

Session 3: Bayesian mixed-effect models

Angeline Tsui

1

Learning goals

1. What is a mixed-effect model
2. Running Bayesian mixed-effect models in R (with brms package)
3. Setting priors and running posterior predictive checks

2

Overview: Regression model

Regression: model a relationship between independent variable (IV_i) and dependent variable (DV).

$$DV = \beta_0 + \beta_i IV_i + \epsilon$$

3

Assumptions of a regression:

1. Linearity between the DV and IV_i
2. Residuals are uncorrelated with one another
3. Residuals have an expected value of zero
4. Variance of residuals is constant across IV_i , also known as homoscedasticity
5. No multicollinearity
6. (Optional) Residuals are normally distributed

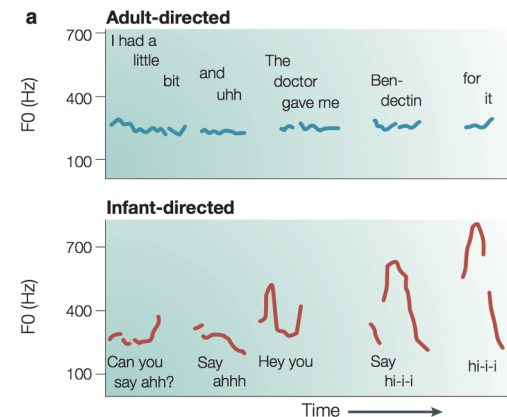
4

Mixed-effect model

- Statistical model that contains both fixed effects and random effects
- Widely used to model repeated measurements
- Fixed effect: population average effect of the IV on the DV
- Random effect: idiosyncratic individual differences between groups (participants)
- We will use the MB1 dataset (large-scale study that examine infant-directed speech)

5

Infant-directed speech

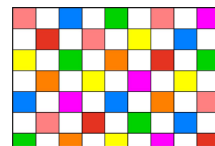


Kuhl (2004) - originally Fernald & Kuhl (1987)

6

Design

- Modeled on Cooper & Aslin (1990)
- DV: Looking time to checkerboard
- IDS vs. ADS played in pairs of trials within subjects
- 16 test trials (8 pairs)
- North American English stimuli



Visual fixation stimulus

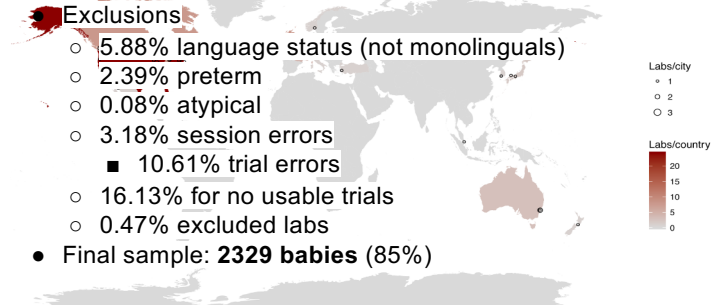


Attention-getter

7

Sample in MB1

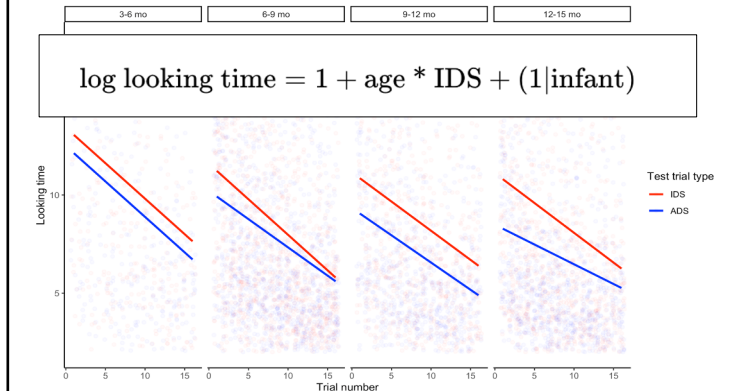
- 67 labs, 17 countries, 4 continents
- Initial sample: **2754 babies**
- Exclusions
 - 5.88% language status (not monolinguals)
 - 2.39% preterm
 - 0.08% atypical
 - 3.18% session errors
 - 10.61% trial errors
 - 16.13% for no usable trials
 - 0.47% excluded labs
- Final sample: **2329 babies (85%)**



