Bayesian Data Analysis, class 4b

Andrew Gelman

Chapter 5: Hierarchical models (part 1)

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
- $\bar{y}$ does *not* have approx normal distribution, but $p(\theta|y)$ is approximately normal

Theory problem

- Normal approximation to the posterior distribution from Cauchy data
- Second derivative, plotting the normal density
- $\bar{y}$ does *not* have approx normal distribution, but $p(\theta|y)$ is approximately normal

Theory problem

- Normal approximation to the posterior distribution from Cauchy data
- Second derivative, plotting the normal density
- $\bar{y}$ does *not* have approx normal distribution, but $p(\theta|y)$ is approximately normal

Theory problem

- Normal approximation to the posterior distribution from Cauchy data
- Second derivative, plotting the normal density
- $\bar{y}$ does *not* have approx normal distribution, but $p(\theta|y)$ is approximately normal

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 1)

- The rat tumor example
- The algebra of conjugate hierarchical models
- The hierarchical normal model
- The 8 schools example

5. Hierarchical models (part 1)

- The rat tumor example
- The algebra of conjugate hierarchical models
- The hierarchical normal model
- The 8 schools example

5. Hierarchical models (part 1)

- The rat tumor example
- The algebra of conjugate hierarchical models
- The hierarchical normal model
- The 8 schools example

5. Hierarchical models (part 1)

- The rat tumor example
- The algebra of conjugate hierarchical models
- The hierarchical normal model
- The 8 schools example

5. Hierarchical models (part 1)

- The rat tumor example
- The algebra of conjugate hierarchical models
- The hierarchical normal model
- The 8 schools example

Rat tumor data

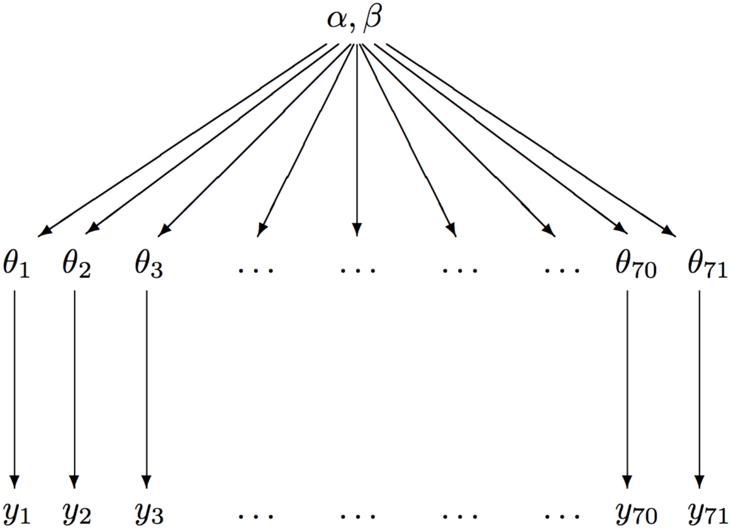
Previous experiments:

0/20	0/20	0/20	0/20	0/20	0/20	0/20	0/19	0/19	0/19
0/19	0/18	0/18	0/17	1/20	1/20	1/20	1/20	1/19	1/19
1/18	1/18	2/25	2/24	2/23	2/20	2/20	2/20	2/20	2/20
2/20	1/10	5/49	2/19	5/46	3/27	2/17	7/49	7/47	3/20
3/20	2/13	9/48	10/50	4/20	4/20	4/20	4/20	4/20	4/20
4/20	10/48	4/19	4/19	4/19	5/22	11/46	12/49	5/20	5/20
6/23	5/19	6/22	6/20	6/20	6/20	16/52	15/47	15/46	9/24

Current experiment:

4/14

Rat tumor model



5.1. Constructing a parameterized prior distribution

- The model:
  - $y \sim \text{Binomial}(n, \theta)$
  - $\theta \sim \text{Beta}(\alpha, \beta)$
- Data: $y = 4$, $n = 14$
- Inference: $\theta | y \sim \text{Beta}(\alpha + 4, \beta + 10)$
- Set α, β based on historical data
- Hierarchical model:
- Need to choose α, β

5.1. Constructing a parameterized prior distribution

- The model:
  - $y \sim \text{Binomial}(n, \theta)$
  - $\theta \sim \text{Beta}(\alpha, \beta)$
- Data: $y = 4$, $n = 14$
- Inference: $\theta | y \sim \text{Beta}(\alpha + 4, \beta + 10)$
- Set α, β based on historical data
- Hierarchical model:
- Need to choose α, β

5.1. Constructing a parameterized prior distribution

- The model:
  - $y \sim \text{Binomial}(n, \theta)$
  - $\theta \sim \text{Beta}(\alpha, \beta)$
- Data: $y = 4$, $n = 14$
- Inference: $\theta | y \sim \text{Beta}(\alpha + 4, \beta + 10)$
- Set α, β based on historical data
- Hierarchical model:
- Need to choose α, β

5.1. Constructing a parameterized prior distribution

- The model:
  - $y \sim \text{Binomial}(n, \theta)$
  - $\theta \sim \text{Beta}(\alpha, \beta)$
- Data: $y = 4$, $n = 14$
- Inference: $\theta | y \sim \text{Beta}(\alpha + 4, \beta + 10)$
- Set α, β based on historical data
- Hierarchical model:
- Need to choose α, β

5.1. Constructing a parameterized prior distribution

- The model:
  - $y \sim \text{Binomial}(n, \theta)$
  - $\theta \sim \text{Beta}(\alpha, \beta)$
- Data: $y = 4$, $n = 14$
- Inference: $\theta | y \sim \text{Beta}(\alpha + 4, \beta + 10)$
- Set α, β based on historical data
- Hierarchical model:
- Need to choose α, β

5.1. Constructing a parameterized prior distribution

- The model:
  - $y \sim \text{Binomial}(n, \theta)$
  - $\theta \sim \text{Beta}(\alpha, \beta)$
- Data: $y = 4,\ n = 14$
- Inference: $\theta | y \sim \text{Beta}(\alpha + 4, \beta + 10)$
- Set α, β based on historical data
- Hierarchical model:
  - $y_j \sim \text{Binomial}(n_j, \theta_j)$, for $j = 1, \ldots, 71$
    $\theta_1, \ldots, \theta_{71} \sim \text{Beta}(\alpha, \beta)$
- Need to choose α, β

