Revisit Sampling

- Why sample?
  - Draw samples to approximate a distribution
  - Intractable pdfs (infinite summation in integrals at high dimensions/ The integral has no closed-form solution)

Revisit Sampling

- Why sample?
  - Draw samples to approximate a distribution
  - Intractable pdfs (infinite summation in integrals at high dimensions/ The integral has no closed-form solution)
- Proposal Distribution in Metropolis-Hasting Sampling
  - acceptance ratio includes ratio of proposal distributions

$$r = \frac{p(\theta^*|y)/J_t(\theta^*|\theta^{t-1})}{p(\theta^{t-1}|y)/J_t(\theta^{t-1}|\theta^*)} = \frac{p(\theta^*|y)J_t(\theta^{t-1}|\theta^*)}{p(\theta^{t-1}|y)J_t(\theta^*|\theta^{t-1})}$$

  - $$\theta^t = \begin{cases} \theta^* & \text{with probability } \min(r, 1) \\ \theta^{t-1} & \text{otherwise} \end{cases}$$

Hierarchical Model

- Exchangeability
- Bayesian analysis of hierarchical models
- Hierarchical normal model
- Weakly informative priors for hierarchical variance parameters

Hierarchical model

- Example: Treatment effectiveness
  - in hospital j the survival probability is θ_j
  - observations y_{ij} tell whether patient i survived in hospital j

$$
\begin{tikzcd}
\theta_1 \arrow[d] & \theta_2 \arrow[d] & \cdots & \theta_n \arrow[d] \\
y_{i1} & y_{i2} & & y_{in}
\end{tikzcd}
$$

Hierarchical model

- Example: Treatment effectiveness
  - in hospital j the survival probability is θ_j
  - observations y_{ij} tell whether patient i survived in hospital j

$$\begin{tikzcd}
\theta_1 \arrow[d] & \theta_2 \arrow[d] & \cdots & \theta_n \arrow[d] \\
y_{i1} & y_{i2} & & y_{in}
\end{tikzcd}$$

  - sensible to assume that θ_j are similar

$$\begin{tikzcd}
& \tau \arrow[dl] \arrow[d] \arrow[drr] & & \\
\theta_1 \arrow[d] & \theta_2 \arrow[d] & \cdots & \theta_n \arrow[d] \\
y_{i1} & y_{i2} & & y_{in}
\end{tikzcd}$$

  - natural to think that θ_j have common population distribution
  - θ_j is not directly observed and the population distribution is unknown

Compare

- "Separate model" (model with separate/independent effects)

$$
\begin{tikzcd}
\theta_1 \arrow[d] & \theta_2 \arrow[d] & \cdots & \theta_n \arrow[d] \\
y_1 & y_2 & & y_n
\end{tikzcd}
$$

Compare

- ”Separate model” (model with separate/independent effects)

$$\begin{tikzcd} \theta_1 \arrow[d] & \theta_2 \arrow[d] & \cdots & \theta_n \arrow[d] \\ y_1 & y_2 & & y_n \end{tikzcd}$$

- ”Joint model” (model with a common effect / pooled model)

$$\begin{tikzcd} & \theta \arrow[dl] \arrow[d] \arrow[drr] & & \\ y_1 & y_2 & \cdots & y_n \end{tikzcd}$$

Compare

- "Separate model" (model with separate/independent effects)

$$\begin{tikzcd} \theta_1 \arrow[d] & \theta_2 \arrow[d] & \cdots & \theta_n \arrow[d] \\ y_1 & y_2 & & y_n \end{tikzcd}$$

- "Joint model" (model with a common effect / pooled model)

$$\begin{tikzcd} & \theta \arrow[dl] \arrow[d] \arrow[drr] & & \\ y_1 & y_2 & \cdots & y_n \end{tikzcd}$$

- Hierarchical model

$$\begin{tikzcd} & \tau \arrow[dl] \arrow[d] \arrow[drr] & & \\ \theta_1 \arrow[d] & \theta_2 \arrow[d] & \cdots & \theta_n \arrow[d] \\ y_1 & y_2 & & y_n \end{tikzcd}$$

