Bayesian Data Analysis, class 5a

Andrew Gelman

Chapter 5: Hierarchical models (part 2)

Discussion of homework due beginning of Class 4b

- Theory problem
- Computing problem
- Applied problem

Discussion of homework due beginning of Class 4b

- Theory problem
- Computing problem
- Applied problem

Discussion of homework due beginning of Class 4b

- Theory problem
- Computing problem
- Applied problem

Discussion of homework due beginning of Class 4b

- Theory problem
- Computing problem
- Applied problem

Theory problem

- Normal approximation to the posterior distribution from Cauchy data
- Second derivative, plotting the normal density

Theory problem

- Normal approximation to the posterior distribution from Cauchy data
- Second derivative, plotting the normal density

Theory problem

- Normal approximation to the posterior distribution from Cauchy data
- Second derivative, plotting the normal density

Computing problem

- Poisson regression: check that posterior inferences are consistent with true parameter values

Computing problem

- Poisson regression: check that posterior inferences are consistent with true parameter values

Applied problem

- Basketball shooting again: θ_i is improvement in success probability for person i
- Prior distribution for mean and standard deviation of θ_i in the population
- Sidestepping causal questions

Applied problem

- Basketball shooting again: θ_i is improvement in success probability for person i
- Prior distribution for mean and standard deviation of θ_i in the population
- Sidestepping causal questions

Applied problem

- Basketball shooting again: θ_i is improvement in success probability for person i
- Prior distribution for mean and standard deviation of θ_i in the population
- Sidestepping causal questions

Applied problem

- Basketball shooting again: θ_i is improvement in success probability for person i
- Prior distribution for mean and standard deviation of θ_i in the population
- Sidestepping causal questions

5. Hierarchical models (part 2)

- The 8 schools example
- Meta-analysis
- Weakly informative priors for hierarchical variance parameters

- The 8 schools example
- Meta-analysis
- Weakly informative priors for hierarchical variance parameters

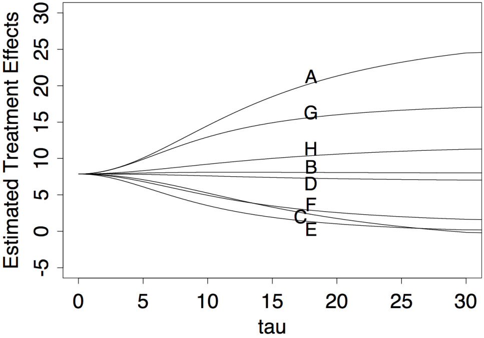
5. Hierarchical models (part 2)

- The 8 schools example
- Meta-analysis
- Weakly informative priors for hierarchical variance parameters

- The 8 schools example
- Meta-analysis
- Weakly informative priors for hierarchical variance parameters

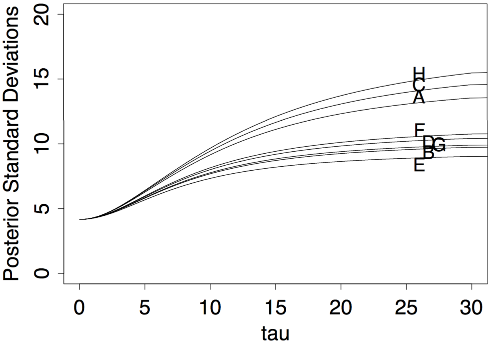
Estimates conditional on the group-level variance

School	Estimated treatment effect, y_j	Standard error of effect estimate, σ_j
A	28	15
B	8	10
C	-3	16
D	7	11
E	-1	9
F	1	11
G	18	10
H	12	18



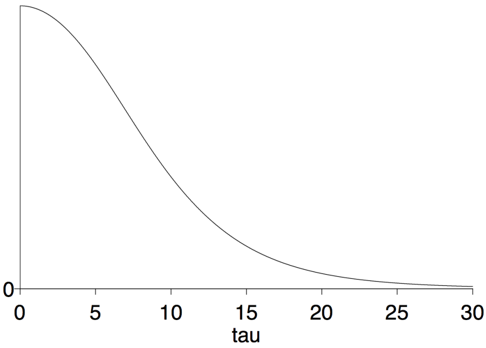
Posterior uncertainties

School	Estimated treatment effect, y_j	Standard error of effect estimate, σ_j
A	28	15
B	8	10
C	-3	16
D	7	11
E	-1	9
F	1	11
G	18	10
H	12	18

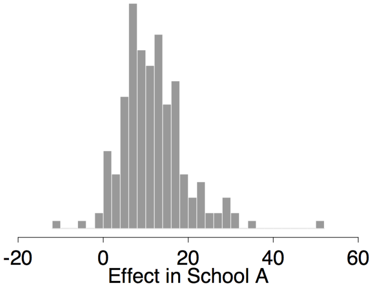


Inference for the group-level variance

School	Estimated treatment effect, y_j	Standard error of effect estimate, σ_j
A	28	15
B	8	10
C	-3	16
D	7	11
E	-1	9
F	1	11
G	18	10
H	12	18



A couple of qois



How do these differ?