5.1. Constructing a parameterized prior distribution

- The model:
  - $y \sim \text{Binomial}(n, \theta)$
  - $\theta \sim \text{Beta}(\alpha, \beta)$
- Data: $y = 4$, $n = 14$
- Inference: $\theta|y \sim \text{Beta}(\alpha + 4, \beta + 10)$
- Set α, β based on historical data
- Hierarchical model:
  - $y_j \sim \text{Binomial}(n_j, \theta_j)$, for $j = 1, \ldots, 71$
  - $\theta_1, \ldots, \theta_{71} \sim \text{Beta}(\alpha, \beta)$
- Need to choose α, β

5.1. Constructing a parameterized prior distribution

- The model:
  - $y \sim \text{Binomial}(n, \theta)$
  - $\theta \sim \text{Beta}(\alpha, \beta)$
- Data: $y = 4$, $n = 14$
- Inference: $\theta | y \sim \text{Beta}(\alpha + 4, \beta + 10)$
- Set α, β based on historical data
- Hierarchical model:
  - $y_j \sim \text{Binomial}(n_j, \theta_j)$, for $j = 1, \ldots, 71$
  - $\theta_1, \ldots, \theta_{71} \sim \text{Beta}(\alpha, \beta)$
- Need to choose α, β

5.1. Constructing a parameterized prior distribution

- The model:
  - $y \sim \text{Binomial}(n, \theta)$
  - $\theta \sim \text{Beta}(\alpha, \beta)$
- Data: $y = 4$, $n = 14$
- Inference: $\theta | y \sim \text{Beta}(\alpha + 4, \beta + 10)$
- Set α, β based on historical data
- Hierarchical model:
  - $y_j \sim \text{Binomial}(n_j, \theta_j)$, for $j = 1, \ldots, 71$
  - $\theta_1, \ldots, \theta_{71} \sim \text{Beta}(\alpha, \beta)$
- Need to choose α, β

5.1. Constructing a parameterized prior distribution

- The model:
  - $y \sim \text{Binomial}(n, \theta)$
  - $\theta \sim \text{Beta}(\alpha, \beta)$
- Data: $y = 4,\ n = 14$
- Inference: $\theta | y \sim \text{Beta}(\alpha + 4, \beta + 10)$
- Set α, β based on historical data
- Hierarchical model:
  - $y_j \sim \text{Binomial}(n_j, \theta_j)$, for $j = 1, \ldots, 71$
  - $\theta_1, \ldots, \theta_{71} \sim \text{Beta}(\alpha, \beta)$
- Need to choose α, β

5.1. Constructing a parameterized prior distribution

- The model:
  - $y \sim \text{Binomial}(n, \theta)$
  - $\theta \sim \text{Beta}(\alpha, \beta)$
- Data: $y = 4$, $n = 14$
- Inference: $\theta | y \sim \text{Beta}(\alpha + 4, \beta + 10)$
- Set α, β based on historical data
- Hierarchical model:
  - $y_j \sim \text{Binomial}(n_j, \theta_j)$, for $j = 1, \ldots, 71$
  - $\theta_1, \ldots, \theta_{71} \sim \text{Beta}(\alpha, \beta)$
- Need to choose α, β

5.2. Exchangeability and setting up hierarchical models

- $\theta_1, \ldots, \theta_J$ are *exchangeable* if $p(\theta_1, \ldots, \theta_J)$ is symmetric
- No information to distinguish the J cases
- "Exchangeable" is not the same as "identical"
- Consider the 71 rat tumor experiments
- Going beyond exchangeability
- Group-level predictors
- Going from the model to the probability density function, $p(\theta)$

5.2. Exchangeability and setting up hierarchical models

- $\theta_1, \ldots, \theta_J$ are *exchangeable* if $p(\theta_1, \ldots, \theta_J)$ is symmetric
- No information to distinguish the J cases
- "Exchangeable" is not the same as "identical"
- Consider the 71 rat tumor experiments
- Going beyond exchangeability
- Group-level predictors
- Going from the model to the probability density function, $p(\theta)$

5.2. Exchangeability and setting up hierarchical models

- $\theta_1, \ldots, \theta_J$ are *exchangeable* if $p(\theta_1, \ldots, \theta_J)$ is symmetric
- No information to distinguish the J cases
- "Exchangeable" is not the same as "identical"
- Consider the 71 rat tumor experiments
- Going beyond exchangeability
- Group-level predictors
- Going from the model to the probability density function, $p(\theta)$

5.2. Exchangeability and setting up hierarchical models

- $\theta_1, \ldots, \theta_J$ are *exchangeable* if $p(\theta_1, \ldots, \theta_J)$ is symmetric
- No information to distinguish the J cases
- "Exchangeable" is not the same as "identical"
- Consider the 71 rat tumor experiments
- Going beyond exchangeability
- Group-level predictors
- Going from the model to the probability density function, $p(\theta)$

5.2. Exchangeability and setting up hierarchical models

- $\theta_1, \ldots, \theta_J$ are *exchangeable* if $p(\theta_1, \ldots, \theta_J)$ is symmetric
- No information to distinguish the J cases
- "Exchangeable" is not the same as "identical"
- Consider the 71 rat tumor experiments
- Going beyond exchangeability
- Group-level predictors
- Going from the model to the probability density function, $p(\theta)$

5.2. Exchangeability and setting up hierarchical models

- $\theta_1, \ldots, \theta_J$ are *exchangeable* if $p(\theta_1, \ldots, \theta_J)$ is symmetric
- No information to distinguish the J cases
- "Exchangeable" is not the same as "identical"
- Consider the 71 rat tumor experiments
- Going beyond exchangeability
- Group-level predictors
- Going from the model to the probability density function, $p(\theta)$

5.2. Exchangeability and setting up hierarchical models

- $\theta_1, \ldots, \theta_J$ are *exchangeable* if $p(\theta_1, \ldots, \theta_J)$ is symmetric
- No information to distinguish the J cases
- "Exchangeable" is not the same as "identical"
- Consider the 71 rat tumor experiments
- Going beyond exchangeability
- Group-level predictors
- Going from the model to the probability density function, $p(\theta)$