Hierarchical model

Level 1: observations given parameters $p(y_{ij}|\theta_j)$

$$\begin{tikzcd}
p(\theta_j|\tau) & \theta_1 \arrow[d] & \theta_2 \arrow[d] & \cdots & \theta_n \arrow[d] & \text{parameters} \\
p(y_{ij}|\theta_j) & y_{i1} & y_{i2} & & y_{in} & \text{observations}
\end{tikzcd}$$

Joint posterior

$$\begin{aligned}
p(\theta, \tau|y) &\propto p(y|\theta, \tau)p(\theta, \tau) \\
&\propto p(y|\theta)p(\theta|\tau)p(\tau)
\end{aligned}$$

Hierarchical model

Level 1: observations given parameters $p(y_{ij}|\theta_j)$

Level 2: parameters given hyperparameters $p(\theta_j|\tau)$

$$\begin{tikzcd}
p(\tau) & & \tau \arrow[dl] \arrow[d] \arrow[drr] & & & \text{hyperparameter} \\
p(\theta_j|\tau) & \theta_1 \arrow[d] & \theta_2 \arrow[d] & \cdots & \theta_n \arrow[d] & \text{parameters} \\
p(y_{ij}|\theta_j) & y_{i1} & y_{i2} & & y_{in} & \text{observations}
\end{tikzcd}$$

Joint posterior

$$\begin{aligned} p(\theta, \tau|y) &\propto p(y|\theta, \tau)p(\theta, \tau) \\ &\propto p(y|\theta)p(\theta|\tau)p(\tau) \end{aligned}$$

Hierarchical binomial model: rats

- Medicine testing
- Type F344 female rats in control group
    - count how many get endometrial stromal polyps (a kind of tumor)
    - Binomial model for number of tumors, given θ

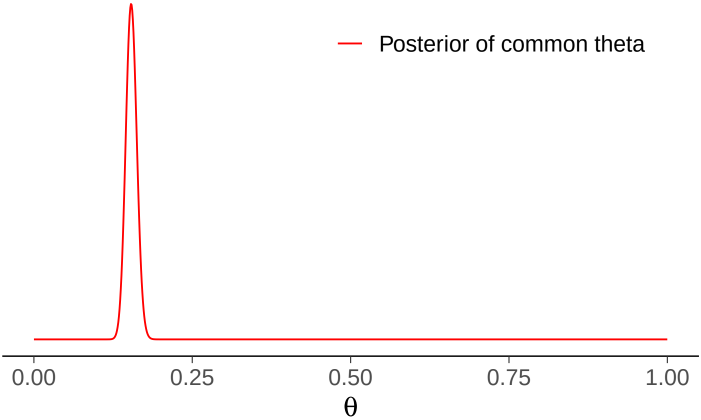
Hierarchical binomial model: rats

- Medicine testing
- Type F344 female rats in control group
  - count how many get endometrial stromal polyps (a kind of tumor)
  - Binomial model for number of tumors, given θ
- Experiment has been repeated 71 times

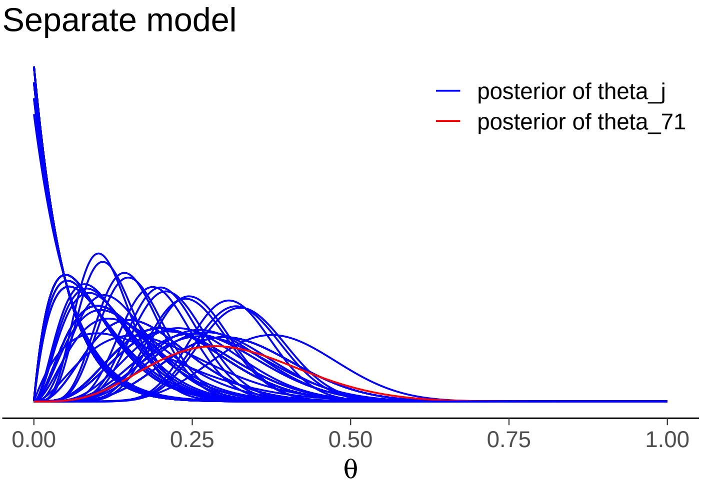
0/20	0/20	0/20	0/20	0/20	0/20	0/20	0/19	0/19	0/19
0/19	0/18	0/18	0/17	1/20	1/20	1/20	1/20	1/19	1/19
1/18	1/18	2/25	2/24	2/23	2/20	2/20	2/20	2/20	2/20
2/20	1/10	5/49	2/19	5/46	3/27	2/17	7/49	7/47	3/20
3/20	2/13	9/48	10/50	4/20	4/20	4/20	4/20	4/20	4/20
4/20	10/48	4/19	4/19	4/19	5/22	11/46	12/49	5/20	5/20
6/23	5/19	6/22	6/20	6/20	6/20	16/52	15/46	15/47	9/24
4/14									