Some lingering questions

- What if the model were applied not to 8 schools but to 8 unrelated objects?
- Inference for μ

Some lingering questions

- What if the model were applied not to 8 schools but to 8 unrelated objects?
- Inference for μ

Some lingering questions

- What if the model were applied not to 8 schools but to 8 unrelated objects?
- Inference for μ

5.6. Hierarchical modeling applied to a meta-analysis

- Transform $\frac{y_2}{n_2} - \frac{y_1}{n_1}$ to log-odds with approximately normal errors
- Apply 8-schools model
- Three levels of inference:
- Write the notation for each inference

5.6. Hierarchical modeling applied to a meta-analysis

- Transform $\frac{y_2}{n_2} - \frac{y_1}{n_1}$ to log-odds with approximately normal errors
- Apply 8-schools model
- Three levels of inference:
  - Average effect
  - Effect in a single study (existing or new)
  - Prediction for a new person (in an existing or new study)
- Write the notation for each inference

5.6. Hierarchical modeling applied to a meta-analysis

- Transform $\frac{y_2}{n_2} - \frac{y_1}{n_1}$ to log-odds with approximately normal errors
- Apply 8-schools model
- Three levels of inference:
  - Average effect
  - Effect in a single study (existing or new)
  - Prediction for a new person (in an existing or new study)
- Write the notation for each inference

5.6. Hierarchical modeling applied to a meta-analysis

- Transform $\frac{y_2}{n_2} - \frac{y_1}{n_1}$ to log-odds with approximately normal errors
- Apply 8-schools model
- Three levels of inference:
  - Average effect
  - Effect in a single study (existing or new)
  - Prediction for a new person (in an existing or new study)
- Write the notation for each inference

5.6. Hierarchical modeling applied to a meta-analysis

- Transform $\frac{y_2}{n_2} - \frac{y_1}{n_1}$ to log-odds with approximately normal errors
- Apply 8-schools model
- Three levels of inference:
  - Average effect
  - Effect in a single study (existing or new)
  - Prediction for a new person (in an existing or new study)
- Write the notation for each inference

5.6. Hierarchical modeling applied to a meta-analysis

- Transform $\frac{y_2}{n_2} - \frac{y_1}{n_1}$ to log-odds with approximately normal errors
- Apply 8-schools model
- Three levels of inference:
  - Average effect
  - Effect in a single study (existing or new)
  - Prediction for a new person (in an existing or new study)
- Write the notation for each inference

5.6. Hierarchical modeling applied to a meta-analysis

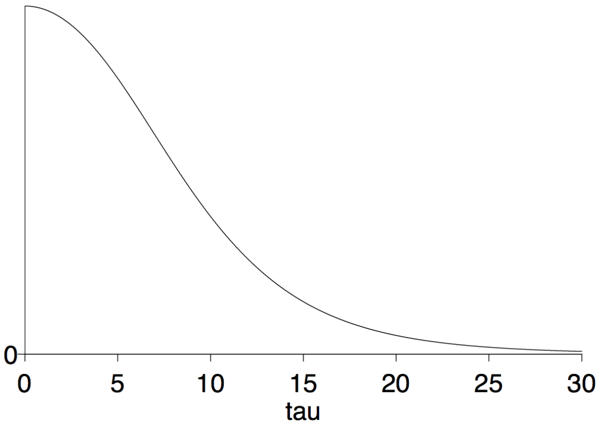
- Transform $\frac{y_2}{n_2} - \frac{y_1}{n_1}$ to log-odds with approximately normal errors
- Apply 8-schools model
- Three levels of inference:
  - Average effect
  - Effect in a single study (existing or new)
  - Prediction for a new person (in an existing or new study)
- Write the notation for each inference

5.6. Hierarchical modeling applied to a meta-analysis

- Transform $\frac{y_2}{n_2} - \frac{y_1}{n_1}$ to log-odds with approximately normal errors
- Apply 8-schools model
- Three levels of inference:
  - Average effect
  - Effect in a single study (existing or new)
  - Prediction for a new person (in an existing or new study)
- Write the notation for each inference