5.2. Exchangeability and setting up hierarchical models

- $\theta_1, \ldots, \theta_J$ are *exchangeable* if $p(\theta_1, \ldots, \theta_J)$ is symmetric
- No information to distinguish the J cases
- "Exchangeable" is not the same as "identical"
- Consider the 71 rat tumor experiments
- Going beyond exchangeability
- Group-level predictors
- Going from the model to the probability density function, $p(\theta)$

5.3. Fully Bayesian analysis of conjugate hierarchical models

- $p(\phi, \theta | y) \propto p(\phi) p(\theta | \phi) p(y | \theta, \phi)$
- Conditional on the hyperparameters is easy:

$$p(\theta | \phi, y) \propto p(\theta | \phi) p(y | \theta, \phi)$$

- Marginal posterior distribution of the hyperparameters:

$$\begin{aligned} p(\phi | y) &= \int p(\phi, \theta | y) d\theta \\ &\propto p(\phi) \int p(\theta | \phi) p(y | \theta, \phi) d\theta \end{aligned}$$

- If you can do the integral, computation is direct:

5.3. Fully Bayesian analysis of conjugate hierarchical models

- $p(\phi, \theta|y) \propto p(\phi)p(\theta|\phi)p(y|\theta, \phi)$
- Conditional on the hyperparameters is easy:

$$p(\theta|\phi, y) \propto p(\theta|\phi)p(y|\theta, \phi)$$

- Marginal posterior distribution of the hyperparameters:

$$\begin{aligned} p(\phi|y) &= \int p(\phi, \theta|y)d\theta \\ &\propto p(\phi)\int p(\theta|\phi)p(y|\theta, \phi)d\theta \end{aligned}$$

- If you can do the integral, computation is direct:

5.3. Fully Bayesian analysis of conjugate hierarchical models

- $p(\phi, \theta | y) \propto p(\phi) p(\theta | \phi) p(y | \theta, \phi)$
- Conditional on the hyperparameters is easy:

$$p(\theta | \phi, y) \propto p(\theta | \phi) p(y | \theta, \phi)$$

- Marginal posterior distribution of the hyperparameters:

$$\begin{aligned} p(\phi | y) &= \int p(\phi, \theta | y) d\theta \\ &\propto p(\phi) \int p(\theta | \phi) p(y | \theta, \phi) d\theta \end{aligned}$$

- If you can do the integral, computation is direct:
  - Compute $p(\phi | y)$ on a grid of ϕ
  - For $s = 1, \ldots, S$:

5.3. Fully Bayesian analysis of conjugate hierarchical models

- $p(\phi, \theta | y) \propto p(\phi) p(\theta | \phi) p(y | \theta, \phi)$
- Conditional on the hyperparameters is easy:

$$p(\theta | \phi, y) \propto p(\theta | \phi) p(y | \theta, \phi)$$

- Marginal posterior distribution of the hyperparameters:

$$\begin{aligned} p(\phi | y) &= \int p(\phi, \theta | y) d\theta \\ &\propto p(\phi) \int p(\theta | \phi) p(y | \theta, \phi) d\theta \end{aligned}$$

- If you can do the integral, computation is direct:
  - Compute $p(\phi | y)$ on a grid of ϕ
  - For $s = 1, \ldots, S$

5.3. Fully Bayesian analysis of conjugate hierarchical models

- $p(\phi, \theta|y) \propto p(\phi)p(\theta|\phi)p(y|\theta, \phi)$
- Conditional on the hyperparameters is easy:

$$p(\theta|\phi, y) \propto p(\theta|\phi)p(y|\theta, \phi)$$

- Marginal posterior distribution of the hyperparameters:

$$\begin{aligned} p(\phi|y) &= \int p(\phi, \theta|y)d\theta \\ &\propto p(\phi) \int p(\theta|\phi)p(y|\theta, \phi)d\theta \end{aligned}$$

- If you can do the integral, computation is direct:
  - Compute $p(\phi|y)$ on a grid of ϕ
  - For $s = 1, \ldots, S$:

5.3. Fully Bayesian analysis of conjugate hierarchical models

- $p(\phi, \theta|y) \propto p(\phi)p(\theta|\phi)p(y|\theta, \phi)$
- Conditional on the hyperparameters is easy:

$$p(\theta|\phi, y) \propto p(\theta|\phi)p(y|\theta, \phi)$$

- Marginal posterior distribution of the hyperparameters:

$$\begin{aligned} p(\phi|y) &= \int p(\phi, \theta|y)d\theta \\ &\propto p(\phi)\int p(\theta|\phi)p(y|\theta, \phi)d\theta \end{aligned}$$

- If you can do the integral, computation is direct:
  - Compute $p(\phi|y)$ on a grid of ϕ
  - For $s = 1, \ldots, S$:
    - Sample ϕ^s from grid
    - Sample θ^s from $p(\theta^s|\phi^s, y)$

5.3. Fully Bayesian analysis of conjugate hierarchical models

- $p(\phi, \theta|y) \propto p(\phi)p(\theta|\phi)p(y|\theta, \phi)$
- Conditional on the hyperparameters is easy:

$$p(\theta|\phi, y) \propto p(\theta|\phi)p(y|\theta, \phi)$$

- Marginal posterior distribution of the hyperparameters:

$$\begin{aligned} p(\phi|y) &= \int p(\phi, \theta|y)d\theta \\ &\propto p(\phi) \int p(\theta|\phi)p(y|\theta, \phi)d\theta \end{aligned}$$

- If you can do the integral, computation is direct:
  - Compute $p(\phi|y)$ on a grid of ϕ
  - For $s = 1, \ldots, S$:
    - Sample ϕ^s from grid
    - Sample θ^s from $p(\theta^s|\phi^s, y)$

5.3. Fully Bayesian analysis of conjugate hierarchical models

- $p(\phi, \theta|y) \propto p(\phi)p(\theta|\phi)p(y|\theta, \phi)$
- Conditional on the hyperparameters is easy:

$$p(\theta|\phi, y) \propto p(\theta|\phi)p(y|\theta, \phi)$$

- Marginal posterior distribution of the hyperparameters:

$$\begin{aligned} p(\phi|y) &= \int p(\phi, \theta|y)d\theta \\ &\propto p(\phi)\int p(\theta|\phi)p(y|\theta, \phi)d\theta \end{aligned}$$

