



UPPSALA  
UNIVERSITET

- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

# Bayesian Statistics and Data Analysis

## Lecture 7

Måns Magnusson

Department of Statistics, Uppsala University  
Thanks to Aki Vehtari, Aalto University



UPPSALA  
UNIVERSITET

- **Hierarchical models**
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

## Section 1

# Hierarchical models



- Hierarchical models

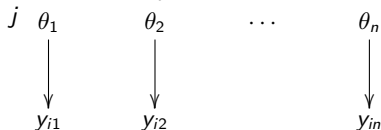
- Rats example
- Factory example
- 8 schools example

- Exchangeability

# Hierarchical model

- Example: Treatment effectiveness

- in hospital  $j$  the survival probability is  $\theta_j$
- observations  $y_{ij}$  tell whether patient  $i$  survived in hospital





- Hierarchical models

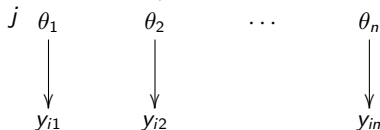
- Rats example
- Factory example
- 8 schools example

- Exchangeability

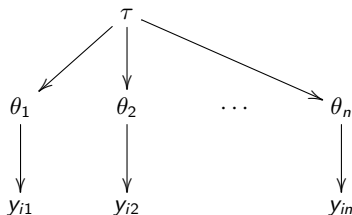
# Hierarchical model

- Example: Treatment effectiveness

- in hospital  $j$  the survival probability is  $\theta_j$
- observations  $y_{ij}$  tell whether patient  $i$  survived in hospital



- sensible to assume that  $\theta_j$  are similar



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



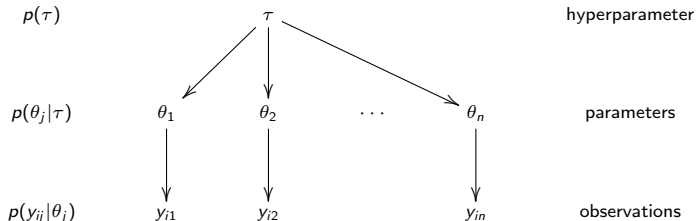
• Hierarchical models

- Rats example
- Factory example
- 8 schools example

• Exchangeability

# Hierarchical model: terms

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



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 models

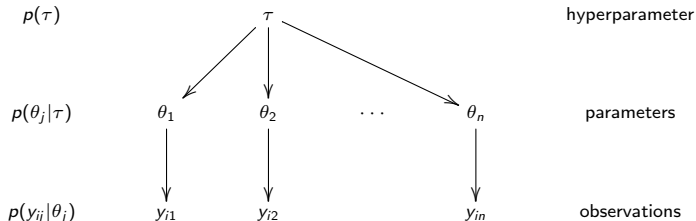
- Rats example
- Factory example
- 8 schools example

- Exchangeability

# Hierarchical model: terms

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

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



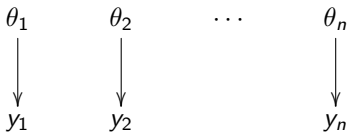
Joint posterior

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



# Comparisons

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



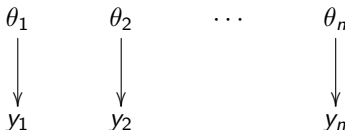
- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability



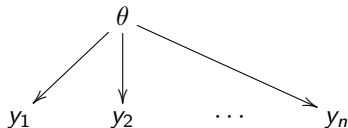
- **Hierarchical models**
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

## Comparisons

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



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

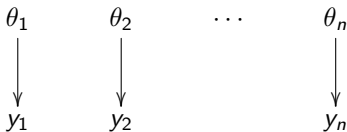




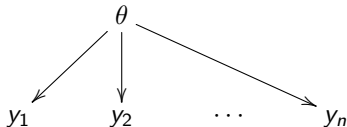


## Comparisons

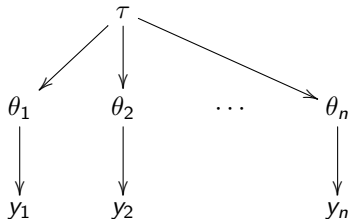
- "Separate model" (model with separate/independent effects)