8

MB1 dataset

- Let us focus on labs used HPP and tested North American English infants



9

Frequentist: lmer model summary

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	1.853	0.015	731.343	123.733	0
center_age_mo	-0.035	0.005	746.421	-6.950	0
IDS	0.158	0.013	6005.720	11.762	0
center_age_mo:IDS	0.021	0.004	6007.955	4.609	0

(fit using lme4)

10

Bayesian: MCMC to estimate posterior

- Markov Chain Monte Carlo (MCMC)
 - An iterative process that approximates the posterior distribution by randomly sampling parameters
 - Process is a sequential random walk – longer iteration usually leads to better approximation
 - Key: can get statistics like mean and 95% CI by averaging across samples
- When evaluating posterior, we look at whether the model has converged or not

11

Bayesian brms model (default priors)

```

Bayes_model_default <- brm(looking_time ~ center_age_mo * IDS +
  (1 |subid_unique),
  family = lognormal(),
  data = mb1_data,
  # extra settings to make things go faster and to save data
  warmup = 1000,
  iter = 4000,
  cores = 2,
  chains = 2,
  seed = 123,
  save_pars = save_pars(all = T),
  control = list(adapt_delta = 0.95),
  file="mb1_bayes.mod")

```

Always good to use a random seed so you can reproduce your results

Settings to control how long MCMC is run for

Parameter to change details of MCMC ... brms has really good error messages and will usually tell you if you need to put this

12

brms model (default priors) output

```
Family: lognormal
Links: mu = identity; sigma = identity
Formula: looking_time ~ center_age_mo * IDS + (1 | subid_unique)
Data: mb1_data (Number of observations: 6464)
Samples: 2 chains, each with iter = 4000; warmup = 1000; thin = 1;
         total post-warmup samples = 6000
```

Just like an LMER model but with all the benefits of Bayesian models

Group-Level Effects:
~subid_unique (Number of levels: 483)

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sd(Intercept)	0.251	0.011	0.230	0.274	1.001	2030	3146

Population-Level Effects:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	1.853	0.015	1.823	1.882	1.000	2547	3591
center_age_mo	-0.035	0.005	-0.044	-0.025	1.000	2702	3476
IDS	0.158	0.014	0.132	0.184	1.000	8313	4472
center_age_mo:IDS	0.021	0.005	0.012	0.029	1.000	11967	4542

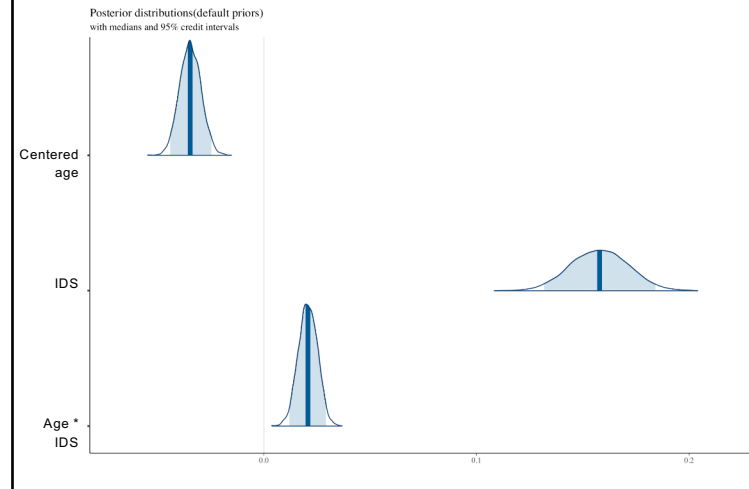
Family Specific Parameters:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sigma	0.539	0.005	0.529	0.548	1.000	7887	4430

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).

