Multilevel Bayesian Models

Eric Guntermann
Université de Montréal

February 20th, 2014

Centre for the Study of Democratic Citizenship
Centre pour l'étude de la citoyenneté démocratique

What you need

- R
- RStudio
- JAGS
- R code file
- R2jags, coda, R2WinBUGS, lattice, and rjags (R packages)
- Datasets
- You can find all of this at:
  `http://ericguntermann.com/multilevel.html`

What we will learn

- Why use multilevel (hierarchical) models?
- Fitting frequentist multilevel models
- Introduction to Bayesian analysis
- Bayesian two-level models
- Bayesian three-level model

Let's start with an example

- Van der Eijk (2006) "Rethinking the dependent variable in voting behavior: On the measurement and analysis of electoral utilities". *Electoral Studies* 25 (2006)
- Propose studying voting behaviour with a continuous dependent variable
- Use complicated conditional logit model, which is hard to interpret and does not fully account for uncertainty
- We have like-dislike ratings of political parties from the Comparative Study of Electoral systems (CSES) module 3
- Data on 66 parties running in 11 elections

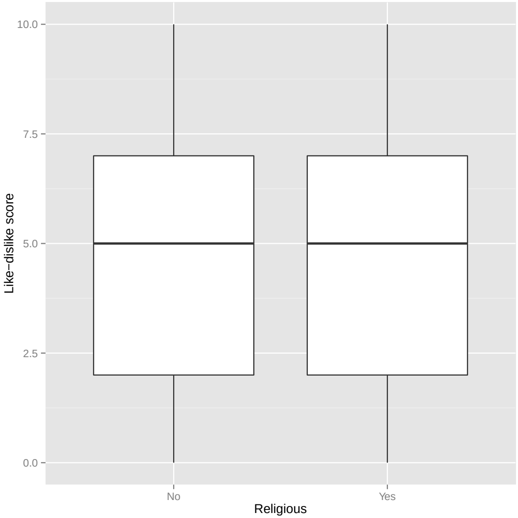
Like-dislike question

"I'd like to know what you think about each of our political parties. After I read the name of a political party, please rate it on a scale from 0 to 10, where 0 means you strongly dislike that party and 10 means that you strongly like that party. If I come to a party you haven't heard of or you feel you do not know enough about, just say so."

What's wrong with this approach?

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.69	0.10	56.90	0.00
log(age)	-0.35	0.03	-13.45	0.00
femaleFemale	0.09	0.02	4.44	0.00
incomemedium	0.11	0.02	4.76	0.00
incomehigh	0.16	0.02	6.48	0.00
religiousYes	0.07	0.02	3.30	0.00

Example: Explaining Party Like-Dislike Scores



What's wrong with this approach?

- We make a lot of assumptions.
- Notably:
  - The relationship between sex and like-dislike (ld) scores is the same for all parties
  - Average ld scores are the same for all parties
  - The relationship between sex and ld scores is the same in all elections
  - Average ld scores are the same in all elections

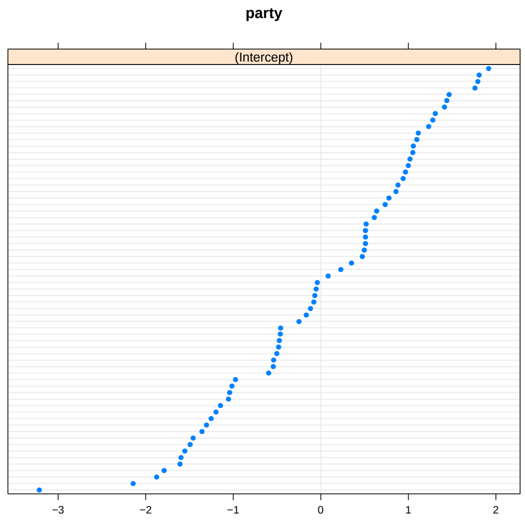
Multilevel data strucures are common in the social sciences

- Students in schools
- Survey responses for different parties (in different elections and even countries)
- Test results at different ages (panel data)
- ...

What happens if we don't take into consideration the multiple levels (i.e. if we pool data)?