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



- Hierarchical model

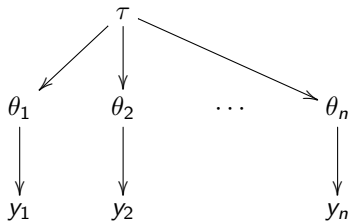




# Predictive distributions for hierarchical models

- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

- Two types of predictive distributions
  1. A new observation in **an existing group**
  2. A new observation in **a new group**





UPPSALA  
UNIVERSITET

# Hierarchical binomial model: rats

---

- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability
- Medicine testing
- Type F344 female rats in control group given placebo
  - count how many get endometrial stromal polyps
  - familiar binomial model example



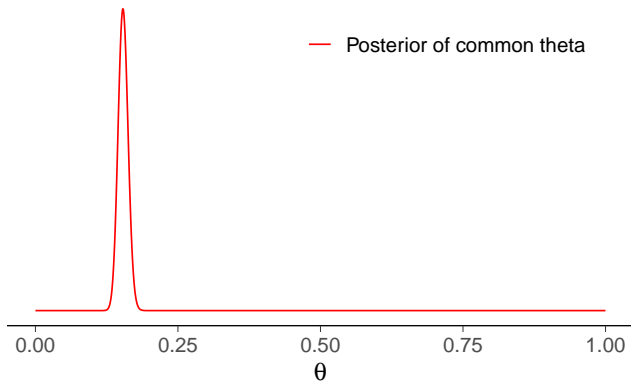
- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

- Medicine testing
- Type F344 female rats in control group given placebo
  - count how many get endometrial stromal polyps
  - familiar binomial model example
- 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									



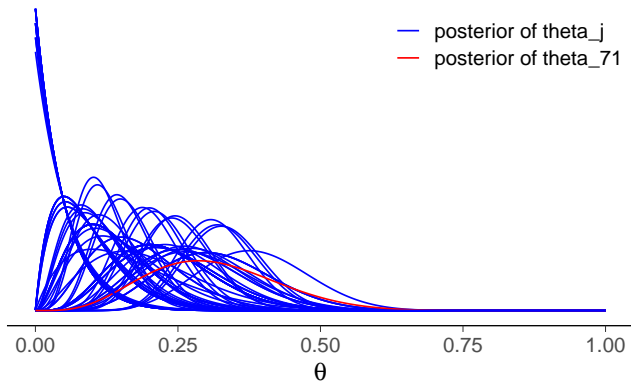
## Pooled model



- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability



## Separate model



- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability



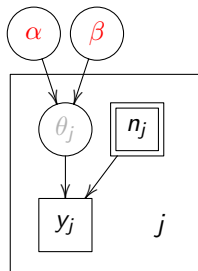
- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

# Hierarchical binomial model: rats

- Hierarchical binomial model for rats
  - prior parameters  $\alpha$  and  $\beta$  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  $p(\theta_1, \dots, \theta_J, \alpha, \beta | y)$ 
  - multiple parameters



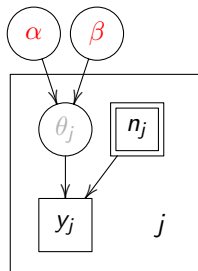
# Hierarchical binomial model: rats

- Hierarchical binomial model for rats  
prior parameters  $\alpha$  and  $\beta$  are unknown

- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

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

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



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





- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

- Population prior  $\text{Beta}(\theta_j | \alpha, \beta)$
- Hyperprior  $p(\alpha, \beta)$ ?
  - $\alpha, \beta$  both affect the location and scale
  - BDA3 (p. 110) has (vague)  $p(\alpha, \beta) \propto (\alpha + \beta)^{-5/2}$



- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

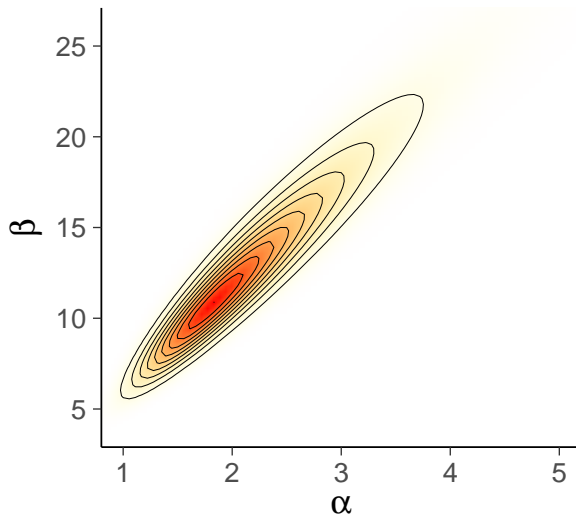
- Population prior  $\text{Beta}(\theta_j | \alpha, \beta)$
- Hyperprior  $p(\alpha, \beta)$ ?
  - $\alpha, \beta$  both affect the location and scale
  - BDA3 (p. 110) has (vague)  $p(\alpha, \beta) \propto (\alpha + \beta)^{-5/2}$
- What type of predictive distributions can we have?



- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

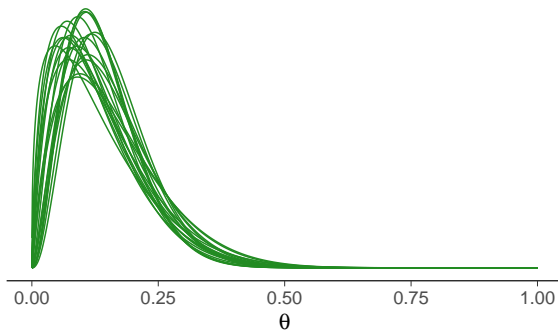
## Hierarchical binomial model: rats

### The marginal of $\alpha$ and $\beta$





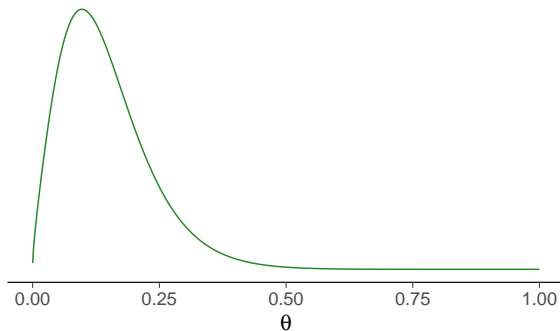
Beta( $\alpha, \beta$ ) given posterior draws of  $\alpha$  and  $\beta$



- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability



Population distribution (prior) for  $\theta_j$



- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

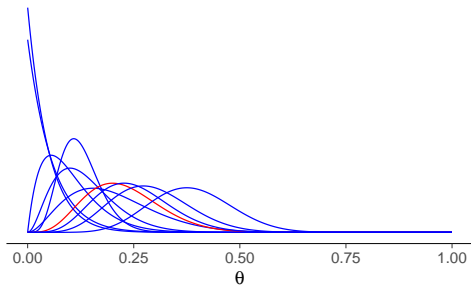


UPPSALA  
UNIVERSITET

# Hierarchical binomial model: rats

- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

## Separate model



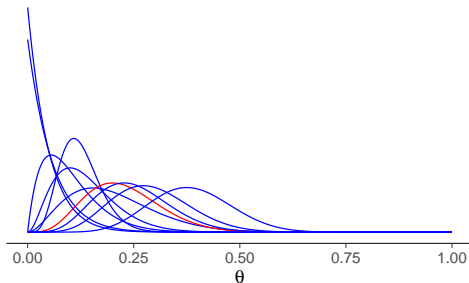


UPPSALA  
UNIVERSITET

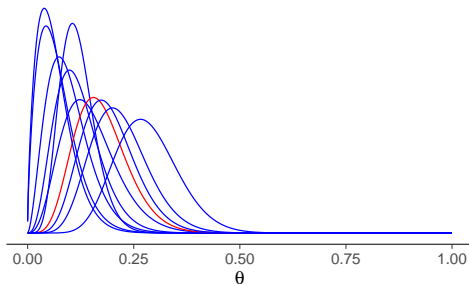
- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