13

Posterior distributions



14

Brms model (informative priors) output

```
library(brms)
get_prior(looking_time ~ center_age_mo * IDS, data = mb1_data)
```

	prior	class	coef	group	resp	dpar	nlpar	bound	source
	(flat)	b							default
	(flat)	b	center_age_mo						(vectorized)
	(flat)	b	center_age_mo:IDS						(vectorized)
	(flat)	b	IDS						(vectorized)
student_t(3, 6.6, 4.4)	Intercept								default
student_t(3, 0, 4.4)	sigma								default

```
priors <- c(prior(normal(2, 1), class = Intercept),
            prior(normal(0, 0.5), class = b),
            prior(normal(0, 1), class = sigma))
```

15

brms model (informative priors) output

```
Family: lognormal
Links: mu = identity; sigma = identity
Formula: looking_time ~ center_age_mo * IDS + (1 | subid_unique)
Data: mb1_data (Number of observations: 6464)
Samples: 2 chains, each with iter = 4000; warmup = 1000; thin = 1;
         total post-warmup samples = 6000
```

Group-Level Effects:
~subid_unique (Number of levels: 483)

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sd(Intercept)	0.251	0.011	0.231	0.274	1.000	2533	4190

Population-Level Effects:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	1.854	0.015	1.824	1.882	1.001	2974	3987
center_age_mo	-0.035	0.005	-0.045	-0.025	1.003	2549	3651
IDS	0.158	0.014	0.131	0.184	1.000	11257	4245
center_age_mo:IDS	0.021	0.005	0.012	0.030	1.000	12142	4680

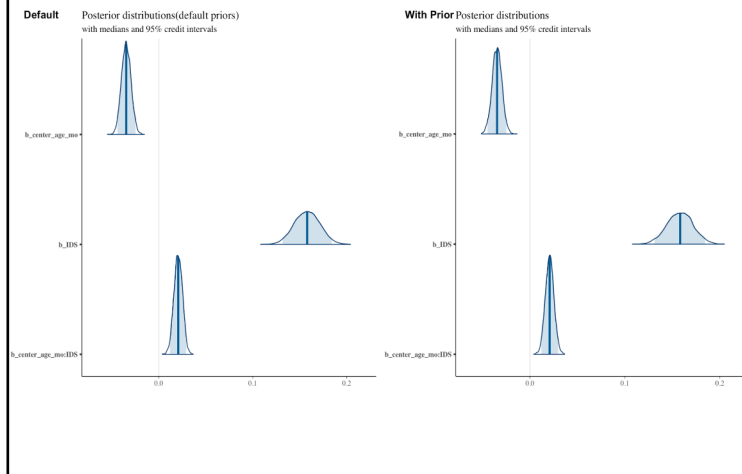
Family Specific Parameters:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sigma	0.539	0.005	0.529	0.548	1.000	10944	4367

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).

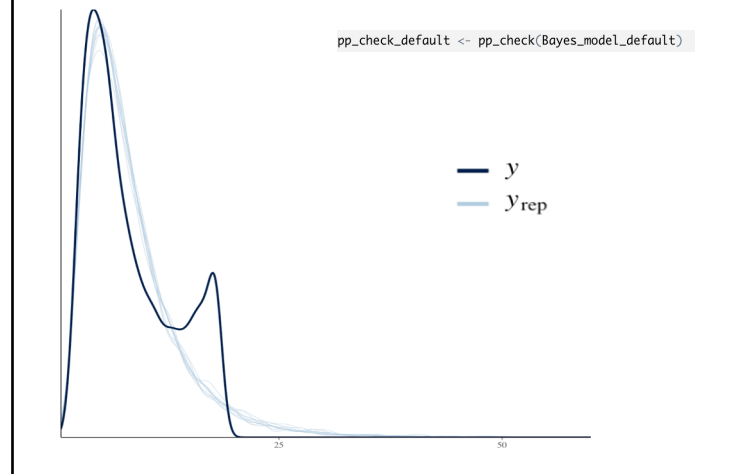
16

Posterior distributions (default vs priors)



17

Posterior predictive check (default)



18

Conclusions

1. What is a mixed-effect model: has both fixed and random effects (e.g., subject effects)
2. Running Bayesian mixed-effect models in R (with brms package): almost as easy as changing "lmer" to "brm"
3. Setting priors and running posterior predictive checks: visualize your model

19