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
- $\tilde{y}$ does *not* have approx normal distribution, but $p(\theta|y)$ is approximately normal
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
Poisson regression: check that posterior inferences are consistent with true parameter values
Poisson regression: check that posterior inferences are consistent with true parameter values
Basketball shooting again: $\theta_i$ is improvement in success probability for person $i$

Prior distribution for mean and standard deviation of $\theta_i$ in the population

Sidestepping causal questions
Basketball shooting again: $\theta_i$ is improvement in success probability for person $i$.

Prior distribution for mean and standard deviation of $\theta_i$ in the population.

Sidestepping causal questions.
Basketball shooting again: $\theta_i$ is improvement in success probability for person $i$

Prior distribution for mean and standard deviation of $\theta_i$ in the population

Sidestepping causal questions
Basketball shooting again: $\theta_i$ is improvement in success probability for person $i$

Prior distribution for mean and standard deviation of $\theta_i$ 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:

<table>
<thead>
<tr>
<th>0/20</th>
<th>0/20</th>
<th>0/20</th>
<th>0/20</th>
<th>0/20</th>
<th>0/20</th>
<th>0/20</th>
<th>0/20</th>
<th>0/19</th>
<th>0/19</th>
<th>0/19</th>
</tr>
</thead>
<tbody>
<tr>
<td>0/19</td>
<td>0/18</td>
<td>0/18</td>
<td>0/17</td>
<td>1/20</td>
<td>1/20</td>
<td>1/20</td>
<td>1/20</td>
<td>1/19</td>
<td>1/19</td>
<td>1/19</td>
</tr>
<tr>
<td>1/18</td>
<td>1/18</td>
<td>2/25</td>
<td>2/24</td>
<td>2/23</td>
<td>2/20</td>
<td>2/20</td>
<td>2/20</td>
<td>2/19</td>
<td>2/20</td>
<td>2/20</td>
</tr>
</tbody>
</table>

Current experiment:

4/14
Rat tumor model

\[ \alpha, \beta \quad \theta_1 \quad \theta_2 \quad \theta_3 \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \theta_{70} \quad \theta_{71} \quad y_{1} \quad y_{2} \quad y_{3} \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad y_{70} \quad y_{71} \]
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 $\alpha, \beta$ 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 $\alpha, \beta$
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 $\alpha, \beta$ based on historical data
- Hierarchical model:
  - Need to choose $\alpha, \beta$
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** $\alpha, \beta$ based on historical data

- **Hierarchical model:**

- **Need to choose** $\alpha, \beta$
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 \( \alpha, \beta \) 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 \( \alpha, \beta \)
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 $\alpha, \beta$ 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 $\alpha, \beta$
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 $\alpha, \beta$ 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 $\alpha, \beta$
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 $\alpha, \beta$ 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 $\alpha, \beta$
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 $\alpha, \beta$ 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 $\alpha, \beta$
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 $\alpha, \beta$ based on historical data
- Hierarchical model:
  - $y_j \sim \text{Binomial}(n_j, \theta_j), \text{ for } j = 1, \ldots, 71$
  - $\theta_1, \ldots, \theta_{71} \sim \text{Beta}(\alpha, \beta)$
- Need to choose $\alpha, \beta$
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** \( \alpha, \beta \) **based on historical data**

- **Hierarchical model:**
  - \( y_j \sim \text{Binomial}(n_j, \theta_j), \text{ for } j = 1, \ldots, 71 \)
  - \( \theta_1, \ldots, \theta_{71} \sim \text{Beta}(\alpha, \beta) \)

- **Need to choose** \( \alpha, \beta \)
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 \( \alpha, \beta \) based on historical data

▶ Hierarchical model:
  ▶ \( y_j \sim \text{Binomial}(n_j, \theta_j), \text{ for } j = 1, \ldots, 71 \)
  ▶ \( \theta_1, \ldots, \theta_{71} \sim \text{Beta}(\alpha, \beta) \)

▶ Need to choose \( \alpha, \beta \)
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, $p(\theta)$
\( \theta_1, \ldots, \theta_J \) 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:
  \[ p(\phi|y) = \int p(\phi, \theta|y) d\theta \]
  \[ \propto p(\phi) \int p(\theta|\phi)p(y|\theta, \phi) d\theta \]
- 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:

\[ p(\phi | y) = \int p(\phi, \theta | y) d\theta \]

\[ \propto p(\phi) \int p(\theta | \phi)p(y | \theta, \phi) d\theta \]