Hierarchical binomial model: rats

Pooled model



Hierarchical binomial model: rats

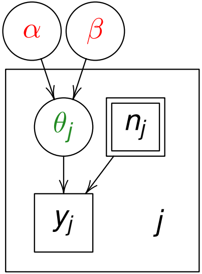


Hierarchical binomial model: rats

- Hierarchical binomial model for rats
  prior parameters α and β are unknown

$$\theta_j|\alpha,\beta \sim \text{Beta}(\theta_j|\alpha,\beta)$$

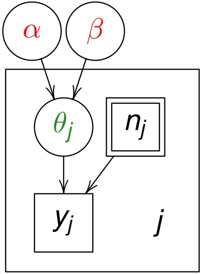
$$y_j|n_j,\theta_j \sim \text{Bin}(y_j|n_j,\theta_j)$$



- Joint posterior $p(\theta_1,\ldots,\theta_J,\alpha,\beta|y)$
  - multiple parameters

Hierarchical binomial model: rats

- Hierarchical binomial model for rats
  prior parameters α and β are unknown



$$\theta_j|\alpha,\beta \sim \mathrm{Beta}(\theta_j|\alpha,\beta)$$

$$y_j|n_j,\theta_j \sim \mathrm{Bin}(y_j|n_j,\theta_j)$$

- Joint posterior $p(\theta_1,\ldots,\theta_J,\alpha,\beta|y)$
  - multiple parameters
  - factorize $\prod_{j=1}^{J} p(\theta_j|\alpha,\beta,y)p(\alpha,\beta|y)$

Hierarchical binomial model: rats

- Write the joint posterior density, $p(\alpha, \beta, \theta | y)$, in unnormalized form as a product of the hyperprior distribution $p(\alpha, \beta)$, the population distribution $p(\theta | \alpha, \beta)$, and the likelihood $p(y | \theta)$.

Hierarchical binomial model: rats

- Write the joint posterior density, $p(\alpha, \beta, \theta | y)$, in unnormalized form as a product of the hyperprior distribution $p(\alpha, \beta)$, the population distribution $p(\theta | \alpha, \beta)$, and the likelihood $p(y | \theta)$.
- Determine analytically the conditional posterior density of θ given the hyperparameters α, β; $p(\theta | \alpha, \beta, y)$.

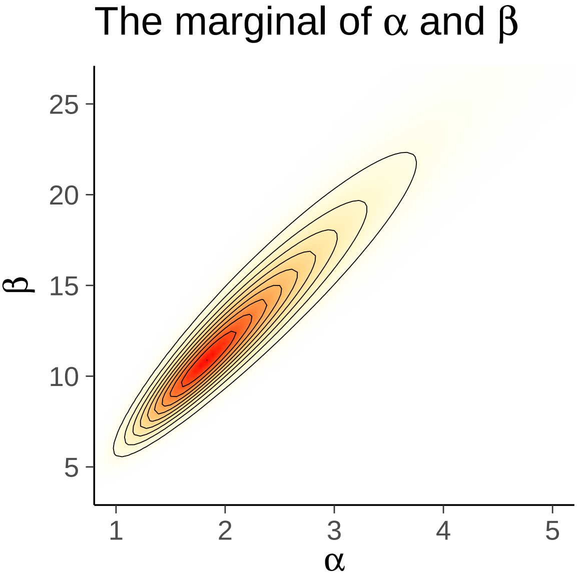
Hierarchical binomial model: rats

- Write the joint posterior density, $p(\alpha, \beta, \theta | y)$, in unnormalized form as a product of the hyperprior distribution $p(\alpha, \beta)$, the population distribution $p(\theta | \alpha, \beta)$, and the likelihood $p(y | \theta)$.
- Determine analytically the conditional posterior density of θ given the hyperparameters α, β; $p(\theta | \alpha, \beta, y)$.
- Estimate α, β using the Bayesian paradigm; that is, obtain its marginal posterior distribution, $p(\alpha, \beta | y)$