- We might find no relationship (or a weak effect) when there actually is one in some groups
- We might find a relationship at the individual-level that is actually a group-level effect.
- At the very least, our estimates would be biased

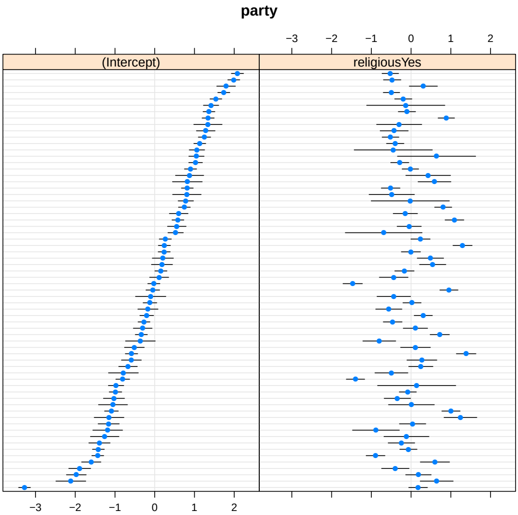
Solution 1: Varying Intercepts (unmodelled)



Solution 2: Varying Intercepts (modelled)

	Estimate	Std. Error	t value
(Intercept)	5.60	0.21	27.03
log(age)	-0.35	0.02	-14.91
femaleFemale	0.11	0.02	5.94
incomemedium	0.12	0.02	5.54
incomehigh	0.17	0.02	7.56
religiousYes	0.16	0.02	8.19
lr	-0.16	0.29	-0.56

Solution 3: Varying intercepts and slopes (unmodelled)



Solution 4: Varying Intercepts and slopes (modelled)

	Estimate	Std. Error	t value
(Intercept)	5.72	0.20	28.14
log(age)	-0.36	0.02	-15.18
femaleFemale	0.10	0.02	5.76
incomemedium	0.11	0.02	5.29
incomehigh	0.16	0.02	7.30
religiousYes	-0.03	0.10	-0.32
lr	-0.42	0.29	-1.48
religiousYes:lr	0.45	0.16	2.76

Other Common Terminology (which can confuse people)

- Fixed Effects: coefficients that don't vary, mean of coefficient across groups, or separate unmodelled intercepts for each group ("within effects")
- Random Effects: varying coefficients or variation of coefficients from overall mean
- Bayesians just forget about all this. All parameters are random!

Introduction to Bayesian Analysis

What do we know before we run a classical frequentist model?

Introduction to Bayesian Analysis

- Frequentists pretend they know nothing
- Bayesians sometimes assert we know nothing
- But we at least are clear about that
- Bayesian analysis is all about combining prior information with information from our data
- We combine a prior (what we know at the beginning) and a likelihood (what we learn) to come up with a posterior

Introduction to Bayesian Analysis

$$Posterior \propto Prior * Likelihood$$

Introduction to Bayesian Analysis

$$P(\theta|X) \propto p(\theta)L(\theta|X)$$

Key differences

- Frequentists
  - Assume data are a random sample from a larger population.
  - Parameters exist in the population.
  - All about replication: what would happen if we took a large number of samples
  - Example: confidence intervals are one of large number of possible intervals. They have no meaning independent of infinite replication.
  - Null-hypothesis significance tests (NHST): significant results are those that are unlikely in imagined sampling distribution.
- Bayesians
  - Bayesians assume data are fixed, while parameters are random.
  - Bayesians describe posterior distributions.
  - Credible intervals: intervals that contain parameter with x % probability

Where does a prior come from?