- If you can do the integral, computation is direct:
  - Compute $p(\phi|y)$ on a grid of ϕ
  - For $s = 1, \ldots, S$:
    - Sample ϕ^s from grid
    - Sample θ^s from $p(\theta^s|\phi^s, y)$

5.3. Fully Bayesian analysis of conjugate hierarchical models

- $p(\phi, \theta|y) \propto p(\phi)p(\theta|\phi)p(y|\theta, \phi)$
- Conditional on the hyperparameters is easy:

$$p(\theta|\phi, y) \propto p(\theta|\phi)p(y|\theta, \phi)$$

- Marginal posterior distribution of the hyperparameters:

$$\begin{aligned} p(\phi|y) &= \int p(\phi, \theta|y)d\theta \\ &\propto p(\phi) \int p(\theta|\phi)p(y|\theta, \phi)d\theta \end{aligned}$$

- If you can do the integral, computation is direct:
  - Compute $p(\phi|y)$ on a grid of ϕ
  - For $s = 1, \ldots, S$:
    - Sample ϕ^s from grid
    - Sample θ^s from $p(\theta^s|\phi^s, y)$

Rat tumor model: algebra

- The model:
  - $y_j \sim \text{Binomial}(n_j, \theta_j)$
  - $\theta_j \sim \text{Beta}(\alpha, \beta)$
  - What are the assumptions?
- Conditional posterior density:

$$p(\theta|\alpha,\beta,y) \propto \prod_{j=1}^{J} \theta_j^{\alpha+y_j-1}(1-\theta_j)^{\beta+n_j-y_j-1}$$

- Joint posterior density:

$$p(\theta,\alpha,\beta|y) \propto p(\alpha,\beta)\prod_{j=1}^{J}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta_j^{\alpha-1}(1-\theta_j)^{\beta-1}\prod_{j=1}^{J}\theta_j^{y_j}(1-\theta_j)^{n_j-y_j}$$

- Marginal posterior density (integrate out the J-dimensional θ):

$$p(\alpha,\beta|y) \propto p(\alpha,\beta)\prod_{j=1}^{J}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha+y_j)\Gamma(\beta+n_j-y_j)}{\Gamma(\alpha+\beta+n_j)}$$

Rat tumor model: algebra

- The model:
  - $y_j \sim \text{Binomial}(n_j, \theta_j)$
  - $\theta_j \sim \text{Beta}(\alpha, \beta)$
  - What are the assumptions?
- Conditional posterior density:

$$p(\theta|\alpha,\beta,y) \propto \prod_{j=1}^{J} \theta_j^{\alpha+y_j-1}(1-\theta_j)^{\beta+n_j-y_j-1}$$

- Joint posterior density:

$$p(\theta,\alpha,\beta|y) \propto p(\alpha,\beta)\prod_{j=1}^{J}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta_j^{\alpha-1}(1-\theta_j)^{\beta-1}\prod_{j=1}^{J}\theta_j^{y_j}(1-\theta_j)^{n_j-y_j}$$

- Marginal posterior density (integrate out the J-dimensional θ):

$$p(\alpha,\beta|y) \propto p(\alpha,\beta)\prod_{j=1}^{J}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha+y_j)\Gamma(\beta+n_j-y_j)}{\Gamma(\alpha+\beta+n_j)}$$

Rat tumor model: algebra

- The model:
  - $y_j \sim \text{Binomial}(n_j, \theta_j)$
  - $\theta_j \sim \text{Beta}(\alpha, \beta)$
  - What are the assumptions?
- Conditional posterior density:

$$p(\theta|\alpha, \beta, y) \propto \prod_{j=1}^{J} \theta_j^{\alpha+y_j-1}(1-\theta_j)^{\beta+n_j-y_j-1}$$

- Joint posterior density:

$$p(\theta, \alpha, \beta|y) \propto p(\alpha, \beta) \prod_{j=1}^{J} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_j^{\alpha-1}(1-\theta_j)^{\beta-1} \prod_{j=1}^{J} \theta_j^{y_j}(1-\theta_j)^{n_j-y_j}$$

- Marginal posterior density (integrate out the J-dimensional θ):

$$p(\alpha, \beta|y) \propto p(\alpha, \beta) \prod_{j=1}^{J} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+y_j)\Gamma(\beta+n_j-y_j)}{\Gamma(\alpha+\beta+n_j)}$$

Rat tumor model: algebra

- The model:
  - $y_j \sim \text{Binomial}(n_j, \theta_j)$
  - $\theta_j \sim \text{Beta}(\alpha, \beta)$
  - What are the assumptions?
- Conditional posterior density:

$$p(\theta|\alpha, \beta, y) \propto \prod_{j=1}^{J} \theta_j^{\alpha+y_j-1}(1-\theta_j)^{\beta+n_j-y_j-1}$$

- Joint posterior density:

$$p(\theta, \alpha, \beta|y) \propto p(\alpha, \beta) \prod_{j=1}^{J} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_j^{\alpha-1}(1-\theta_j)^{\beta-1} \prod_{j=1}^{J} \theta_j^{y_j}(1-\theta_j)^{n_j-y_j}$$

- Marginal posterior density (integrate out the J-dimensional θ):

$$p(\alpha, \beta|y) \propto p(\alpha, \beta) \prod_{j=1}^{J} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+y_j)\Gamma(\beta+n_j-y_j)}{\Gamma(\alpha+\beta+n_j)}$$

Rat tumor model: algebra

- The model:
  - $y_j \sim \text{Binomial}(n_j, \theta_j)$
  - $\theta_j \sim \text{Beta}(\alpha, \beta)$
  - What are the assumptions?
- Conditional posterior density:

$$p(\theta|\alpha,\beta,y) \propto \prod_{j=1}^{J} \theta_j^{\alpha+y_j-1}(1-\theta_j)^{\beta+n_j-y_j-1}$$

- Joint posterior density:

$$p(\theta,\alpha,\beta|y) \propto p(\alpha,\beta)\prod_{j=1}^{J}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta_j^{\alpha-1}(1-\theta_j)^{\beta-1}\prod_{j=1}^{J}\theta_j^{y_j}(1-\theta_j)^{n_j-y_j}$$

- Marginal posterior density (integrate out the J-dimensional θ):

$$p(\alpha,\beta|y) \propto p(\alpha,\beta)\prod_{j=1}^{J}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha+y_j)\Gamma(\beta+n_j-y_j)}{\Gamma(\alpha+\beta+n_j)}$$