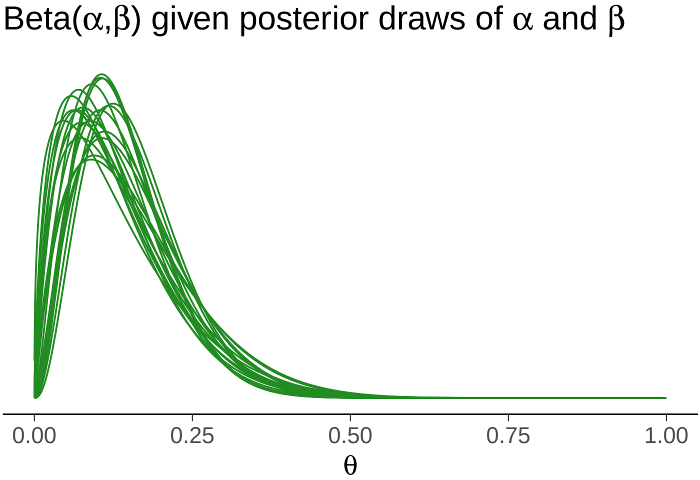
Hierarchical binomial model: rats

- Hyperprior $p(\alpha, \beta)$?
  - α, β both affect the location and scale
  - BDA3 has $p(\alpha, \beta) \propto (\alpha + \beta)^{-5/2}$
    - diffuse prior for location and scale (BDA3 p. 110)

