

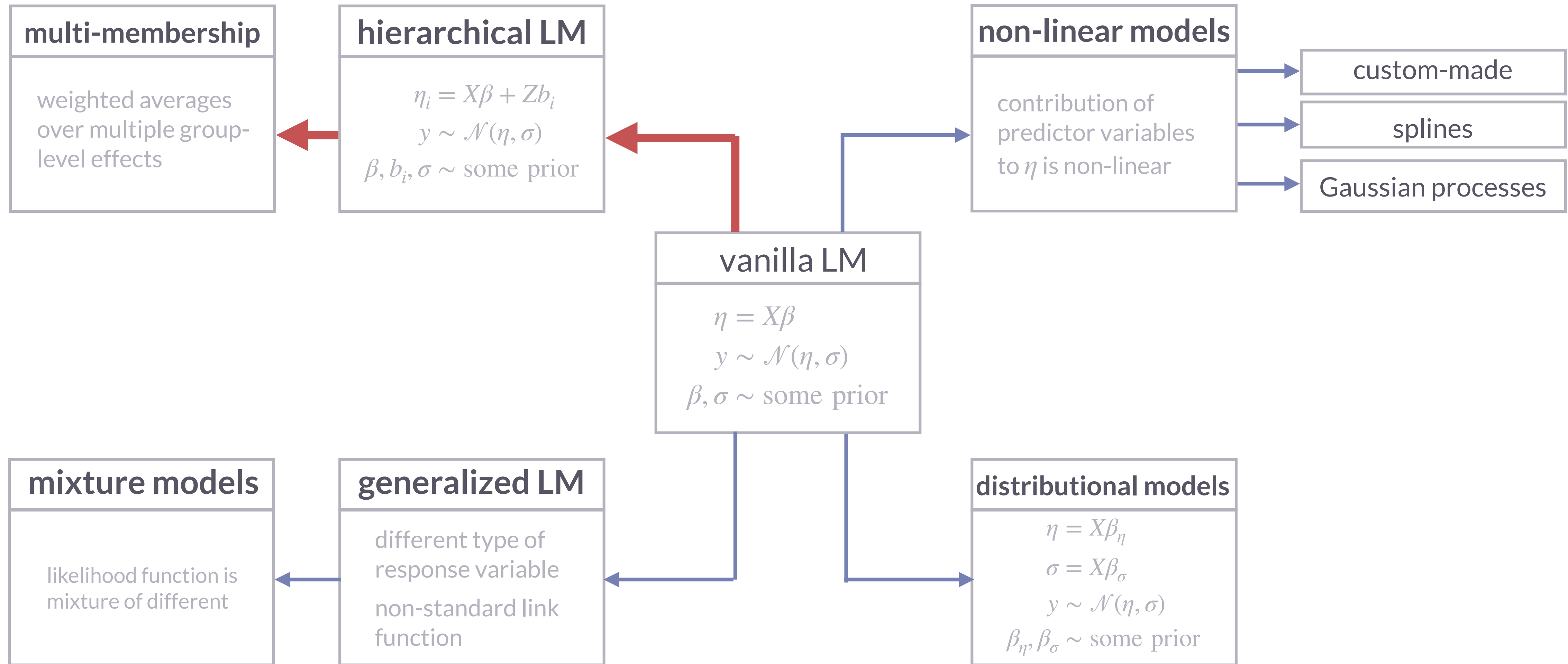
Bayesian data analysis: Theory & practice

Part 3b: Multi-level models

Michael Franke

Roadmap “beyond vanilla”

common extensions of linear regression modeling



Prediction, please!

participant	condition	response
Alex	A	1
Alex	B	3
Alex	C	0
Bo	A	2
Bo	B	5
Bo	C	???

Group-level effects

motivation

Two ways to motivate group-level effects:

1. breaking incorrect independence assumptions
2. “adaptive, regularizing priors”

These are not in competition, just two ways of looking at the same thing.



Case study: MLMs

Case study: processing relative clauses

motivation

Subject relative clause

The senator who interrogated the journalist ...

Object relative clause

The senator who the journalist interrogated ...