- Could be prior knowledge (e.g. we know that a coefficient is in a particular range, is positive, etc.)
- Could be no knowledge, then we use a non-informative prior
- Could be knowledge from upper level (multilevel)

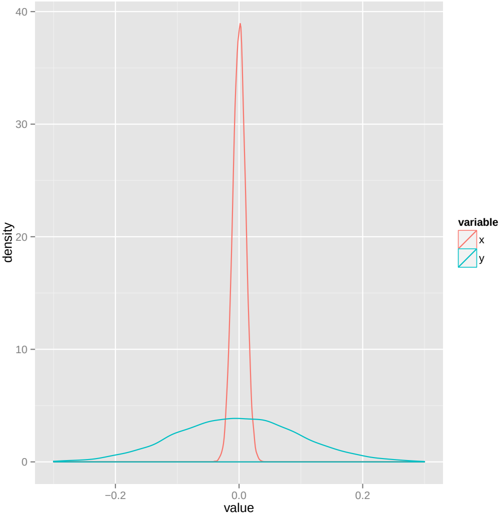
How do we get to a posterior?

- Problem: often posterior calculations are hard or impossible to calculate
- MCMC (Markov Chain Monte Carlo) sampling
- Eventually after sampling for a while, Markov chain reaches posterior
- In other words, our model has converged
- Once our model has converged, our samples are from the posterior distribution
- We summarize results with distributions
- Not reliant on asymptotic properties

Where do posterior estimates come from?

- It all depends on how much information is in the prior vs the data
- Prior variance vs. data variance (and N)

How informative are our priors?



x = Normal(0,0.01) y= Normal(0,1)

How do Bayesian methods help us with multilvel models?

- We use group-level knowledge as priors!
- In a single level model: a prior for a coefficient could be $\beta \sim \mathcal{N}(0, 10000)$
- In a multilevel model, we use a second level as a prior for the first level coefficients
  - $\beta_i \sim \mathcal{N}(mu.b, var.beta)$
  - $mu.b \sim \mathcal{N}(0, 10000)$ (unmodelled)
- If we model the first-level coefficients:
  - $\beta_j \sim \mathcal{N}(mu.b_j, var.beta)$
  - $mu.b_j = \beta_0 + \beta_u * uvar_j$

Where do posterior estimates come from in multilevel models?

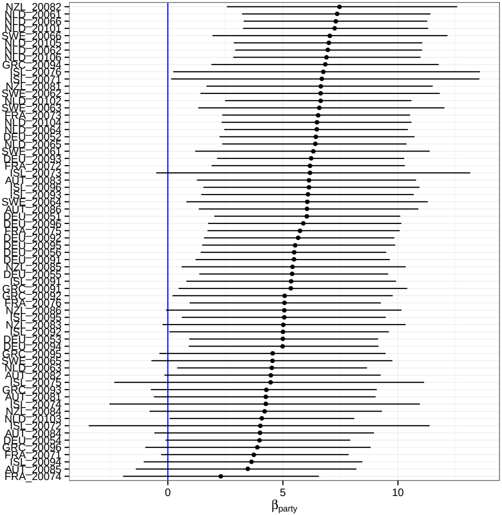
- Since the prior for level-1 parameters comes from a higher level, the variance at that higher level matters.
- The less variance at a higher level (the greater precision), the more the higher level matters for determining a coefficient's posterior
- The less variance at the lower level and also the greater the N at the lower level, the more data in the particular group matters for determining a coefficient's posterior

Single-level Bayesian Model

Table: Model 1

	mean	sd	2.5%	97.5%	Rhat
b.age	-0.35	0.03	-0.40	-0.30	1.00
b.female	0.08	0.02	0.04	0.12	1.00
b.income2	0.50	31.39	-64.04	63.82	1.01
b.income3	0.10	0.02	0.06	0.14	1.01
b.religious	0.07	0.02	0.03	0.10	1.01
deviance	411505.22	3.38	411500.60	411513.74	1.00