Hierarchical binomial model: rats

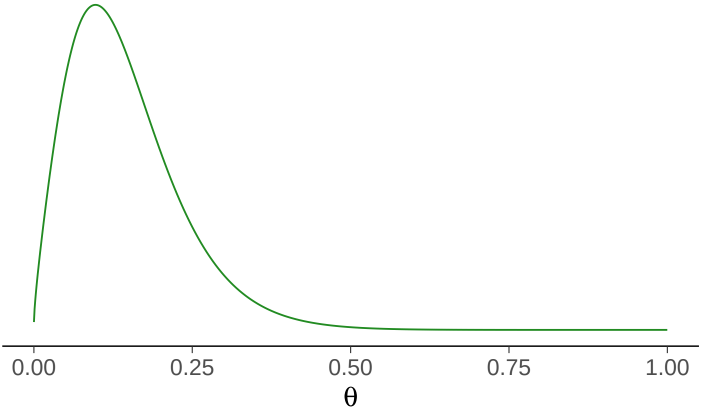


Hierarchical binomial model: rats



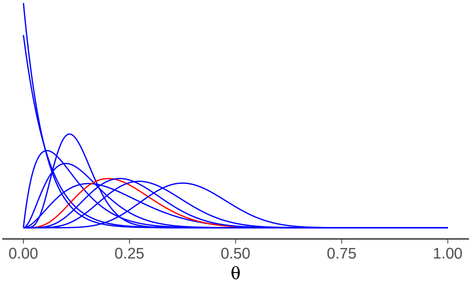
Hierarchical binomial model: rats

Population distribution (prior) for θ_j



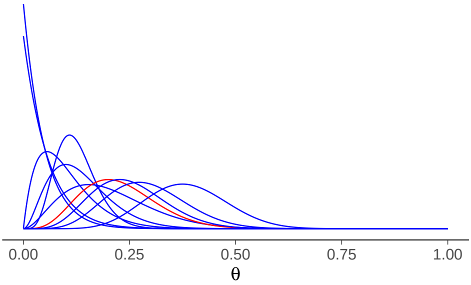
Hierarchical binomial model: rats

Separate model

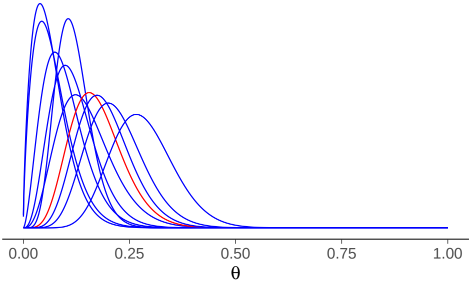


Hierarchical binomial model: rats

Separate model

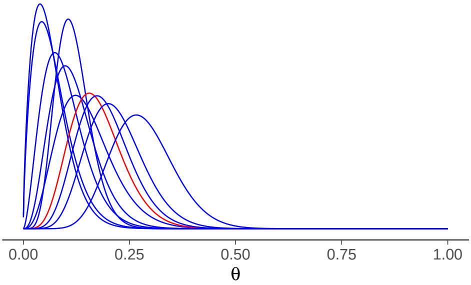


Hierarchical model

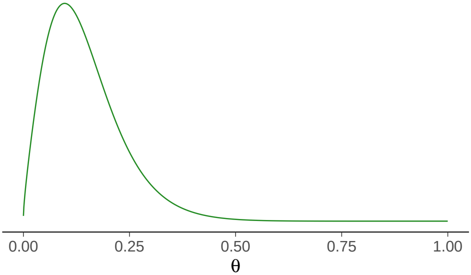


Hierarchical binomial model: rats

Hierarchical model



Population distribution (prior) for θ_j

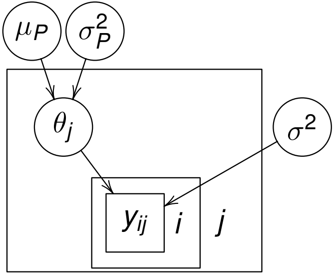


Hierarchical normal model: factory

- Factory has 6 machines which quality is evaluated
- Assume hierarchical model
  - each machine has its own (average) quality θ_j and common variance σ^2

$$\theta_j|\mu_P,\sigma_P^2 \sim \mathrm{N}(\mu_P,\sigma_P^2)$$

$$y_{ij}|\theta_j \sim \mathrm{N}(\theta_j,\sigma^2)$$



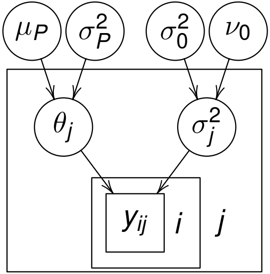
- Can be used to predict the future quality produced by each machine and quality produced by a new similar machine

Hierarchical normal model: factory

- Factory has 6 machines which quality is evaluated
- Assume hierarchical model
  - each machine has its own (average) quality θ_j and own variance σ_j^2

$$\theta_j | \mu_P, \sigma_P^2 \sim \mathrm{N}(\mu_P, \sigma_P^2)$$

$$y_{ij} | \theta_j \sim \mathrm{N}(\theta_j, \sigma_j^2)$$



$$\sigma_j^2 | \sigma_0^2, \nu_0 \sim \text{Inv-}\chi^2(\sigma_0^2, \nu_0)$$

- Can be used to predict the future quality produced by each machine and quality produced by a new similar machine

Hierarchical normal model: 8 schools

- Example: SAT coaching effectiveness
  - in USA commonly used Scholastic Aptitude Test (SAT) is designed so that short term practice should not improve the results significantly
  - schools have coaching courses
  - test the effectiveness of the coaching courses
- SAT
  - standardized multiple choice test
  - mean about 500 and standard deviation about 100
  - scores between 200 and 800
  - different topics, e.g., V=Verbal, M=Mathematics
  - pre-test PSAT

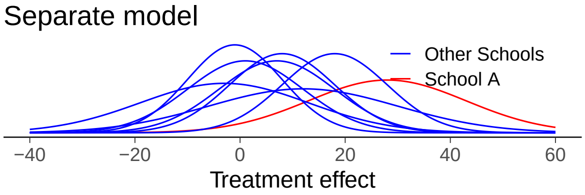
Hierarchical normal model: 8 schools

- Effectiveness of the SAT coaching
  - students had made pre-tests PSAT-M and PSAT-V
  - part of students were coached
  - linear regression was used to estimate the coaching effect y_j for the school j (could be denoted with $\bar{y}_{.j}$, too) and variances σ_j^2
  - y_j approximately normally distributed, with variances assumed to be known based on about 30 students per school
  - data is group means and variances (not personal results)
- Data:

School	A	B	C	D	E	F	G	H
y_j	28	8	-3	7	-1	1	18	12
σ_j	15	10	16	11	9	22	20	28

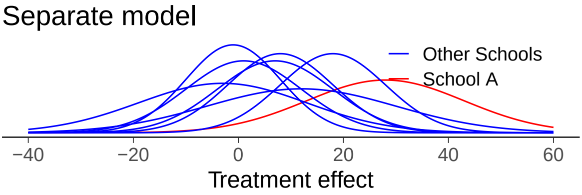
Hierarchical normal model: 8 schools

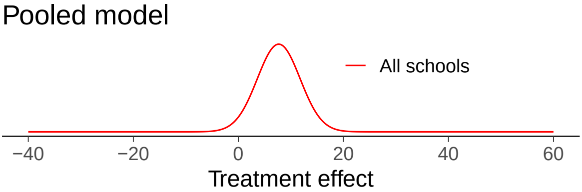
Separate or Pooled Model?



Hierarchical normal model: 8 schools

Separate or Pooled Model?





Hierarchical normal model for group means

- J experiments, unknown θ_j and known σ^2

$$y_{ij}|\theta_j \sim \mathrm{N}(\theta_j, \sigma^2), \quad i = 1, \ldots, n_j; \quad j = 1, \ldots, J$$

- Group j sample mean and sample variance

$$\begin{aligned} \bar{y}_{.j} &= \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij} \\ \sigma_j^2 &= \frac{\sigma^2}{n_j} \end{aligned}$$

- Use model

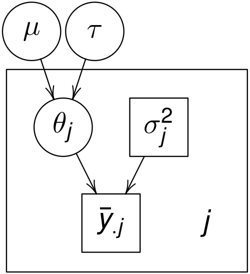
$$\bar{y}_{.j}|\theta_j \sim \mathrm{N}(\theta_j, \sigma_j^2)$$

this model can be generalized so that, σ_j^2 can be different from each other for other reasons than n_j

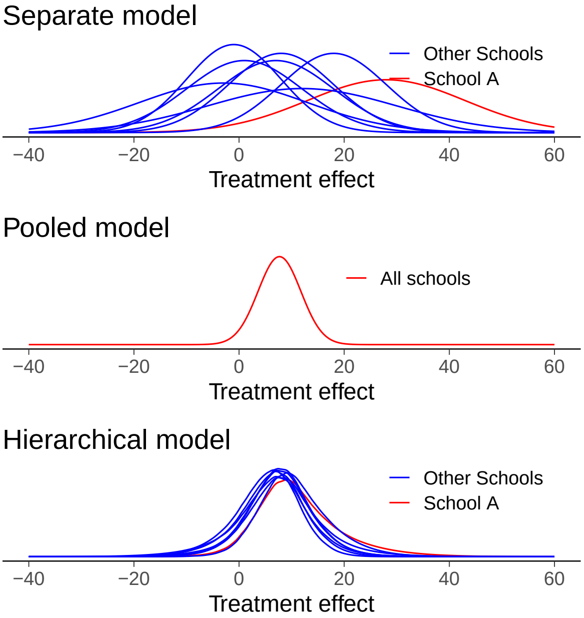
Hierarchical normal model for group means

$\theta_j|\mu, \tau \sim \mathrm{N}(\mu, \tau)$

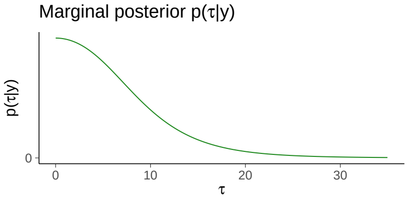
$\bar{y}_{.j}|\theta_j \sim \mathrm{N}(\theta_j, \sigma_j^2)$



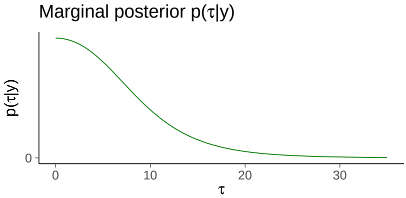
Hierarchical normal model: 8 schools

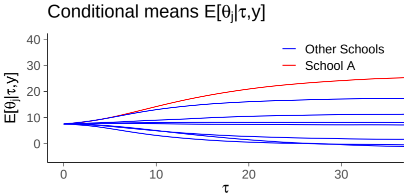


Hierarchical normal model: 8 schools



Hierarchical normal model: 8 schools





Hierarchical normal model: 8 schools