5.7. Weakly informative priors for hierarchical variance parameters

Result of the uniform prior distribution, $p(\tau) \propto 1$, for $\tau \in (0, \infty)$:



Prior distributions for the hierarchical variance parameter

- What is a good "weakly informative prior"?
  - $\log \tau \sim \text{Uniform}(-\infty, \infty)$
  - $\tau \sim \text{Uniform}(0, \infty)$
  - $\tau \sim \text{Inverse-gamma}(0.001, 0.001)$
  - $\tau \sim \text{Cauchy}^{+}(0, A)$
- Polson and Scott (2011):
  "The half-Cauchy occupies a sensible 'middle ground' ... it performs very well near the origin, but does not lead to drastic compromises in other parts of the parameter space."

Prior distributions for the hierarchical variance parameter

- What is a good "weakly informative prior"?
  - $\log \tau \sim \text{Uniform}(-\infty, \infty)$
  - $\tau \sim \text{Uniform}(0, \infty)$
  - $\tau \sim \text{Inverse-gamma}(0.001, 0.001)$
  - $\tau \sim \text{Cauchy}^{+}(0, A)$
- Polson and Scott (2011):
  "The half-Cauchy occupies a sensible 'middle ground' . . . it performs very well near the origin, but does not lead to drastic compromises in other parts of the parameter space."

Prior distributions for the hierarchical variance parameter

- What is a good "weakly informative prior"?
  - $\log \tau \sim \text{Uniform}(-\infty, \infty)$
  - $\tau \sim \text{Uniform}(0, \infty)$
  - $\tau \sim \text{Inverse-gamma}(0.001, 0.001)$
  - $\tau \sim \text{Cauchy}^{+}(0, A)$
- Polson and Scott (2011):
  "The half-Cauchy occupies a sensible 'middle ground' . . . it performs very well near the origin, but does not lead to drastic compromises in other parts of the parameter space."

Prior distributions for the hierarchical variance parameter

- What is a good "weakly informative prior"?
  - $\log \tau \sim \text{Uniform}(-\infty, \infty)$
  - $\tau \sim \text{Uniform}(0, \infty)$
  - $\tau \sim \text{Inverse-gamma}(0.001, 0.001)$
  - $\tau \sim \text{Cauchy}^{+}(0, A)$
- Polson and Scott (2011):
  "The half-Cauchy occupies a sensible 'middle ground' ... it performs very well near the origin, but does not lead to drastic compromises in other parts of the parameter space."

Prior distributions for the hierarchical variance parameter

- What is a good "weakly informative prior"?
  - $\log \tau \sim \text{Uniform}(-\infty, \infty)$
  - $\tau \sim \text{Uniform}(0, \infty)$
  - $\tau \sim \text{Inverse-gamma}(0.001, 0.001)$
  - $\tau \sim \text{Cauchy}^+(0, A)$
- Polson and Scott (2011):
  "The half-Cauchy occupies a sensible 'middle ground' . . . it performs very well near the origin, but does not lead to drastic compromises in other parts of the parameter space."

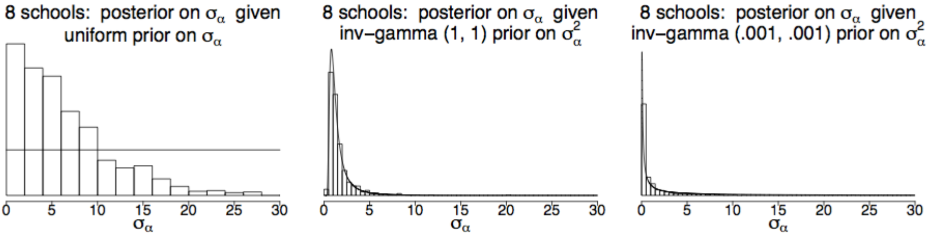
Prior distributions for the hierarchical variance parameter

- What is a good "weakly informative prior"?
  - $\log \tau \sim \text{Uniform}(-\infty, \infty)$
  - $\tau \sim \text{Uniform}(0, \infty)$
  - $\tau \sim \text{Inverse-gamma}(0.001, 0.001)$
  - $\tau \sim \text{Cauchy}^{+}(0, A)$
- Polson and Scott (2011):
  "The half-Cauchy occupies a sensible 'middle ground' ... it performs very well near the origin, but does not lead to drastic compromises in other parts of the parameter space."