Bayesian Unmodelled Varying Intercepts Model

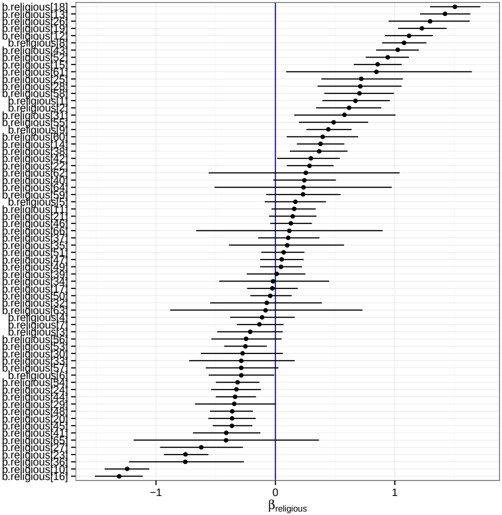


Bayesian modelled Varying Intercepts Model

Table: Model 3

	mean	sd	2.5%	97.5%	Rhat
b.age	-0.35	0.02	-0.40	-0.31	1.00
b.female	0.11	0.02	0.07	0.14	1.00
b.income2	0.12	0.02	0.08	0.16	1.00
b.income3	0.17	0.02	0.12	0.21	1.00
b.lr	0.37	0.30	-0.21	0.97	1.00
b.religious	0.16	0.02	0.12	0.19	1.01
deviance	395657.07	12.15	395634.21	395682.96	1.00

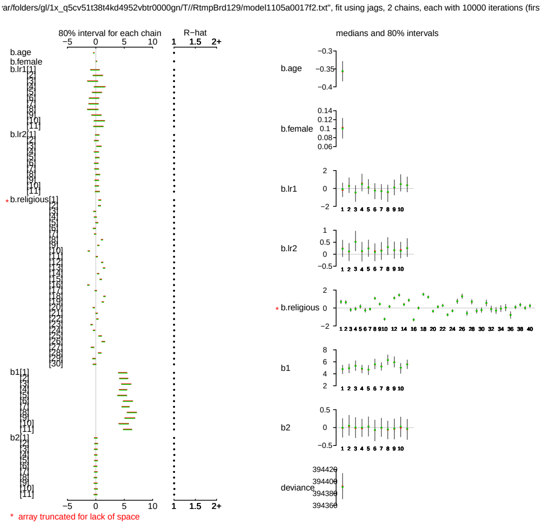
Bayesian Unmodelled Varying Intercepts and Slopes Model



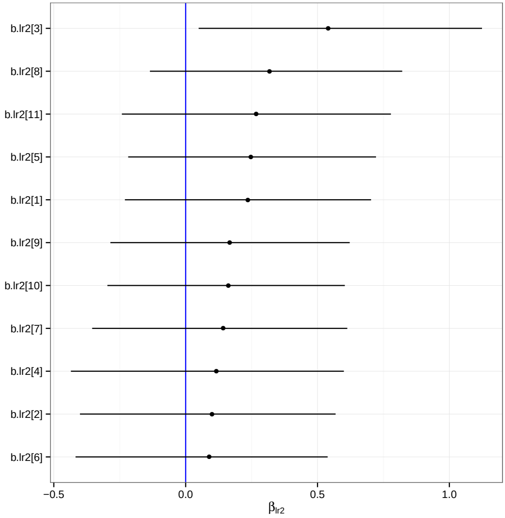
Bayesian modelled Varying Intercepts and Slopes Model

	mean	sd	2.5%	97.5%	Rhat
b.age	-0.36	0.02	-0.40	-0.31	1.01
b.income2	0.11	0.02	0.07	0.15	1.02
b.income3	0.16	0.02	0.12	0.20	1.01
b.lr1	0.29	0.31	-0.31	0.92	1.00
b.lr2	0.20	0.18	-0.16	0.55	1.00
b1	5.37	0.27	4.86	5.90	1.01
b2	0.00	0.14	-0.27	0.29	1.00
deviance	394388.21	16.16	394359.26	394421.67	1.00

Three Level Model: First level Intercepts and Slopes are Modelled. Second level coefficients are unmodelled



Three-Level Model: First-level intercepts and slopes are modelled. Second-level coefficients are unmodelled



Any questions?

Feel free to email me: eric.guntermann@umontreal.ca

To learn more about Multilevel Models and Bayesian Analysis

- Gelman and Hill (2006) *Data Analysis Using Regression and Multilevel/Hierarchical Models*
- Gelman et al. (2013) *Bayesian Data Analysis*
- Gill (2014) *Bayesian Methods: A Social and Behavioral Sciences Approach*
- Jackman (2009) *Bayesian Analysis for the Social Sciences*