Hierarchical normal model: 8 schools

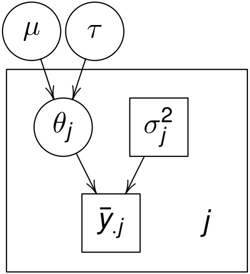
- Effectiveness of the SAT coaching
  - students had made pre-tests PSAT-M and PSAT-V
  - part of students were coached
  - linear regression was used to estimate the coaching effect y_j for the school j (could be denoted with $\bar{y}_{.j}$, too) and variances σ_j^2
  - y_j approximately normally distributed, 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 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

The necessary conditional and marginal posteriors are presented in section 5.4 of BDA. Let

$$\overline{y}_{\cdot j} = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij} \qquad \text{and} \qquad \sigma_j^2 = \sigma^2 / n_j$$

Then

$$p(\tau|y) \propto p(\tau) V_\mu^{1/2} \prod_{j=1}^{J} (\sigma_j^2 + \tau^2)^{-1/2} \exp\left(-\frac{(\overline{y}_{\cdot j} - \hat{\mu})^2}{2(\sigma_j^2 + \tau^2)}\right)$$

$$\mu|\tau, y \sim N(\hat{\mu}, V_\mu)$$

$$\theta_j|\mu, \tau, y \sim N(\hat{\theta}_j, V_j)$$

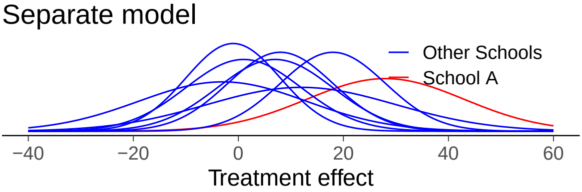
$V_\mu^{-1} = \sum_{j=1}^{J} \frac{1}{s_j^2+\tau^2} \qquad \hat{\mu} = V_\mu(\sum_{j=1}^{J} \frac{\overline{y}_{\cdot j}}{s_i^2+\tau^2})$

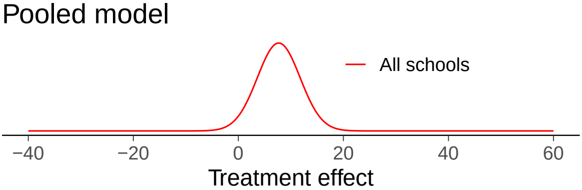
$V_j^{-1} = \frac{1}{s_i^2} + \frac{1}{\tau^2} \qquad \hat{\theta}_j = V_j(\frac{\overline{y}_{i\cdot}}{s_i^2} + \frac{\mu}{\tau^2})$

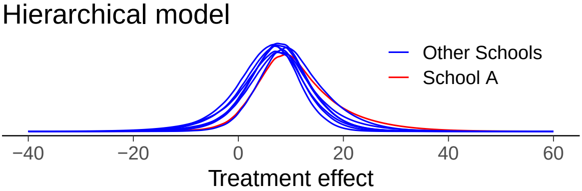
Hierarchical normal model: Computation strategy

1. $\tau^{(k)} \sim p(\tau|y)$
2. $\mu^{(k)} \sim p(\mu|\tau^{(k)}, y)$
3. $\theta_j^{(k)} \sim p(\theta|\mu^{(k)}, \tau^{(k)}, y)$ for $j = 1, \ldots, J$.

Hierarchical normal model: 8 schools

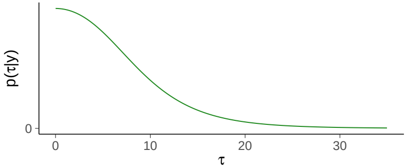




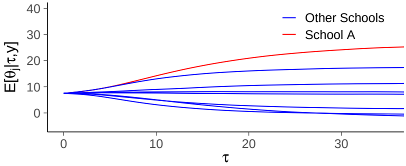


Hierarchical normal model: 8 schools

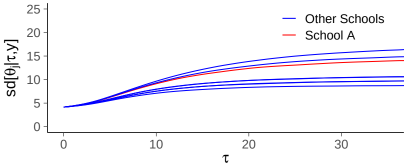
Marginal posterior p(τ|y)



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



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



Summary - hierarchical examples

- Allow the data to inform us about similarities across groups
- Provide data driven shrinkage toward a grand mean
  - lots of shrinkage when means are similar
  - little shrinkage when means are different

Decomposition used in computation

$$p(\theta, \mu, \tau|y) = p(\theta|\mu, \tau, y)p(\mu|\tau, y)p(\tau|y)$$

which allowed for simulation from τ then μ and then θ to obtain samples from the posterior.

Summary - extension to more levels

Three-level hierarchical model:

$$y \sim p(y|\theta) \qquad \theta \sim p(\theta|\phi) \qquad \phi \sim p(\phi|\psi) \qquad \psi \sim p(\psi)$$

When deriving posteriors, remember the conditional independence structure, e.g.