Rat tumor model: algebra

- The model:
  - $y_j \sim \text{Binomial}(n_j, \theta_j)$
  - $\theta_j \sim \text{Beta}(\alpha, \beta)$
  - What are the assumptions?
- Conditional posterior density:

$$p(\theta|\alpha, \beta, y) \propto \prod_{j=1}^{J} \theta_j^{\alpha+y_j-1}(1-\theta_j)^{\beta+n_j-y_j-1}$$

- Joint posterior density:

$$p(\theta, \alpha, \beta|y) \propto p(\alpha, \beta) \prod_{j=1}^{J} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_j^{\alpha-1}(1-\theta_j)^{\beta-1} \prod_{j=1}^{J} \theta_j^{y_j}(1-\theta_j)^{n_j-y_j}$$

- Marginal posterior density (integrate out the J-dimensional θ):

$$p(\alpha, \beta|y) \propto p(\alpha, \beta) \prod_{j=1}^{J} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+y_j)\Gamma(\beta+n_j-y_j)}{\Gamma(\alpha+\beta+n_j)}$$

Rat tumor model: algebra

- The model:
  - $y_j \sim \text{Binomial}(n_j, \theta_j)$
  - $\theta_j \sim \text{Beta}(\alpha, \beta)$
  - What are the assumptions?
- Conditional posterior density:

$$p(\theta|\alpha, \beta, y) \propto \prod_{j=1}^{J} \theta_j^{\alpha+y_j-1}(1-\theta_j)^{\beta+n_j-y_j-1}$$

- Joint posterior density:

$$p(\theta, \alpha, \beta|y) \propto p(\alpha, \beta) \prod_{j=1}^{J} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_j^{\alpha-1}(1-\theta_j)^{\beta-1} \prod_{j=1}^{J} \theta_j^{y_j}(1-\theta_j)^{n_j-y_j}$$

- Marginal posterior density (integrate out the J-dimensional θ):

$$p(\alpha, \beta|y) \propto p(\alpha, \beta) \prod_{j=1}^{J} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+y_j)\Gamma(\beta+n_j-y_j)}{\Gamma(\alpha+\beta+n_j)}$$

Rat tumor model: algebra

- The model:
  - $y_j \sim \text{Binomial}(n_j, \theta_j)$
  - $\theta_j \sim \text{Beta}(\alpha, \beta)$
  - What are the assumptions?
- Conditional posterior density:

$$p(\theta|\alpha,\beta,y) \propto \prod_{j=1}^{J} \theta_j^{\alpha+y_j-1}(1-\theta_j)^{\beta+n_j-y_j-1}$$

- Joint posterior density:

$$p(\theta,\alpha,\beta|y) \propto p(\alpha,\beta)\prod_{j=1}^{J}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta_j^{\alpha-1}(1-\theta_j)^{\beta-1}\prod_{j=1}^{J}\theta_j^{y_j}(1-\theta_j)^{n_j-y_j}$$

- Marginal posterior density (integrate out the J-dimensional θ):

$$p(\alpha,\beta|y) \propto p(\alpha,\beta)\prod_{j=1}^{J}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha+y_j)\Gamma(\beta+n_j-y_j)}{\Gamma(\alpha+\beta+n_j)}$$

Rat tumor model: prior distribution on (α, β)

- $p(\theta|\alpha, \beta)$ already set
- $p(\alpha, \beta) = ?$
- Reparameterize to $\text{logit}(\frac{\alpha}{\alpha+\beta}) = \log(\frac{\alpha}{\beta})$ and $\log(\alpha+\beta)$
- Logit of prior mean, and prior "sample size"
- $p(\log(\frac{\alpha}{\beta}), \log(\alpha+\beta)) \propto 1$ doesn't work (improper posterior)
- Uniform on $[-10^{10}, 10^{10}] \times [-10^{10}, 10^{10}]$ wouldn't work either!
- Instead, try $p(\frac{\alpha}{\alpha+\beta}, (\alpha+\beta)^{-1/2}) \propto 1$
- Don't forget the Jacobian
- Noninformative prior distribution as placeholder

Rat tumor model: prior distribution on (α, β)

- $p(\theta|\alpha, \beta)$ already set
- $p(\alpha, \beta) = ?$
- Reparameterize to $\text{logit}(\frac{\alpha}{\alpha+\beta}) = \log(\frac{\alpha}{\beta})$ and $\log(\alpha+\beta)$
- Logit of prior mean, and prior "sample size"
- $p(\log(\frac{\alpha}{\beta}), \log(\alpha+\beta)) \propto 1$ doesn't work (improper posterior)
- Uniform on $[-10^{10}, 10^{10}] \times [-10^{10}, 10^{10}]$ wouldn't work either!
- Instead, try $p(\frac{\alpha}{\alpha+\beta}, (\alpha+\beta)^{-1/2}) \propto 1$
- Don't forget the Jacobian
- Noninformative prior distribution as placeholder

Rat tumor model: prior distribution on (α, β)

- $p(\theta|\alpha, \beta)$ already set
- $p(\alpha, \beta) = ?$
- Reparameterize to $\text{logit}(\frac{\alpha}{\alpha+\beta}) = \log(\frac{\alpha}{\beta})$ and $\log(\alpha+\beta)$
- Logit of prior mean, and prior "sample size"
- $p(\log(\frac{\alpha}{\beta}), \log(\alpha+\beta)) \propto 1$ doesn't work (improper posterior)
- Uniform on $[-10^{10}, 10^{10}] \times [-10^{10}, 10^{10}]$ wouldn't work either!
- Instead, try $p(\frac{\alpha}{\alpha+\beta}, (\alpha+\beta)^{-1/2}) \propto 1$
- Don't forget the Jacobian
- Noninformative prior distribution as placeholder

Rat tumor model: prior distribution on (α, β)

- $p(\theta|\alpha, \beta)$ already set
- $p(\alpha, \beta) = ?$
- Reparameterize to $\text{logit}(\frac{\alpha}{\alpha+\beta}) = \log(\frac{\alpha}{\beta})$ and $\log(\alpha+\beta)$
- Logit of prior mean, and prior "sample size"
- $p(\log(\frac{\alpha}{\beta}), \log(\alpha+\beta)) \propto 1$ doesn't work (improper posterior)
- Uniform on $[-10^{10}, 10^{10}] \times [-10^{10}, 10^{10}]$ wouldn't work either!
- Instead, try $p(\frac{\alpha}{\alpha+\beta}, (\alpha+\beta)^{-1/2}) \propto 1$
- Don't forget the Jacobian
- Noninformative prior distribution as placeholder

