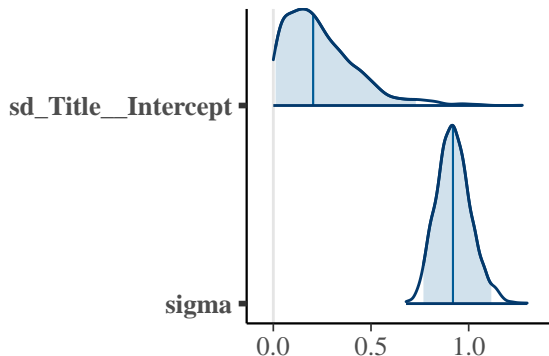
Posterior plots cont'd



Posterior plots cont'd

- Within group variability τ vs between group variability σ .

```
library(bayesplot)
mcmc_areas(post_ml,
  pars = c("sd_Title__Intercept", "sigma"),
  prob = 0.95)
```



MCMC diagnostics: overview

- Theory proves that if a Gibbs sampler iterates enough, the draws will be from the joint posterior distribution (called the target or stationary distribution).
 - Do initial values matter? Should they matter?
 - Markov chain indicate dependence of draws. How to create independent parameter draws?
 - How long do we need to run the MCMC to adequately explore the posterior distribution?
 - How can we tell if the chain is not converging?

MCMC diagnostics: overview cont'd

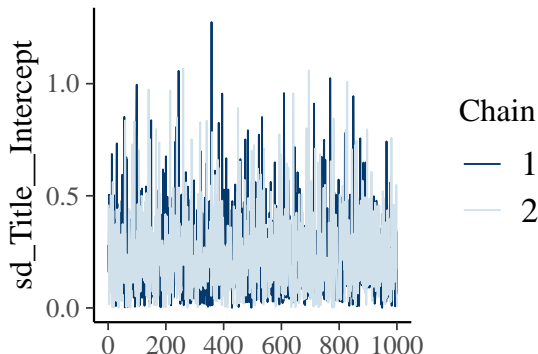
```
ml_fit <- brm(data = MovieRatings, family = gaussian,  
             rating ~ 1 + (1 | Title),  
             prior = c(prior(normal(3, 1), class = Intercept),  
                       prior(cauchy(0, 1), class = sd),  
                       prior(cauchy(0, 1), class = sigma)),  
             iter = 20000, warmup = 10000, thin = 10, chains = 2,  
             seed = 1234)
```

- `iter`: total number of iterations.
- `warmup`: the number of iterations to be discarded (beginning iterations are not converged).
- `thin`: the number of draws to thin for saving.
- `chains`: the number of MCMC chains (some diagnostics can only be done for more than one chain).

MCMC diagnostics: traceplot

- Function `mcmc_trace()` displays a traceplot of the simulated posterior draws for each chain.

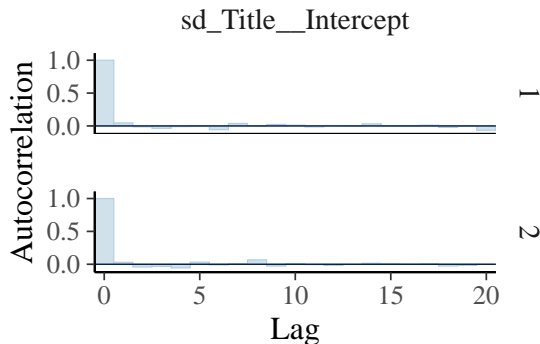
```
mcmc_trace(ml_fit, pars = c("sd_Title__Intercept"))
```



MCMC diagnostics: autocorrelation plot

- Function `mcmc_acf()` displays an autocorrelation plot of the simulated posterior draws.

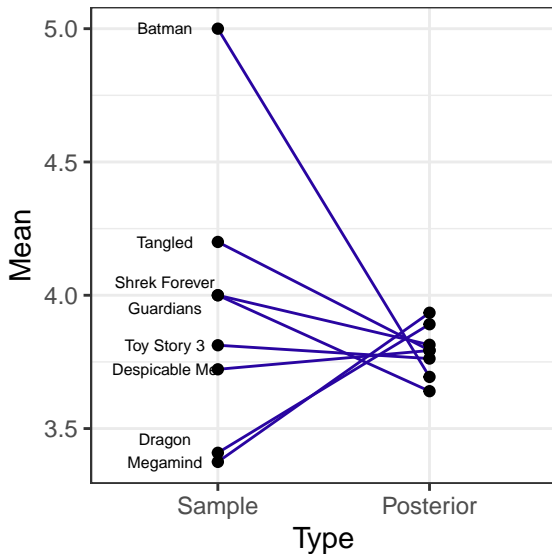
```
mcmc_acf_bar(ml_fit, pars = c("sd_Title__Intercept"))
```



Section 4

Additional Bayesian inferential questions

Shrinkage / pooling effects



Sources of variability

- Two sources of variability in Y_{ij} :

$$\begin{aligned} Y_{ij} &\stackrel{i.i.d.}{\sim} \text{Normal}(\mu, \sigma) \text{ [within-group variability]} \\ \mu_j &\sim \text{Normal}(\mu, \tau) \text{ [between-group variability]} \end{aligned}$$

- To compare these two sources of variability, one can compute the fraction

$$R = \frac{\tau^2}{\tau^2 + \sigma^2}$$

from the posterior draws of τ and σ .

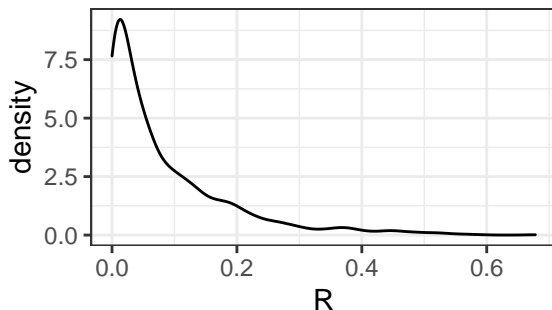
- If $R \rightarrow 1$, the higher the between-group variability.

Sources of variability: results

```
tau_draws <- post_ml[, "sd_Title__Intercept"]  
sigma_draws <- post_ml[, "sigma"]  
R <- tau_draws^2 / (tau_draws^2 + sigma_draws^2)  
quantile(R, c(0.025, 0.975))
```

```
##           2.5%           97.5%  
## 0.0001599452 0.3870263760
```

Density of R



Wrap-up

- Bayesian multilevel modeling:
 - A two-stage prior; interpretation of parameters and hyperparameters.
 - Using the `brms` package; prior choices.
- Additional Bayesian inferences:
 - Shrinkage / pooling effects.
 - Sources of variability.