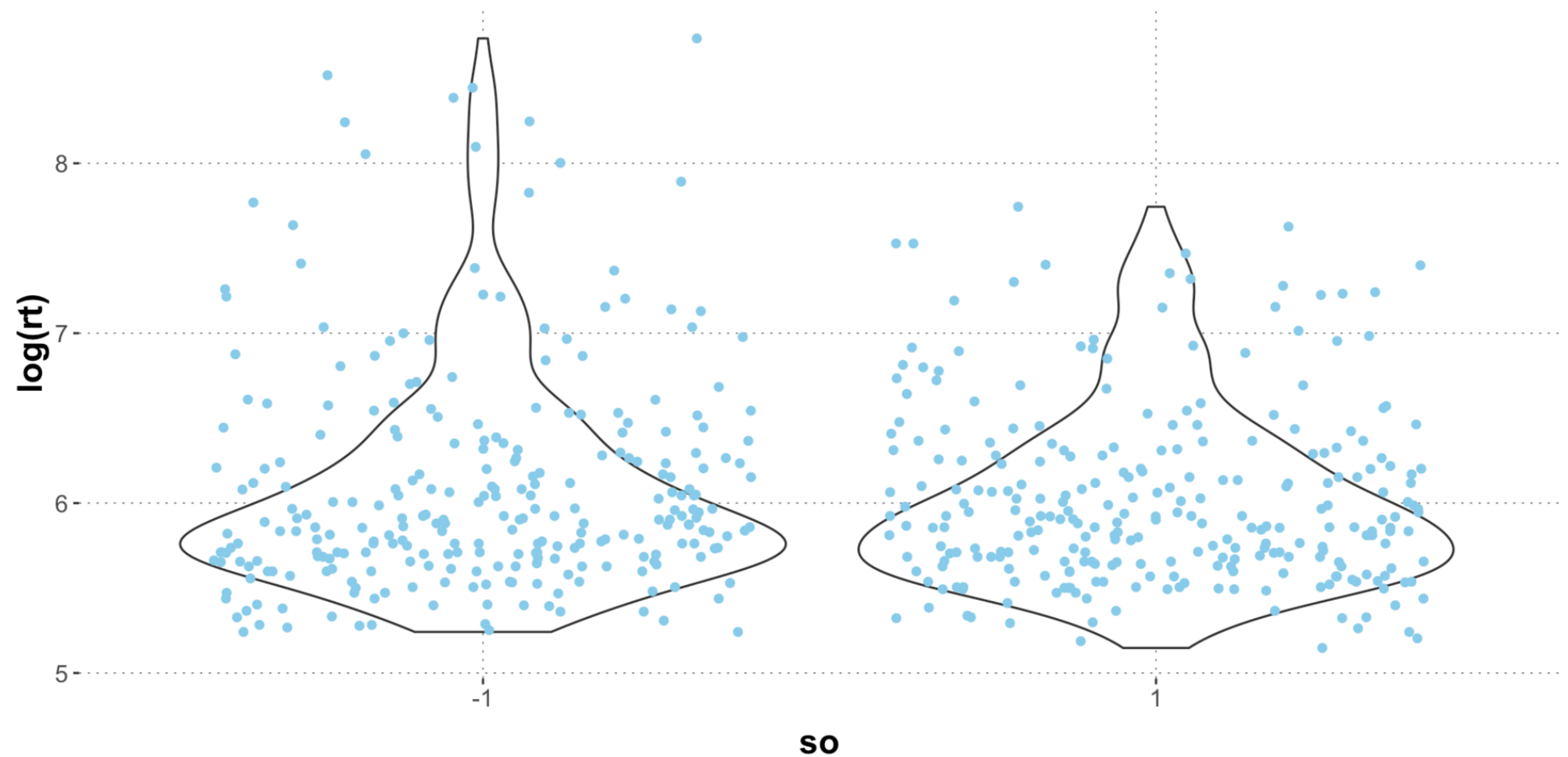
English: SRCs << ORCs 

Chinese: ORCs << SRCs ?

Self-paced reading times

37 subjects read 15 sentences either with an SRC or an ORC

	subj	item	so	rt
	<dbl>	<dbl>	<chr>	<dbl>
1	1	13	1	1561
2	1	6	-1	959
3	1	5	1	582
4	1	9	1	294
5	1	14	-1	438
6	1	4	-1	286
7	1	8	-1	438
8	1	10	-1	278
9	1	2	-1	542
10	1	11	1	494
11	1	7	1	270
12	1	3	1	406
13	1	16	-1	374
14	1	15	1	286
15	1	1	1	246



NB: deviation coding

Fixed-effects model

model specs

- ▶ predict log rt in terms of factor so
- ▶ improper prior on all parameters

$$\log(\text{rt}_i) \sim \mathcal{N}(\eta_i, \sigma_{err})$$
$$\eta_i = \beta_0 + \beta_1 \text{so}_i$$

$$\sigma_{err} \sim \mathcal{U}(0, \infty)$$
$$\beta_0, \beta_1 \sim \mathcal{U}(-\infty, \infty)$$

Fixed-effects model

results

Family: gaussian
Links: mu = identity; sigma = identity
Formula: log(rt) ~ so
Data: rt_data (Number of observations: 547)
Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
total post-warmup draws = 4000

Population-Level Effects:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	6.10	0.04	6.03	6.17	1.00	4030	2692
so1	-0.08	0.05	-0.17	0.02	1.00	4150	2620

Family Specific Parameters:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sigma	0.60	0.02	0.57	0.64	1.00	3898	2821

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).