# Hierarchical binomial model: rats

## Separate model



## Hierarchical model



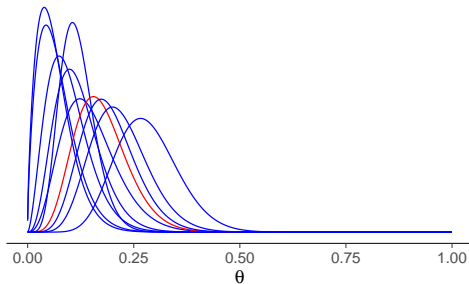


UPPSALA  
UNIVERSITET

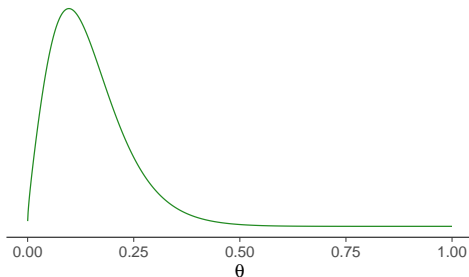
- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

# Hierarchical binomial model: rats

## Hierarchical model



Population distribution (prior) for  $\theta_j$







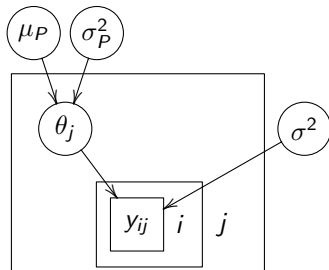
- Hierarchical models
  - Rats example
  - **Factory example**
  - 8 schools example
- Exchangeability

## Hierarchical normal model: factory

- Factory has 6 machines which quality is evaluated
- Assume hierarchical model
  - each machine has its own (average) quality  $\theta_j$  and **common variance**  $\sigma^2$

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

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



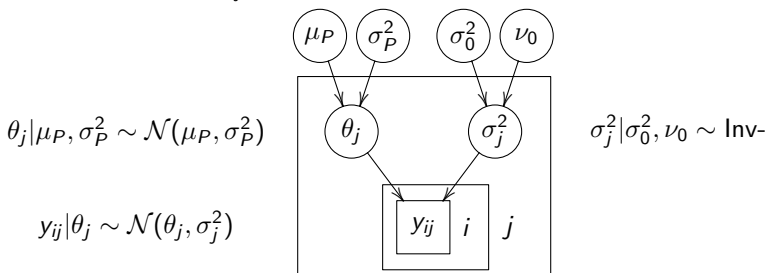
- Can be used to predict the future quality produced by each machine and quality produced by a new similar machine



- Hierarchical models
  - Rats example
  - **Factory example**
  - 8 schools example
- Exchangeability

## Hierarchical normal model: factory

- Factory has 6 machines which quality is evaluated
- Assume hierarchical model
  - each machine has its own (average) quality  $\theta_j$  and **own variance**  $\sigma_j^2$

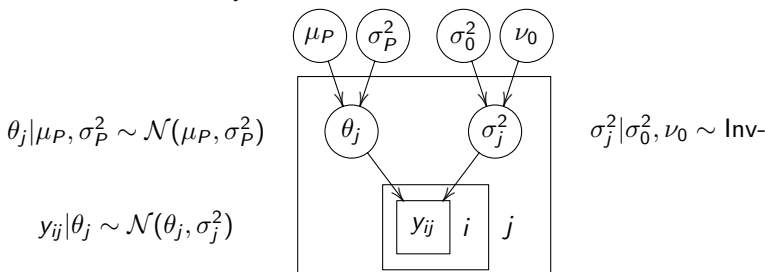




- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

## Hierarchical normal model: factory

- Factory has 6 machines which quality is evaluated
- Assume hierarchical model
  - each machine has its own (average) quality  $\theta_j$  and **own variance**  $\sigma_j^2$