Rat tumor model: prior distribution on (α, β)

- $p(\theta|\alpha, \beta)$ already set
- $p(\alpha, \beta) = ?$
- Reparameterize to $\text{logit}(\frac{\alpha}{\alpha+\beta}) = \log(\frac{\alpha}{\beta})$ and $\log(\alpha+\beta)$
- Logit of prior mean, and prior "sample size"
- $p(\log(\frac{\alpha}{\beta}), \log(\alpha+\beta)) \propto 1$ doesn't work (improper posterior)
- Uniform on $[-10^{10}, 10^{10}] \times [-10^{10}, 10^{10}]$ wouldn't work either!
- Instead, try $p(\frac{\alpha}{\alpha+\beta}, (\alpha+\beta)^{-1/2}) \propto 1$
- Don't forget the Jacobian
- Noninformative prior distribution as placeholder

Rat tumor model: prior distribution on (α, β)

- $p(\theta|\alpha, \beta)$ already set
- $p(\alpha, \beta) = ?$
- Reparameterize to $\text{logit}(\frac{\alpha}{\alpha+\beta}) = \log(\frac{\alpha}{\beta})$ and $\log(\alpha+\beta)$
- Logit of prior mean, and prior "sample size"
- $p(\log(\frac{\alpha}{\beta}), \log(\alpha+\beta)) \propto 1$ doesn't work (improper posterior)
- Uniform on $[-10^{10}, 10^{10}] \times [-10^{10}, 10^{10}]$ wouldn't work either!
- Instead, try $p(\frac{\alpha}{\alpha+\beta}, (\alpha+\beta)^{-1/2}) \propto 1$
- Don't forget the Jacobian
- Noninformative prior distribution as placeholder

Rat tumor model: prior distribution on (α, β)

- $p(\theta|\alpha, \beta)$ already set
- $p(\alpha, \beta) = ?$
- Reparameterize to $\text{logit}(\frac{\alpha}{\alpha+\beta}) = \log(\frac{\alpha}{\beta})$ and $\log(\alpha+\beta)$
- Logit of prior mean, and prior "sample size"
- $p(\log(\frac{\alpha}{\beta}), \log(\alpha+\beta)) \propto 1$ doesn't work (improper posterior)
- Uniform on $[-10^{10}, 10^{10}] \times [-10^{10}, 10^{10}]$ wouldn't work either!
- Instead, try $p(\frac{\alpha}{\alpha+\beta}, (\alpha+\beta)^{-1/2}) \propto 1$
- Don't forget the Jacobian
- Noninformative prior distribution as placeholder

Rat tumor model: prior distribution on (α, β)

- $p(\theta|\alpha, \beta)$ already set
- $p(\alpha, \beta) = ?$
- Reparameterize to $\text{logit}(\frac{\alpha}{\alpha+\beta}) = \log(\frac{\alpha}{\beta})$ and $\log(\alpha+\beta)$
- Logit of prior mean, and prior "sample size"
- $p(\log(\frac{\alpha}{\beta}), \log(\alpha+\beta)) \propto 1$ doesn't work (improper posterior)
- Uniform on $[-10^{10}, 10^{10}] \times [-10^{10}, 10^{10}]$ wouldn't work either!
- Instead, try $p(\frac{\alpha}{\alpha+\beta}, (\alpha+\beta)^{-1/2}) \propto 1$
- Don't forget the Jacobian
- Noninformative prior distribution as placeholder

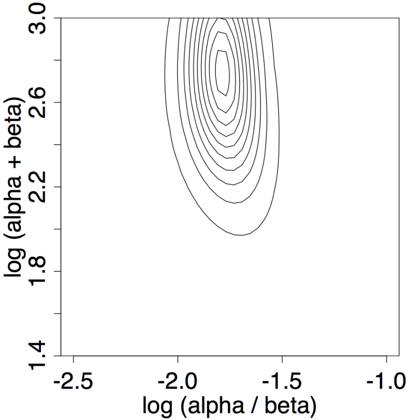
Rat tumor model: prior distribution on (α, β)

- $p(\theta|\alpha, \beta)$ already set
- $p(\alpha, \beta) =?$
- Reparameterize to $\text{logit}(\frac{\alpha}{\alpha+\beta}) = \log(\frac{\alpha}{\beta})$ and $\log(\alpha+\beta)$
- Logit of prior mean, and prior "sample size"
- $p(\log(\frac{\alpha}{\beta}), \log(\alpha+\beta)) \propto 1$ doesn't work (improper posterior)
- Uniform on $[-10^{10}, 10^{10}] \times [-10^{10}, 10^{10}]$ wouldn't work either!
- Instead, try $p(\frac{\alpha}{\alpha+\beta}, (\alpha+\beta)^{-1/2}) \propto 1$
- Don't forget the Jacobian
- Noninformative prior distribution as placeholder

Rat tumor model: prior distribution on (α, β)

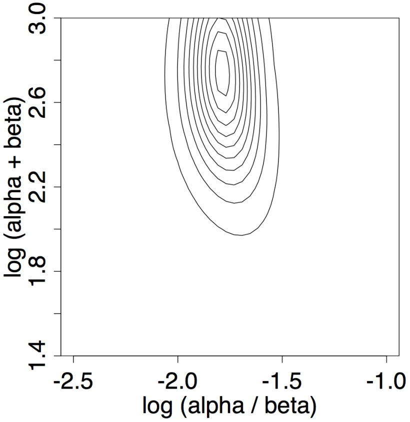
- $p(\theta|\alpha, \beta)$ already set
- $p(\alpha, \beta) = ?$
- Reparameterize to $\text{logit}(\frac{\alpha}{\alpha+\beta}) = \log(\frac{\alpha}{\beta})$ and $\log(\alpha+\beta)$
- Logit of prior mean, and prior "sample size"
- $p(\log(\frac{\alpha}{\beta}), \log(\alpha+\beta)) \propto 1$ doesn't work (improper posterior)
- Uniform on $[-10^{10}, 10^{10}] \times [-10^{10}, 10^{10}]$ wouldn't work either!
- Instead, try $p(\frac{\alpha}{\alpha+\beta}, (\alpha+\beta)^{-1/2}) \propto 1$
- Don't forget the Jacobian
- Noninformative prior distribution as placeholder