Varying-intercepts model

model specs

- ▶ predict log rt in terms of factor so
- ▶ improper prior on all parameters
- ▶ different subjects / items could be faster or slower *tout court*

$$\log(\text{rt}_i) \sim \mathcal{N}(\eta_i, \sigma_{err})$$

$$\eta_i = \beta_0 + \underbrace{u_{0,\text{subj}_i} + w_{0,\text{item}_i}}_{\text{varying intercepts}} + \beta_1 \text{so}_i$$

$$u_{0,\text{subj}_i} \sim \mathcal{N}(0, \sigma_{u_0})$$

$$w_{0,\text{subj}_i} \sim \mathcal{N}(0, \sigma_{w_0})$$

Varying-intercepts model

results

Family: gaussian

Links: mu = identity; sigma = identity

Formula: log(rt) ~ (1 | subj + item) + so

Data: rt_data (Number of observations: 547)

Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
total post-warmup draws = 4000

Group-Level Effects:

~item (Number of levels: 15)

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sd(Intercept)	0.20	0.05	0.12	0.32	1.00	1475	2392

~subj (Number of levels: 37)

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sd(Intercept)	0.25	0.04	0.18	0.34	1.00	1593	2503

Population-Level Effects:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	6.10	0.08	5.95	6.24	1.00	1145	1379
so1	-0.07	0.04	-0.16	0.01	1.00	7191	2986

Family Specific Parameters:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sigma	0.52	0.02	0.49	0.55	1.00	5466	3115

Varying-intercepts & varying slopes model

model specs

- ▶ predict log rt in terms of factor so
- ▶ improper prior on all parameters
- ▶ different subjects / items could be faster or slower *tout court*
- ▶ different subjects / items could be more or less sensitive to factor so

$$\log(\text{rt}_i) \sim \mathcal{N}(\eta_i, \sigma_{err})$$

$$\eta_i = \underbrace{\beta_0 + u_{0,\text{subj}_i} + w_{0,\text{item}_i}}_{\text{varying intercepts}} + \underbrace{(\beta_1 + u_{1,\text{subj}_i} + w_{1,\text{item}_i})}_{\text{varying slopes}} \text{so}_i$$

$$u_{0,\text{subj}_i} \sim \mathcal{N}(0, \sigma_{u_0}) \quad w_{0,\text{subj}_i} \sim \mathcal{N}(0, \sigma_{w_0})$$

$$u_{1,\text{subj}_i} \sim \mathcal{N}(0, \sigma_{u_1}) \quad w_{1,\text{subj}_i} \sim \mathcal{N}(0, \sigma_{w_1})$$

Varying-intercepts & varying slopes model

results

Group-Level Effects:

~item (Number of levels: 15)

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sd(Intercept)	0.20	0.05	0.12	0.31	1.00	1514	2007
sd(so1)	0.07	0.05	0.00	0.20	1.00	1653	2508

~subj (Number of levels: 37)

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sd(Intercept)	0.26	0.04	0.19	0.35	1.00	1743	2015
sd(so1)	0.07	0.05	0.00	0.17	1.00	1288	1966

Population-Level Effects:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	6.10	0.07	5.95	6.24	1.00	1565	2375
so1	-0.07	0.05	-0.17	0.03	1.00	4754	2930

Family Specific Parameters:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sigma	0.52	0.02	0.49	0.55	1.00	5597	2604

Varying-intercepts & varying slopes model w/ correlation

model specs

- ▶ predict log rt in terms of factor `so`
- ▶ improper prior on all parameters
- ▶ different subjects / items could be faster or slower *tout court*
- ▶ different subjects / items could be more or less sensitive to factor `so`

$$\eta_i = \beta_0 + u_{0,\text{subj}_i} + w_{0,\text{item}_i} + (\beta_1 + u_{1,\text{subj}_i} + w_{1,\text{item}_i}) \text{so}_i$$
$$\begin{pmatrix} u_{0,\text{subj}_i} \\ u_{1,\text{subj}_i} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_u \right)$$
$$\Sigma_u = \begin{pmatrix} \sigma_{u_0}^2 & \rho_u \sigma_{u_0} \sigma_{u_1} \\ \rho_u \sigma_{u_0} \sigma_{u_1} & \sigma_{u_1}^2 \end{pmatrix} \quad \text{same for item}$$