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:
  \[ p(\phi|y) = \int p(\phi, \theta|y) d\theta \]
  \[ \propto p(\phi) \int p(\theta|\phi)p(y|\theta, \phi) d\theta \]
- If you can do the integral, computation is direct:
  - Compute \( p(\phi|y) \) on a grid of \( \phi \)
  - 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:
  \[
p(\phi|y) = \int p(\phi, \theta|y) d\theta
  \propto p(\phi) \int p(\theta|\phi)p(y|\theta, \phi)d\theta
  \]
- If you can do the integral, computation is direct:
  - Compute \( p(\phi|y) \) on a grid of \( \phi \)
  - 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:

\[
p(\phi | y) = \int p(\phi, \theta | y) d\theta
\]

\[
\propto p(\phi) \int p(\theta | \phi)p(y | \theta, \phi)d\theta
\]

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

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

\[
p(\phi | y) = \int p(\phi, \theta | y) d\theta
\]

\[
\propto p(\phi) \int p(\theta | \phi)p(y | \theta, \phi) d\theta
\]

- If you can do the integral, computation is direct:
  - Compute \( p(\phi | y) \) on a grid of \( \phi \)
  - For \( s = 1, \ldots, S \):
    - Sample \( \phi^s \) from grid
    - Sample \( \theta^s \) 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:
  \[ p(\phi | y) = \int p(\phi, \theta | y) d\theta \]
  \[ \propto p(\phi) \int p(\theta | \phi)p(y | \theta, \phi) d\theta \]
- If you can do the integral, computation is direct:
  - Compute \( p(\phi | y) \) on a grid of \( \phi \)
  - For \( s = 1, \ldots, S \):
    - Sample \( \phi^s \) from grid
    - Sample \( \theta^s \) 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:
  \[
p(\phi|y) = \int p(\phi, \theta|y)d\theta
  \]
    \[
  \propto p(\phi) \int p(\theta|\phi)p(y|\theta, \phi)d\theta
  \]
- If you can do the integral, computation is direct:
  - Compute \( p(\phi|y) \) on a grid of \( \phi \)
  - For \( s = 1, \ldots, S \):
    - Sample \( \phi^s \) from grid
    - Sample \( \theta^s \) 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 $\theta$):
  $$p(\alpha, \beta|y) \propto p(\alpha, \beta) \prod_{j=1}^{J} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha + y_j)\Gamma(\beta + n_j - y_j)}{\Gamma(\alpha + \beta + n_j)}$$
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 $\theta$):

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

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

\[
p(\alpha, \beta|y) \propto p(\alpha, \beta) \prod_{j=1}^{J} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+y_j)\Gamma(\beta+n_j-y_j)}{\Gamma(\alpha+\beta+n_j)}
\]
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 \( \theta \)):
  \[
p(\alpha, \beta|y) \propto p(\alpha, \beta) \prod_{j=1}^{J} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+y_j)\Gamma(\beta+n_j-y_j)}{\Gamma(\alpha+\beta+n_j)}
  \]
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 \( \theta \)):
  \[
p(\alpha, \beta|y) \propto p(\alpha, \beta) \prod_{j=1}^{J} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha + y_j)\Gamma(\beta + n_j - y_j)}{\Gamma(\alpha + \beta + n_j)}
  \]
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 \( \theta \)):
  \[
p(\alpha, \beta | y) \propto p(\alpha, \beta) \prod_{j=1}^{J} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} \frac{\Gamma(\alpha + y_j) \Gamma(\beta + n_j - y_j)}{\Gamma(\alpha + \beta + n_j)}
\]
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 \( \theta \)):
  \[
p(\alpha, \beta|y) \propto p(\alpha, \beta) \prod_{j=1}^{J} \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+y_j)\Gamma(\beta+n_j-y_j)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta+n_j)}
\]
Rat tumor model: prior distribution on \((\alpha, \beta)\)

- \(p(\theta|\alpha, \beta)\) 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 \((\alpha, \beta)\)

- \(p(\theta|\alpha, \beta)\) 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 \((\alpha, \beta)\)

- \(p(\theta|\alpha, \beta)\) 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 \((\alpha, \beta)\)