Rat tumor model: first try



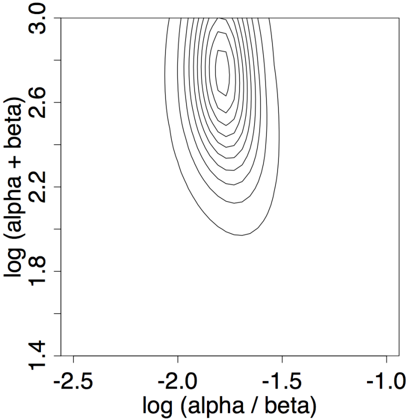
- Computed on grid
- Centered and scaled based on crude estimate and s.e.

Rat tumor model: first try



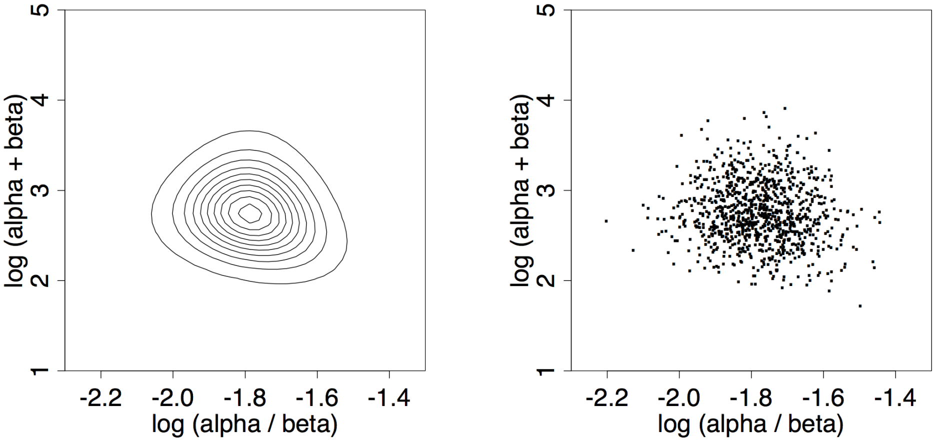
- Computed on grid
- Centered and scaled based on crude estimate and s.e.

Rat tumor model: first try

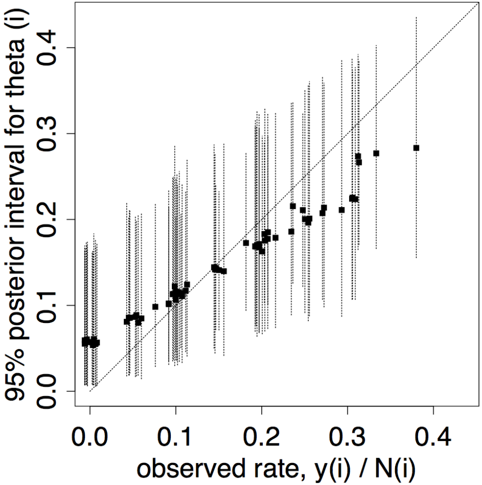


- Computed on grid
- Centered and scaled based on crude estimate and s.e.

Rat tumor model: contour plots and simulations



Rat tumor model: partial pooling



5.4. Exchangeable parameters from a normal model

- The model:
  - $\bar{y}_j \sim \mathrm{N}(\theta_j, \sigma_j^2)$
  - $\theta_j \sim \mathrm{N}(\mu, \tau^2)$
  - What are the assumptions?
- Conditional posterior density:

$$\theta|\mu, \tau, y \sim \mathrm{N}\left(\frac{\frac{1}{\cdots}\cdots + \frac{1}{\cdots}\cdots}{\frac{1}{\cdots} + \frac{1}{\cdots}}, \frac{1}{\frac{1}{\cdots} + \frac{1}{\cdots}}\right)$$

- Average over marginal posterior density of μ, τ
- Problems with simple point estimates of μ, τ

5.4. Exchangeable parameters from a normal model

- The model:
  - $\bar{y}_j \sim \mathrm{N}(\theta_j, \sigma_j^2)$
  - $\theta_j \sim \mathrm{N}(\mu, \tau^2)$
  - What are the assumptions?
- Conditional posterior density:

$$\theta|\mu, \tau, y \sim \mathrm{N}\left(\frac{\frac{1}{\cdots}\cdots + \frac{1}{\cdots}\cdots}{\frac{1}{\cdots} + \frac{1}{\cdots}}, \frac{1}{\frac{1}{\cdots} + \frac{1}{\cdots}}\right)$$

- Average over marginal posterior density of μ, τ
- Problems with simple point estimates of μ, τ

5.4. Exchangeable parameters from a normal model

- The model:
  - $\bar{y}_j \sim \mathrm{N}(\theta_j, \sigma_j^2)$
  - $\theta_j \sim \mathrm{N}(\mu, \tau^2)$
  - What are the assumptions?
- Conditional posterior density:

$$\theta|\mu, \tau, y \sim \mathrm{N}\left(\frac{\frac{1}{\cdots}\cdots + \frac{1}{\cdots}\cdots}{\frac{1}{\cdots} + \frac{1}{\cdots}}, \frac{1}{\frac{1}{\cdots} + \frac{1}{\cdots}}\right)$$

- Average over marginal posterior density of μ, τ
- Problems with simple point estimates of μ, τ

5.4. Exchangeable parameters from a normal model

- The model:
  - $\bar{y}_j \sim \mathrm{N}(\theta_j, \sigma_j^2)$
  - $\theta_j \sim \mathrm{N}(\mu, \tau^2)$
  - What are the assumptions?
- Conditional posterior density:

$$\theta|\mu, \tau, y \sim \mathrm{N}\left(\frac{\frac{1}{\cdots}\cdots + \frac{1}{\cdots}\cdots}{\frac{1}{\cdots} + \frac{1}{\cdots}}, \frac{1}{\frac{1}{\cdots} + \frac{1}{\cdots}}\right)$$

  - Partial pooling (shrinkage) determined by τ
- Average over marginal posterior density of μ, τ
- Problems with simple point estimates of μ, τ