Varying-intercepts & varying slopes model w/ correlation

results

Group-Level Effects:							
~item (Number of levels: 15)							
	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sd(Intercept)	0.21	0.06	0.12	0.35	1.00	1683	2489
sd(so1)	0.07	0.05	0.00	0.20	1.00	1344	2006
cor(Intercept,so1)	-0.06	0.53	-0.94	0.92	1.00	4241	2651
~subj (Number of levels: 37)							
	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sd(Intercept)	0.30	0.05	0.21	0.41	1.00	1435	2261
sd(so1)	0.13	0.07	0.01	0.26	1.00	1066	1030
cor(Intercept,so1)	-0.67	0.33	-0.99	0.28	1.00	2084	1626
Population-Level Effects:							
	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	6.10	0.08	5.94	6.25	1.00	1484	2214
so1	-0.07	0.06	-0.18	0.04	1.00	4210	2934
Family Specific Parameters:							
	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sigma	0.51	0.02	0.48	0.55	1.00	4003	3093



**How to choose
RE-structure?**

How to choose multi-level architecture?

two approaches

(1) keep it maximal

- include maximum RE structure that makes sense
 - what “makes sense” may depend on *a priori* considerations
 - data might not be sufficient to estimate otherwise conceivable MLMs

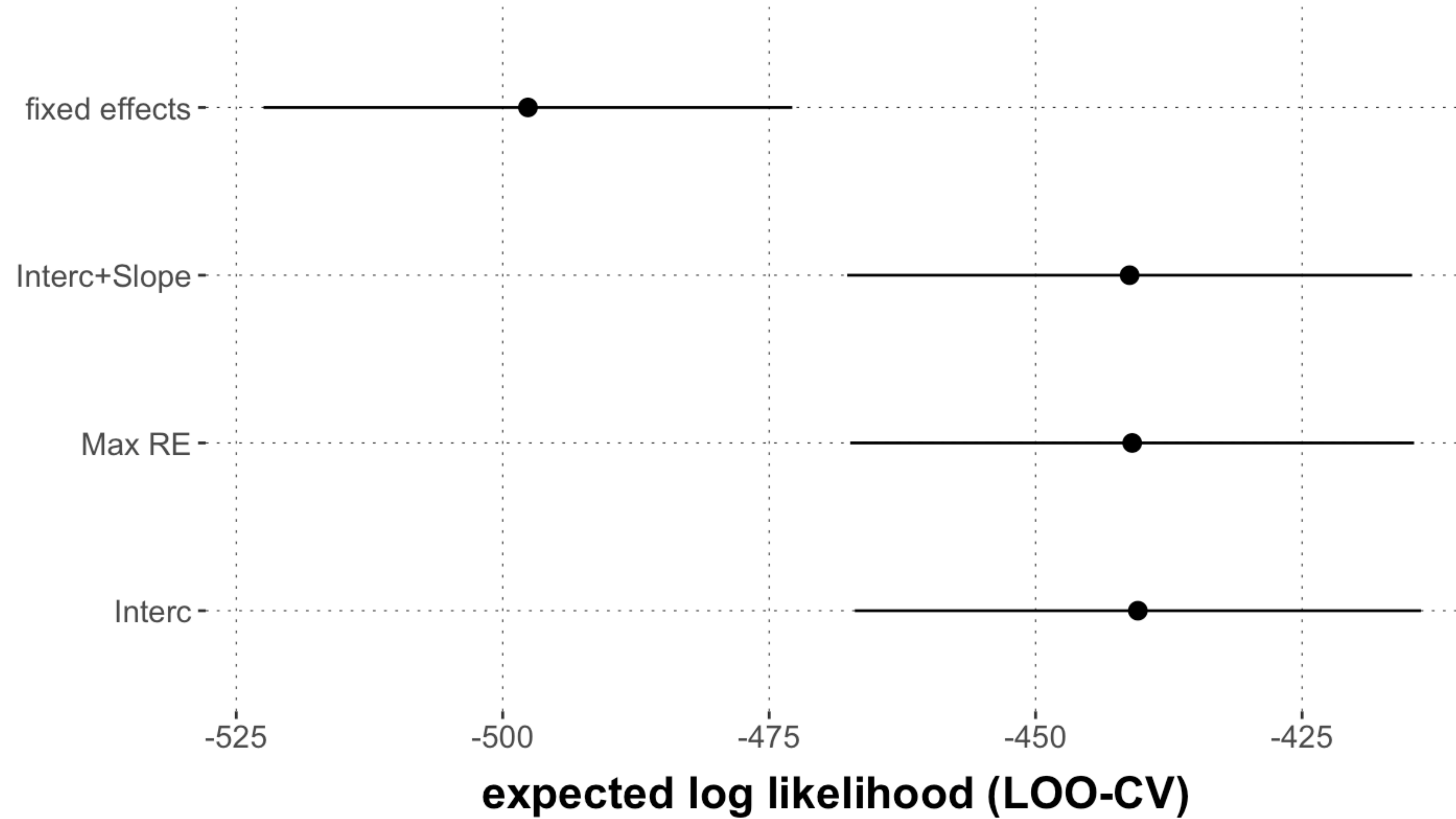
(2) let the data decide

- compare different MLMs and choose the best

- (1) is more careful / prudent in a science context (learning about the world from the model and the data)
- (2) may be more adequate for an engineering context (predicting well enough with efficient models)

LOO-based model comparison

results






priors for MLMs

Priors in MLMs

```
# define the model as a "brmsformula" object
myFormula <- brms::bf(RT ~ 1 + condition + (1 + condition | submission_id))

# get prior information
brms::get_prior(
  formula = myFormula,
  data     = data_MC,
  family   = gaussian()
)
```

	prior	class	coef	group	resp	dpar	nlpar	lb	ub	source
	(flat)	b								default
	(flat)	b	conditiondiscrimination							(vectorized)
	(flat)	b	conditionreaction							(vectorized)
	lkj(1)	cor								default
	lkj(1)	cor		submission_id						(vectorized)
student_t(3, 385, 133.4)		Intercept								default