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

Theoretical justification for hierarchical models

- Exchangeability
  - Justifies why we can use
    - a joint model for data
    - a joint prior for a set of parameters
  - Less strict than independence
- de Finetti's theorem
- Application to hierarchical models

Exchangeability

Definition

The set $Y_1, Y_2, \ldots, Y_n$ is exchangeable if the joint probability $p(y_1, \ldots, y_n)$ is invariant to permutation of the indices. That is, for any permutation π,

$$p(y_1, \ldots, y_n) = p(y_{\pi_1}, \ldots, y_{\pi_n}).$$

Some examples:

1. A box has one black ball and one white ball. We pick a ball y1 at random, put it back, and pick another ball y2 at random.
2. A box has one black ball and one white ball. We pick a ball y1 at random, we do not put it back, then we pick ball y2.
3. A box has a million black balls and a million white balls. We pick a ball y1 at random, we do not put it back, then we pick ball y2 at random.

Exchangeability

Definition

The set $Y_1, Y_2, \ldots, Y_n$ is exchangeable if the joint probability $p(y_1, \ldots, y_n)$ is invariant to permutation of the indices. That is, for any permutation π,

$$p(y_1, \ldots, y_n) = p(y_{\pi_1}, \ldots, y_{\pi_n}).$$

An exchangeable but not iid example:

- Consider an box with one black ball and one white ball with probability 1/2 of drawing either.
- Draw without replacement from the urn.
- Let $Y_i = 1$ if the ith ball is black and otherwise $Y_i = 0$.
- Since $1/2 = P(Y_1 = 1, Y_2 = 0) = P(Y_1 = 0, Y_2 = 1) = 1/2$, Y_1 and Y_2 are exchangeable.
- But $0 = P(Y_2 = 1 | Y_1 = 1) \neq P(Y_2 = 1) = 1/2$ and thus Y_1 and Y_2 are not independent.

Exchangeability

Theorem

All independent and identically distributed random variables are exchangeable.

Proof.

Let $y_i \overset{iid}{\sim} p(y)$, then

$$p(y_1, \ldots, y_n) = \prod_{i=1}^{n} p(y_i) = \prod_{i=1}^{n} p(y_{\pi_i}) = p(y_{\pi_1}, \ldots, y_{\pi_n})$$

$\square$

Definition

The sequence $Y_1, Y_2, \ldots$ is infinitely exchangeable if, for any n, $Y_1, Y_2, \ldots, Y_n$ are exchangeable.

de Finetti's theorem

Theorem

A sequence of random variables $(y_1, y_2, \ldots)$ is infinitely exchangeable iff, for all n,

$$p(y_1, y_2, \ldots, y_n) = \int \prod_{i=1}^{n} p(y_i|\theta) P(d\theta),$$

for some measure P on θ.

If the distribution on θ has a density, we can replace $P(d\theta)$ with $p(\theta)d\theta$.

This means that there must exist

- a parameter θ,
- a likelihood $p(y|\theta)$ such that $y_i \overset{ind}{\sim} p(y|\theta)$, and
- a distribution P on θ.

Application to hierarchical models

Assume $(y_1, y_2, \ldots)$ are infinitely exchangeable, then by de Finetti's theorem for the $(y_1, \ldots, y_n)$ that you actually observed, there exists

- a parameter θ,
- a distribution $p(y|\theta)$ such that $y_i \stackrel{ind}{\sim} p(y|\theta)$, and
- a distribution P on θ.

Assume $\theta = (\theta_1, \theta_2, \ldots)$ with θ_i infinitely exchangeable. By de Finetti's theorem for $(\theta_1, \ldots, \theta_n)$, there exists

- a parameter ϕ,
- a distribution $p(\theta|\phi)$ such that $\theta_i \stackrel{ind}{\sim} p(\theta|\phi)$, and
- a distribution P on ϕ.

Assume $\phi = \phi$ with $\phi \sim p(\phi)$.

Hierarchical exchangeability

- Example: hierarchical rats example
  - all rats not exchangeable
  - in a single laboratory rats exchangeable
  - laboratories exchangeable
  - $\rightarrow$ hierarchical model

Exchangeability and additional information

- 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

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

Partial or conditional 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

Exchangeability with covariates

Suppose we observe y_i observations and x_i covariates for each unit i. Now we assume $(y_1, y_2, \ldots)$ are infinitely exchangeable given x_i, then by de Finetti's theorem for the $(y_1, \ldots, y_n)$, there exists

- a parameter θ,
- a distribution $p(y|\theta, x)$ such that $y_i \stackrel{ind}{\sim} p(y|\theta, x_i)$, and
- a distribution P on θ given x.

Assume $\theta = (\theta_1, \theta_2, \ldots)$ with θ_i infinitely exchangeable given x. By de Finetti's theorem for $(\theta_1, \ldots, \theta_n)$, there exists

- a parameter ϕ,
- a distribution $p(\theta|\phi, x)$ such that $\theta_i \stackrel{ind}{\sim} p(\theta|\phi, x_i)$, and
- a distribution P on ϕ given x.

Assume $\phi = \phi$ with $\phi \sim p(\phi|x)$.

Summary

Hierarchical model:

$$y_i \stackrel{ind}{\sim} p(y|\theta_i), \qquad \theta_i \stackrel{ind}{\sim} p(\theta|\phi), \qquad \phi \sim p(\phi)$$

Hierarchical (linear) model:

$$y_i \stackrel{ind}{\sim} p(y|\theta_i, x_i), \qquad \theta_i \stackrel{ind}{\sim} p(\theta|\phi, x_i), \qquad \phi \sim p(\phi|x)$$

Although hierarchical models are typically written using the conditional independence notation above, the assumptions underlying the model are exchangeability and functional forms for the priors.