- \(p(\theta|\alpha, \beta)\) 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 \((\alpha, \beta)\)

- \(p(\theta|\alpha, \beta)\) 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 \((\alpha, \beta)\)

- \(p(\theta|\alpha, \beta)\) already set
- \(p(\alpha, \beta) = ?\)
- Reparameterize to \(\text{logit}(\frac{\alpha}{\alpha+\beta}) = \log(\frac{\alpha}{\beta}) \text{ and } \log(\alpha+\beta)\)
- 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 \((\alpha, \beta)\)

- \(p(\theta|\alpha, \beta)\) 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 \((\alpha, \beta)\)

- \(p(\theta|\alpha, \beta)\) 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 \((\alpha, \beta)\)

- \(p(\theta|\alpha, \beta)\) 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 \((\alpha, \beta)\)

- \(p(\theta|\alpha, \beta)\) 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

![Contour plot](image)

![Scatter plot](image)
Rat tumor model: partial pooling

95% posterior interval for theta (i)

observed rate, y(i) / N(i)
5.4. Exchangeable parameters from a normal model

- The model:
  - \( y_j \sim N(\theta_j, \sigma^2_j) \)
  - \( \theta_j \sim N(\mu, \tau^2) \)
  - What are the assumptions?

- Conditional posterior density:
  \[
  \theta | \mu, \tau, y \sim N\left( \frac{1}{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}, \frac{1}{\frac{1}{\sigma^2} + \frac{1}{\tau^2}} \right)
  \]

- Average over marginal posterior density of \( \mu, \tau \)

- Problems with simple point estimates of \( \mu, \tau \)
5.4. Exchangeable parameters from a normal model

The model:

- $\bar{y}_j \sim N(\theta_j, \sigma_j^2)$
- $\theta_j \sim N(\mu, \tau^2)$
- What are the assumptions?

Conditional posterior density:

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

- Average over marginal posterior density of $\mu, \tau$
- Problems with simple point estimates of $\mu, \tau$
5.4. Exchangeable parameters from a normal model

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

- Conditional posterior density:
  $$\theta | \mu, \tau, y \sim N \left( \frac{1}{\bar{y}_1 + \cdots + \bar{y}_n}, \frac{1}{\tau^2} \right)$$

- Average over marginal posterior density of $\mu, \tau$

- Problems with simple point estimates of $\mu, \tau$
5.4. Exchangeable parameters from a normal model

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

- Conditional posterior density:

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

- Partial pooling (shrinkage) determined by $\tau$
- Average over marginal posterior density of $\mu, \tau$
- Problems with simple point estimates of $\mu, \tau$
5.4. Exchangeable parameters from a normal model

The model:

- $\bar{y}_j \sim N(\theta_j, \sigma_j^2)$
- $\theta_j \sim N(\mu, \tau^2)$
- What are the assumptions?

Conditional posterior density:

$$
\theta|\mu, \tau, y \sim N \left( \frac{1}{\frac{1}{\tau} + \frac{1}{\sigma_j^2}} \cdot \frac{1}{\frac{1}{\tau} + \frac{1}{\sigma_j^2}}, \frac{1}{\frac{1}{\tau} + \frac{1}{\sigma_j^2}} \right)
$$

- Partial pooling (shrinkage) determined by $\tau$
- Average over marginal posterior density of $\mu, \tau$
- Problems with simple point estimates of $\mu, \tau$
5.4. Exchangeable parameters from a normal model

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

- Conditional posterior density:

$$\theta | \mu, \tau, y \sim N \left( \frac{1}{\frac{1}{\bar{y}_j} + \frac{1}{\sigma^2_j}} + \frac{1}{\frac{1}{\mu} + \frac{1}{\tau^2}}, \frac{1}{\frac{1}{\bar{y}_j} + \frac{1}{\sigma^2_j}} \right)$$

- Partial pooling (shrinkage) determined by $\tau$
- Average over marginal posterior density of $\mu, \tau$
- Problems with simple point estimates of $\mu, \tau$
5.4. Exchangeable parameters from a normal model

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

- Conditional posterior density:

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

- Partial pooling (shrinkage) determined by $\tau$
- Average over marginal posterior density of $\mu, \tau$
- Problems with simple point estimates of $\mu, \tau$
5.4. Exchangeable parameters from a normal model