Prior distributions for the hierarchical variance parameter

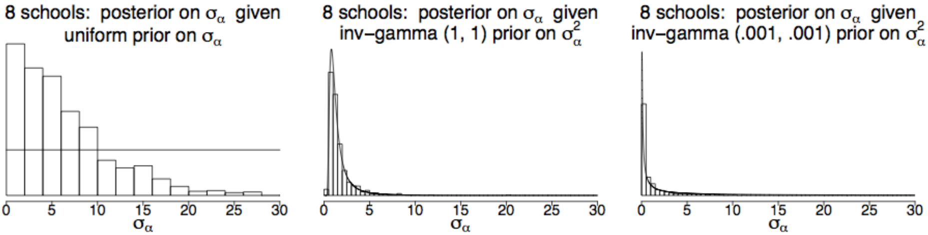
- What is a good "weakly informative prior"?
  - $\log \tau \sim \text{Uniform}(-\infty, \infty)$
  - $\tau \sim \text{Uniform}(0, \infty)$
  - $\tau \sim \text{Inverse-gamma}(0.001, 0.001)$
  - $\tau \sim \text{Cauchy}^{+}(0, A)$
- Polson and Scott (2011):
  "The half-Cauchy occupies a sensible 'middle ground' ... it performs very well near the origin, but does not lead to drastic compromises in other parts of the parameter space."

Problems with inverse-gamma prior



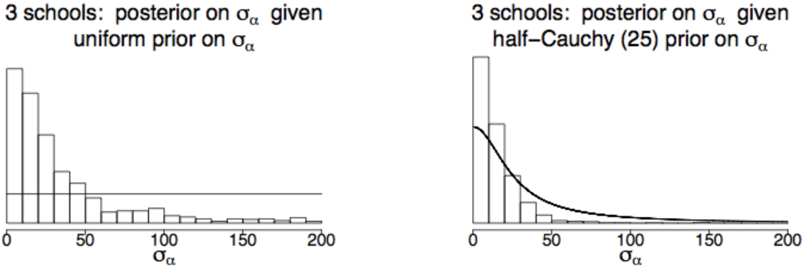
- Inv-gamma prior cuts off at 0

Problems with inverse-gamma prior



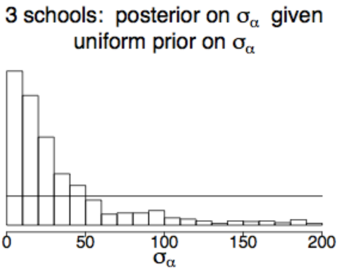
- Inv-gamma prior cuts off at 0

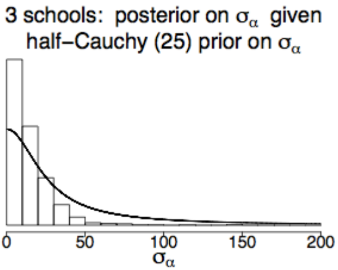
Problems with uniform prior



- Uniform prior doesn't cut off the long tail

Problems with uniform prior





- Uniform prior doesn't cut off the long tail

Priors and distributions for multilevel models

- Variance parameters indistinguishable from 0
- 8 schools
- 3 schools
- 1 or 2 schools??

Priors and distributions for multilevel models

- Variance parameters indistinguishable from 0
- 8 schools
- 3 schools
- 1 or 2 schools??

Priors and distributions for multilevel models

- Variance parameters indistinguishable from 0
- 8 schools
- 3 schools
- 1 or 2 schools??

Priors and distributions for multilevel models

- Variance parameters indistinguishable from 0
- 8 schools
- 3 schools
- 1 or 2 schools??

Priors and distributions for multilevel models

- Variance parameters indistinguishable from 0
- 8 schools
- 3 schools
- 1 or 2 schools??

Summary of Chapter 5

- Estimate hyperparameters from data
- Partial pooling
- Exchangeability and going beyond
- Difficulties when #groups is small

Summary of Chapter 5

- Estimate hyperparameters from data
- Partial pooling
- Exchangeability and going beyond
- Difficulties when #groups is small

Summary of Chapter 5

- Estimate hyperparameters from data
- Partial pooling
- Exchangeability and going beyond
- Difficulties when #groups is small

Summary of Chapter 5

- Estimate hyperparameters from data
- Partial pooling
- Exchangeability and going beyond
- Difficulties when #groups is small

Summary of Chapter 5

- Estimate hyperparameters from data
- Partial pooling
- Exchangeability and going beyond
- Difficulties when #groups is small