5.4. Exchangeable parameters from a normal model

- The model:
  - $\bar{y}_j \sim \mathrm{N}(\theta_j, \sigma_j^2)$
  - $\theta_j \sim \mathrm{N}(\mu, \tau^2)$
  - What are the assumptions?
- Conditional posterior density:

$$\theta|\mu, \tau, y \sim \mathrm{N}\left(\frac{\frac{1}{\cdots}\cdots + \frac{1}{\cdots}\cdots}{\frac{1}{\cdots} + \frac{1}{\cdots}}, \frac{1}{\frac{1}{\cdots} + \frac{1}{\cdots}}\right)$$

  - Partial pooling (shrinkage) determined by τ
- Average over marginal posterior density of μ, τ
- Problems with simple point estimates of μ, τ

5.4. Exchangeable parameters from a normal model

- The model:
  - $\bar{y}_j \sim \mathrm{N}(\theta_j, \sigma_j^2)$
  - $\theta_j \sim \mathrm{N}(\mu, \tau^2)$
  - What are the assumptions?
- Conditional posterior density:

$$\theta|\mu, \tau, y \sim \mathrm{N}\left(\frac{\frac{1}{\cdots}\cdots + \frac{1}{\cdots}\cdots}{\frac{1}{\cdots} + \frac{1}{\cdots}}, \frac{1}{\frac{1}{\cdots} + \frac{1}{\cdots}}\right)$$

  - Partial pooling (shrinkage) determined by τ
- Average over marginal posterior density of μ, τ
- Problems with simple point estimates of μ, τ

5.4. Exchangeable parameters from a normal model

- The model:
  - $\bar{y}_j \sim \mathrm{N}(\theta_j, \sigma_j^2)$
  - $\theta_j \sim \mathrm{N}(\mu, \tau^2)$
  - What are the assumptions?
- Conditional posterior density:

$$\theta|\mu, \tau, y \sim \mathrm{N}\left(\frac{\frac{1}{\cdots}\cdots + \frac{1}{\cdots}\cdots}{\frac{1}{\cdots} + \frac{1}{\cdots}}, \frac{1}{\frac{1}{\cdots} + \frac{1}{\cdots}}\right)$$

  - Partial pooling (shrinkage) determined by τ
- Average over marginal posterior density of μ, τ
- Problems with simple point estimates of μ, τ

5.4. Exchangeable parameters from a normal model

- The model:
  - $\bar{y}_j \sim \mathrm{N}(\theta_j, \sigma_j^2)$
  - $\theta_j \sim \mathrm{N}(\mu, \tau^2)$
  - What are the assumptions?
- Conditional posterior density:

$$\theta|\mu,\tau,y \sim \mathrm{N}\left(\frac{\frac{1}{\dots}\dots + \frac{1}{\dots}\dots}{\frac{1}{\dots} + \frac{1}{\dots}}, \frac{1}{\frac{1}{\dots} + \frac{1}{\dots}}\right)$$

  - Partial pooling (shrinkage) determined by τ
- Average over marginal posterior density of μ, τ
- Problems with simple point estimates of μ, τ

5.4. Exchangeable parameters from a normal model

- The model:
  - $\bar{y}_j \sim \mathrm{N}(\theta_j, \sigma_j^2)$
  - $\theta_j \sim \mathrm{N}(\mu, \tau^2)$
  - What are the assumptions?
- Conditional posterior density:

$$\theta|\mu, \tau, y \sim \mathrm{N}\left(\frac{\frac{1}{\dots}\dots + \frac{1}{\dots}\dots}{\frac{1}{\dots} + \frac{1}{\dots}}, \frac{1}{\frac{1}{\dots} + \frac{1}{\dots}}\right)$$

  - Partial pooling (shrinkage) determined by τ
- Average over marginal posterior density of μ, τ
- Problems with simple point estimates of μ, τ

5.5. Example: parallel experiments in eight schools

- Pre-test, randomized treatment, post-test on each of 8 schools
- Inferences from separate regressions:

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

- Separate estimates
- Pooled estimate

5.5. Example: parallel experiments in eight schools

- Pre-test, randomized treatment, post-test on each of 8 schools
- Inferences from separate regressions:

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

- Separate estimates
- Pooled estimate

5.5. Example: parallel experiments in eight schools

- Pre-test, randomized treatment, post-test on each of 8 schools
- Inferences from separate regressions:

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

- Separate estimates
- Pooled estimate

5.5. Example: parallel experiments in eight schools

- Pre-test, randomized treatment, post-test on each of 8 schools
- Inferences from separate regressions:

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

- Separate estimates
- Pooled estimate

5.5. Example: parallel experiments in eight schools

- Pre-test, randomized treatment, post-test on each of 8 schools
- Inferences from separate regressions:

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

- Separate estimates
- Pooled estimate

Homework due beginning of class 5b

- All assignments are at http://www.stat.columbia.edu/~gelman/bda.course/homeworks.pdf
  - Theory problem: Exchangeable models and conditional independence
  - Computing problem: Simulation of a discrete stochastic process
  - Applied problem: Fitting and checking a stochastic learning model

Homework due beginning of class 5b

- All assignments are at http://www.stat.columbia.edu/~gelman/bda.course/homeworks.pdf
  - Theory problem: Exchangeable models and conditional independence
  - Computing problem: Simulation of a discrete stochastic process
  - Applied problem: Fitting and checking a stochastic learning model

Homework due beginning of class 5b

- All assignments are at http://www.stat.columbia.edu/~gelman/bda.course/homeworks.pdf
  - Theory problem: Exchangeable models and conditional independence
  - Computing problem: Simulation of a discrete stochastic process
  - Applied problem: Fitting and checking a stochastic learning model

Homework due beginning of class 5b

- All assignments are at http://www.stat.columbia.edu/~gelman/bda.course/homeworks.pdf
  - Theory problem: Exchangeable models and conditional independence
  - Computing problem: Simulation of a discrete stochastic process
  - Applied problem: Fitting and checking a stochastic learning model

Homework due beginning of class 5b

- All assignments are at http://www.stat.columbia.edu/~gelman/bda.course/homeworks.pdf
  - Theory problem: Exchangeable models and conditional independence
  - Computing problem: Simulation of a discrete stochastic process
  - Applied problem: Fitting and checking a stochastic learning model