student_t(3, 0, 133.4)		sd						0		default
student_t(3, 0, 133.4)		sd		submission_id				0		(vectorized)
student_t(3, 0, 133.4)		sd	conditiondiscrimination	submission_id				0		(vectorized)
student_t(3, 0, 133.4)		sd	conditionreaction	submission_id				0		(vectorized)
student_t(3, 0, 133.4)		sd	Intercept	submission_id				0		(vectorized)
student_t(3, 0, 133.4)		sigma						0		default



MLM notation & formula syntax

Multi-level models

notation

Cumbersome: $\eta_i = \beta_0 + u_{0,\text{subj}_i} + w_{0,\text{item}_i} + (\beta_1 + u_{1,\text{subj}_i} + w_{1,\text{item}_i}) \text{so}_i$

Compact: $\eta = X\beta + Z\gamma$

$$X = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \\ \vdots & \vdots \end{pmatrix} \quad Z = \begin{pmatrix} 0 & \dots & 0 & 0 & \dots & 1 & \dots \\ 0 & \dots & 0 & 1 & \dots & 0 & \dots \\ 0 & \dots & 1 & 0 & \dots & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

random effects matrix

	subj	item	so	rt
	<dbl>	<dbl>	<chr>	<dbl>
1	1	13	1	1561
2	1	6	-1	959
3	1	5	1	582
4	1	9	1	294
5	1	14	-1	438
6	1	4	-1	286
7	1	8	-1	438
8	1	10	-1	278
9	1	2	-1	542
10	1	11	1	494

BRMS formula syntax for MLMs


in a nutshell

basic form:

$\text{response} \sim \text{pterms} + (\text{gterms} \mid \text{group})$

variations:

- ▶ $(\text{gterms} \parallel \text{group})$: suppress correlation between gterms
- ▶ $(\text{gterms} \mid \text{g1} + \text{g2})$: syntactic sugar for $(\text{gterms} \mid \text{g1}) + (\text{gterms} \mid \text{g2})$
- ▶ $(\text{gterms} \mid \text{g1} : \text{g2})$: all combinations of g1 and g2 (Cartesian product)
- ▶ $(\text{gterms} \mid \text{g1} / \text{g2})$: nesting g2 within g1; equals $(\text{gterms} \mid \text{g1}) + (\text{gterms} \mid \text{g1} : \text{g2})$
- ▶ $(\text{gterms} \mid \text{IDx} \mid \text{group})$: correlation for all group-level categories with IDx
 - useful for multi-formula models (e.g., non-linear models)



**multi-membership &
an alternative motivation
for MLMs**

demo



walkthrough for multi-membership models & an alternative view on group-level effects