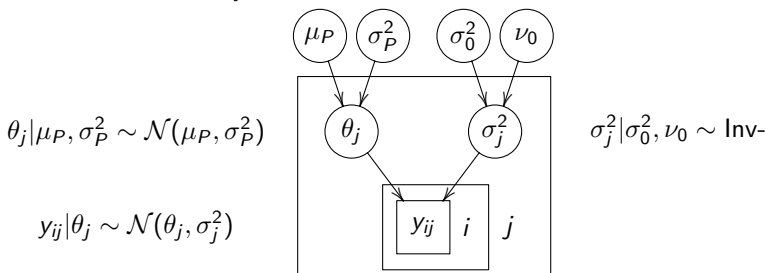
- What type of predictive distributions can we have?



- Hierarchical models
  - Rats example
  - **Factory example**
  - 8 schools example
- Exchangeability

## Hierarchical normal model: factory

- Factory has 6 machines which quality is evaluated
- Assume hierarchical model
  - each machine has its own (average) quality  $\theta_j$  and **own variance**  $\sigma_j^2$



- **What type of predictive distributions can we have?**
- Can be used to predict the future quality produced by each machine and quality produced by a new similar machine



UPPSALA  
UNIVERSITET

# Hierarchical normal model: 8 schools

---

- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

- 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 anyway coaching courses
  - test the effectiveness of the coaching courses



- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

- 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 anyway coaching courses
  - test the effectiveness of the coaching courses
- SAT
  - standardized multiple choice test
  - mean about 500 and standard deviation about 100
  - most scores between 200 and 800
  - different topics, e.g., V=Verbal, M=Mathematics
  - pre-test PSAT



UPPSALA  
UNIVERSITET

# Hierarchical normal model: 8 schools

---

- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

- 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  $\sigma_j^2$
  - $y_j$  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)



- 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  $\sigma_j^2$
  - $y_j$  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
	$\sigma_j$	15	10	16	11	9	22	20	28





# Hierarchical normal model for group means

- $J$  experiments, unknown  $\theta_j$  and known  $\sigma^2$

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

- Group  $j$  sample mean and sample variance

$$\bar{y}_{.j} = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij}$$

$$\sigma_j^2 = \frac{\sigma^2}{n_j}$$

- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability



- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

## Hierarchical normal model for group means

- $J$  experiments, unknown  $\theta_j$  and known  $\sigma^2$

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

- Group  $j$  sample mean and sample variance

$$\bar{y}_{.j} = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij}$$

$$\sigma_j^2 = \frac{\sigma^2}{n_j}$$

- Use model

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

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

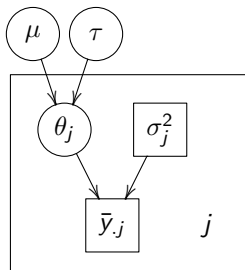


- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

# Hierarchical normal model for group means

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

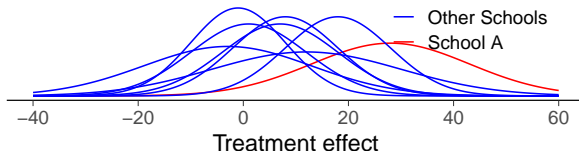
$$\bar{y}_j | \theta_j \sim \mathcal{N}(\theta_j, \sigma_j^2)$$





- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

## Separate model

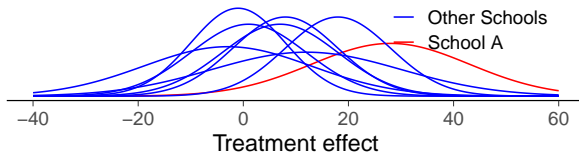




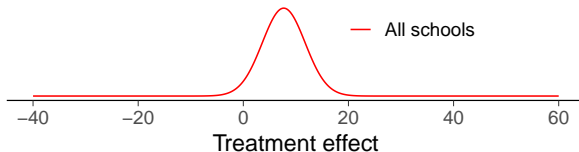
# Hierarchical normal model: 8 schools

- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

## Separate model



## Pooled model

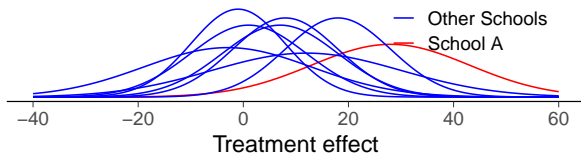




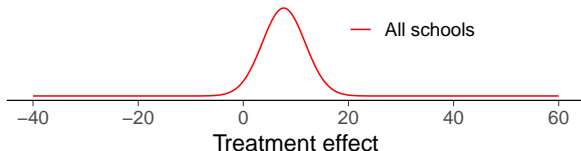
- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

# Hierarchical normal model: 8 schools

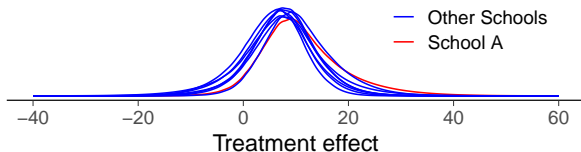
## Separate model



## Pooled model



## Hierarchical model

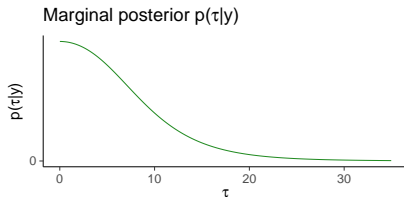




UPPSALA  
UNIVERSITET

# Hierarchical normal model: 8 schools

- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

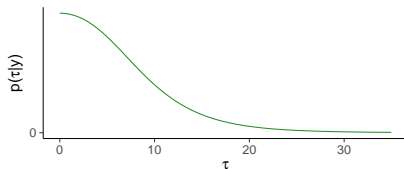




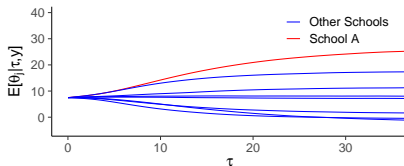
- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

# Hierarchical normal model: 8 schools

Marginal posterior  $p(\tau|y)$



Conditional means  $E[\theta_j|\tau, y]$





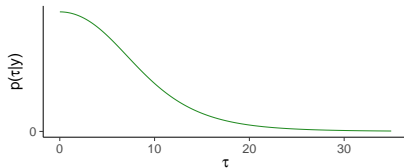


UPPSALA  
UNIVERSITET

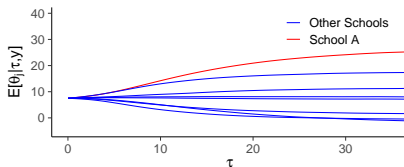
- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

# Hierarchical normal model: 8 schools

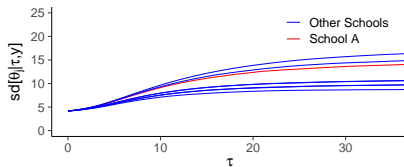
Marginal posterior  $p(\tau|y)$



Conditional means  $E[\theta_j|\tau, y]$



Conditional standard deviations  $sd[\theta_j|\tau, y]$





UPPSALA  
UNIVERSITET

# Exchangeability

---

- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

- Justifies why we can use
  - a joint model for data
  - a joint prior for a set of parameters
- Less strict than independence (IID)
- IID  $\rightarrow$  exchangeability



- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

- *Exchangeability*

Parameters  $\theta_1, \dots, \theta_J$  (or observations  $y_1, \dots, y_J$ ) are exchangeable if the joint distribution  $p$  is invariant to the permutation of indices  $(1, \dots, J)$

e.g.

$$p(\theta_1, \theta_2, \theta_3) = p(\theta_2, \theta_3, \theta_1)$$

- Exchangeability implies symmetry:  
If there is no information which can be used *a priori* to separate  $\theta_j$  from each other, we can assume exchangeability. ("Ignorance implies exchangeability")



- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

- Exchangeability does not mean that the results of the experiments could not be different
  - e.g. if we know that the experiments have been in two different laboratories, and we know that the other laboratory has better conditions for the rats, but we do not know which experiments have been made in which laboratory
  - a priori experiments are exchangeable
  - model could have unknown parameter for the laboratory with a conditional prior for rats assumed to come from the same place (clustering model)



UPPSALA  
UNIVERSITET

# Exchangeability and additional information

---

- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

- Example: bioassay
  - $y_i$  number of dead animals are not exchangeable alone



UPPSALA  
UNIVERSITET

# Exchangeability and additional information

---

- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

- Example: bioassay
  - $y_i$  number of dead animals are not exchangeable alone
  - $x_i$  dose is additional information



- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

- Example: bioassay
  - $y_i$  number of dead animals are not exchangeable alone
  - $x_i$  dose is additional information
  - $(x_i, y_i)$  exchangeable and logistic regression was used

$$p(\alpha, \beta | y, n, x) \propto \prod_{i=1}^n p(y_i | \alpha, \beta, n_i, x_i) p(\alpha, \beta)$$



UPPSALA  
UNIVERSITET

# Hierarchical exchangeability

---

- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

- Example: hierarchical rats example
  - all rats not exchangeable





UPPSALA  
UNIVERSITET

# Hierarchical exchangeability

---

- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

- Example: hierarchical rats example
  - all rats not exchangeable
  - in a single laboratory rats exchangeable



UPPSALA  
UNIVERSITET

# Hierarchical exchangeability

---

- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

- Example: hierarchical rats example
  - all rats not exchangeable
  - in a single laboratory rats exchangeable
  - laboratories exchangeable



UPPSALA  
UNIVERSITET

# Hierarchical exchangeability

---

- Hierarchical models

- Rats example
- Factory example
- 8 schools example

- Exchangeability

- Example: hierarchical rats example
  - all rats not exchangeable
  - in a single laboratory rats exchangeable
  - laboratories exchangeable
  - → hierarchical model can be used



- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

- Conditional exchangeability
  - if  $y_i$  is connected to an additional information  $x_i$ , so that  $y_i$  are not exchangeable, but  $(y_i, x_i)$  exchangeable use joint model or conditional model  $(y_i|x_i)$ .



- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

- Conditional exchangeability
  - if  $y_i$  is connected to an additional information  $x_i$ , so that  $y_i$  are not exchangeable, but  $(y_i, x_i)$  exchangeable use joint model or conditional model  $(y_i|x_i)$ .
- Partial exchangeability
  - if the observations can be grouped (a priori), then use hierarchical model



UPPSALA  
UNIVERSITET

- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

# Exchangeability

---

- The simplest form of the exchangeability (but not the only one) for the parameters  $\theta$  conditional independence

$$p(x_1, \dots, x_J | \theta) = \prod_{j=1}^J p(x_j | \theta)$$



- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

# Exchangeability

- The simplest form of the exchangeability (but not the only one) for the parameters  $\theta$  conditional independence

$$p(x_1, \dots, x_J | \theta) = \prod_{j=1}^J p(x_j | \theta)$$

- Let  $(x_n)_{n=1}^{\infty}$  to be an infinite sequence of exchangeable random variables. De Finetti's theorem then says that there is some random variable  $\theta$  so that  $x_j$  are conditionally independent given  $\theta$ , and joint density for  $x_1, \dots, x_J$  can be written in the *iid mixture* form

$$p(x_1, \dots, x_J) = \int \left[ \prod_{j=1}^J p(x_j | \theta) \right] p(\theta) d\theta$$



- Hierarchical models
  - Rats example
  - Factory example
  - 8 schools example
- Exchangeability

- A six sided die with probabilities (a finite sequence!)  $\theta_1, \dots, \theta_6$ 
  - without additional knowledge  $\theta_1, \dots, \theta_6$  exchangeable
  - due to the constraint  $\sum_{j=1}^6 \theta_j$ , parameters are not independent and thus joint distribution can not be presented as iid mixture