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

\[
\theta | \mu, \tau, y \sim N \left( \frac{1}{\frac{1}{\mu} + \frac{1}{\tau}} + \frac{1}{\frac{1}{\sigma_j^2} + \frac{1}{\tau}}, \frac{1}{\frac{1}{\mu} + \frac{1}{\tau}} \right)
\]

- Partial pooling (shrinkage) determined by \( \tau \)
- Average over marginal posterior density of \( \mu, \tau \)
- Problems with simple point estimates of \( \mu, \tau \)
5.4. Exchangeable parameters from a normal model

- The model:
  - $\bar{y}_j \sim N(\theta_j, \sigma^2_j)$
  - $\theta_j \sim N(\mu, \tau^2)$
  - What are the assumptions?
- Conditional posterior density:
  
  $$
  \theta | \mu, \tau, y \sim N \left( \frac{1}{\ldots + \ldots} \cdot \frac{1}{\ldots + \ldots}, \frac{1}{\ldots + \ldots} \right)
  $$

- Partial pooling (shrinkage) determined by $\tau$
- Average over marginal posterior density of $\mu, \tau$
- Problems with simple point estimates of $\mu, \tau$
5.5. Example: parallel experiments in eight schools

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

<table>
<thead>
<tr>
<th>School</th>
<th>Estimated treatment effect, $y_j$</th>
<th>Standard error of effect estimate, $\sigma_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>28</td>
<td>15</td>
</tr>
<tr>
<td>B</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>C</td>
<td>-3</td>
<td>16</td>
</tr>
<tr>
<td>D</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>E</td>
<td>-1</td>
<td>9</td>
</tr>
<tr>
<td>F</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>G</td>
<td>18</td>
<td>10</td>
</tr>
<tr>
<td>H</td>
<td>12</td>
<td>18</td>
</tr>
</tbody>
</table>

- 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:

<table>
<thead>
<tr>
<th>School</th>
<th>Estimated treatment effect, $y_j$</th>
<th>Standard error of effect estimate, $\sigma_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>28</td>
<td>15</td>
</tr>
<tr>
<td>B</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>C</td>
<td>−3</td>
<td>16</td>
</tr>
<tr>
<td>D</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>E</td>
<td>−1</td>
<td>9</td>
</tr>
<tr>
<td>F</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>G</td>
<td>18</td>
<td>10</td>
</tr>
<tr>
<td>H</td>
<td>12</td>
<td>18</td>
</tr>
</tbody>
</table>

- 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:

<table>
<thead>
<tr>
<th>School</th>
<th>Estimated treatment effect, $y_j$</th>
<th>Standard error of effect estimate, $\sigma_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>28</td>
<td>15</td>
</tr>
<tr>
<td>B</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>C</td>
<td>-3</td>
<td>16</td>
</tr>
<tr>
<td>D</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>E</td>
<td>-1</td>
<td>9</td>
</tr>
<tr>
<td>F</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>G</td>
<td>18</td>
<td>10</td>
</tr>
<tr>
<td>H</td>
<td>12</td>
<td>18</td>
</tr>
</tbody>
</table>

- 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:

<table>
<thead>
<tr>
<th>School</th>
<th>Estimated treatment effect, $y_j$</th>
<th>Standard error of effect estimate, $\sigma_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>28</td>
<td>15</td>
</tr>
<tr>
<td>B</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>C</td>
<td>−3</td>
<td>16</td>
</tr>
<tr>
<td>D</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>E</td>
<td>−1</td>
<td>9</td>
</tr>
<tr>
<td>F</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>G</td>
<td>18</td>
<td>10</td>
</tr>
<tr>
<td>H</td>
<td>12</td>
<td>18</td>
</tr>
</tbody>
</table>

- 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:

<table>
<thead>
<tr>
<th>School</th>
<th>Estimated treatment effect, $y_j$</th>
<th>Standard error of effect estimate, $\sigma_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>28</td>
<td>15</td>
</tr>
<tr>
<td>B</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>C</td>
<td>-3</td>
<td>16</td>
</tr>
<tr>
<td>D</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>E</td>
<td>-1</td>
<td>9</td>
</tr>
<tr>
<td>F</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>G</td>
<td>18</td>
<td>10</td>
</tr>
<tr>
<td>H</td>
<td>12</td>
<td>18</td>
</tr>
</tbody>
</table>

- 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