

Hierarchical Models

PUBH 8442: Bayes Decision Theory and Data Analysis

Eric F. Lock
UMN Division of Biostatistics, SPH
elock@umn.edu

02/21/2024

Bayesian hierarchical models

- ▶ Standard model: $\mathbf{y} \sim p(\mathbf{y} | \theta_1)$, $\theta_1 \sim p(\theta_1)$.
- ▶ The joint distribution of \mathbf{y}, θ_1 is

$$p(\mathbf{y}, \theta_1) = p(\mathbf{y} | \theta_1)p(\theta_1).$$

- ▶ Typically, $p(\theta_1)$ has hyper-parameters θ_2 :

$$p(\mathbf{y}, \theta_1) = p(\mathbf{y} | \theta_1)p(\theta_1 | \theta_2)$$

- ▶ Can give prior for hyper-parameters θ_2 , $p(\theta_2 | \theta_3)$, with hyper-hyper-parameters θ_3 , etc.

$$p(\mathbf{y}, \theta_1, \theta_2, \dots, \theta_l) = p(\mathbf{y} | \theta_1)p(\theta_1 | \theta_2)p(\theta_2 | \theta_3) \cdots p(\theta_l | \eta).$$

- ▶ A Bayesian *hierarchical* model of level l .

- ▶ The marginal prior for θ_1 is

$$p(\theta_1) = \int \cdots \int p(\theta_1 | \theta_2) \cdots p(\theta_l | \eta) d\theta_2 \cdots d\theta_l$$

- ▶ If $p(\theta_1 | \theta_2), \dots, p(\theta_l | \eta)$ are proper, so is $p(\theta_1)$.
- ▶ The marginal, posterior, and posterior predictive can be derived from $p(\theta_1)$ in the standard way for $\mathbf{y} = y_1, \dots, y_n \stackrel{iid}{\sim} p(\cdot | \theta_1)$:

$$p(\mathbf{y}) = \int p(\mathbf{y} | \theta_1) p(\theta_1) d\theta_1,$$

$$p(\theta_1 | \mathbf{y}) = \frac{p(\mathbf{y} | \theta_1) p(\theta_1)}{\int p(\mathbf{y} | \theta_1) p(\theta_1) d\theta_1},$$

$$p(y_{n+1} | \mathbf{y}) = \int p(y_{n+1} | \theta_1) p(\theta_1 | \mathbf{y}) d\theta_1.$$

Example: Liver disease

- ▶ Give a collection of mice a dose of alcohol
- ▶ Observe which mice develop liver disease
- ▶ Each mouse comes from one of 50 genetic strains
- ▶ For $i = 1, \dots, 50$ let
 - ▶ n_i be the number of mice from strain i
 - ▶ y_i be the number of mice with liver disease from strain i .
- ▶ Data: <http://www.ericfrazerlock.com/MiceData.csv>

Example: Liver disease

- ▶ Assume each strain has probability θ_i of developing liver disease:

$$y_i \sim \text{Binomial}(n_i, \theta_i).$$

- ▶ Model θ_i 's as iid from Beta(a, b) distribution.

$$p(\theta_i | y_i, a, b) = \text{Beta}(a + y_i, b + n_i - y_i),$$

$$P(y_i | a, b) = \binom{n_i}{y_i} \frac{B(a + y_i, b + n_i - y_i)}{B(a, b)}$$

and

$$P(\mathbf{y} | a, b) = \prod_{i=1}^n P(y_i | a, b)$$

Example: Liver disease

- ▶ How to choose a and b ?
- ▶ One approach: put a prior on (a, b) .
- ▶ For computational simplicity, we'll use a discrete uniform prior:

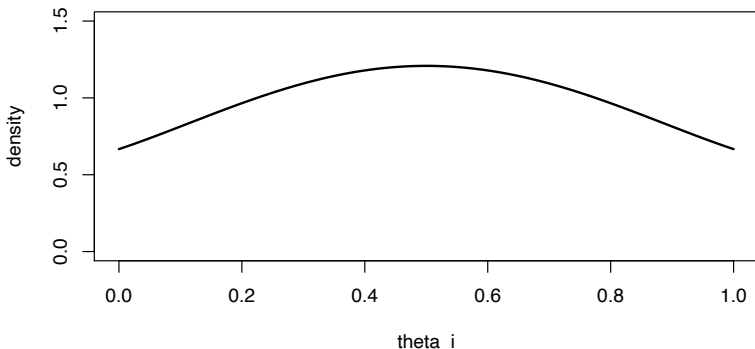
▶

	b=1	b=2	b=3
a=1	1/9	1/9	1/9
a=2	1/9	1/9	1/9
a=3	1/9	1/9	1/9

Example: Liver disease

- The marginal prior (over a,b) for θ_i is

$$p(\theta_i) = \sum_{l=1}^3 \sum_{m=1}^3 \frac{1}{9} \text{Beta}(a = l, b = m)$$



Example: Liver disease

- ▶ The posterior probability of $a = a'$, $b = b'$ is

$$\frac{P(\mathbf{y} | a = a', b = b')}{\sum_{i=1}^3 \sum_{j=1}^3 P(\mathbf{y} | a = i, b = j)}$$

- ▶ Posterior probabilities for a, b :

	b=1	b=2	b=3
a=1	0	0.14	0.86
a=2	0	0	0
a=3	0	0	0

http://www.ericfrazierlock.com/Hierarchical_Models_Rcode1.r

Example: Liver disease

- ▶ The posterior distribution for θ_i is

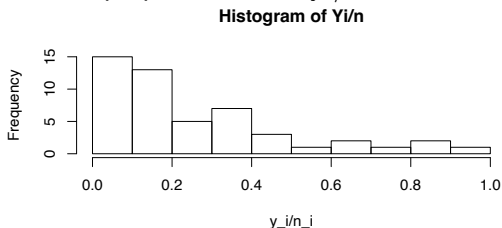
$$p(\theta_i | \mathbf{y}) = 0.14 \cdot \text{Beta}(1+y_i, 2+n_i-y_i) + 0.86 \cdot \text{Beta}(1+y_i, 3+n_i-y_i)$$

- ▶ The posterior expectation for θ_i is

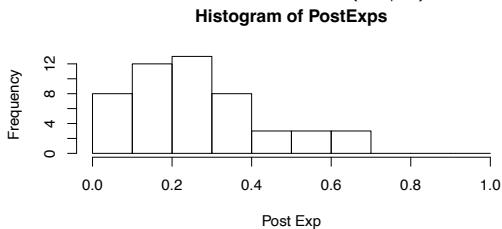
$$E(\theta_i | \mathbf{y}) = 0.14 \left(\frac{1+y_i}{3+n_i} \right) + 0.86 \left(\frac{1+y_i}{4+n_i} \right)$$

Example: Liver disease

- ▶ Histogram of raw proportions $\hat{\theta}_i = y_i/n_i$:



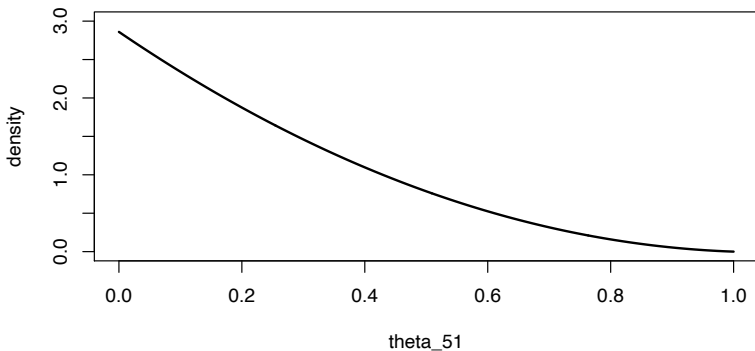
- ▶ Histogram of posterior expectations $E(\theta_i | \mathbf{y})$:



Example: Liver disease

- The posterior for the probability of liver disease in an unobserved strain, θ_{51} , is

$$p(\theta_{51} | \mathbf{y}) =$$



Example: Liver disease

- ▶ The prior predictive pmf $P(y_i = k)$ for n_i mice from strain i is

$$\binom{n_i}{k} \sum_{l=1}^3 \sum_{m=1}^3 \frac{1}{9} \cdot \frac{B(l+k, m+n_i-k)}{B(l, m)}$$

- ▶ The full marginal prior pmf $P(\mathbf{y})$ is

$$\sum_{l=1}^3 \sum_{m=1}^3 \frac{1}{9} \cdot \prod_{i=1}^{50} \binom{n_i}{y_i} \frac{B(l+y_i, m+n_i-y_i)}{B(l, m)}$$

Example: Liver disease

- ▶ For n_{51} mice from a new strain, the predictive pmf $P(y_{51} = k | \mathbf{y})$ is

$$\binom{n_{51}}{k} \left(0.14 \cdot \frac{B(1+k, 2+n_{51}-k)}{B(1,2)} + 0.86 \cdot \frac{B(1+k, 3+n_{51}-k)}{B(1,3)} \right)$$

Example: Liver disease

- ▶ Suppose a mouse's susceptibility to liver disease does not depend on genetics ($\theta_1 = \dots = \theta_{50} = \theta$)

$$y_i \sim \text{Binomial}(n_i, \theta)$$

$$\theta \sim \text{Beta}(1, 1)$$

- ▶ Call this model M_0 , previous hierarchical model M_1
- ▶ Note under M_0 $y \sim \text{Binomial}(n, \theta)$, where

$$n = \sum_{i=1}^{50} n_i, \quad y = \sum_{i=1}^{50} y_i$$

- ▶ Observe $n = 356, y = 92$.

$$p(\theta | \mathbf{y}, M_0) = \text{Beta}(93, 265)$$

Example: Liver disease

- ▶ Marginal for M_0 :

$$P(\mathbf{y} | M_0) = \left[\prod_{i=1}^{50} \binom{n_i}{y_i} \right] B(93, 265)$$

- ▶ Marginal for M_1 :

$$P(\mathbf{y} | M_1) \sum_{l=1}^3 \sum_{m=1}^3 \frac{1}{9} \cdot \prod_{i=1}^{50} \binom{n_i}{y_i} \frac{B(l + y_i, m + n_i - y_i)}{B(l, m)}$$

- ▶ The Bayes factor for M_1 over M_0 is > 10000
- ▶ Convincing evidence that the hierarchical (genetic effect) model is superior

- ▶ Assume

$$y_1, \dots, y_n \sim \text{Normal}(\theta_1, \sigma_1^2)$$

$$\theta_1 \sim \text{Normal}(\theta_2, \sigma_2^2)$$

$$\theta_2 \sim \text{Normal}(\theta_3, \sigma_3^2)$$

$$\vdots$$

$$\theta_l \sim \text{Normal}(\mu^*, \tau^2).$$

- ▶ Then, $p(\theta_1) \sim \text{Normal}(\mu^*, \sigma_2^2 + \sigma_3^2 + \dots + \sigma_l^2 + \tau^2)$

Hierarchical normal model (known variance)

- ▶ Assume $y_{ij} \sim \text{Normal}(\theta_i, \sigma^2)$ for
 - ▶ Groups $i = 1, \dots, m$
 - ▶ Observations $j = 1, \dots, n_i$ for group i
- ▶ Assume σ^2 is known, and θ_i 's are iid $\text{Normal}(\mu, \tau^2)$
- ▶ Assume τ^2 is known, and give prior $p(\mu)$ for μ .
- ▶ Note that for $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})$

$$p(\mathbf{y}_i | \theta_i, \sigma^2) \propto \text{Normal}(\bar{y}_i | \theta_i, \sigma_i^2)$$

where $\sigma_i^2 = \sigma^2/n_i$. So it suffices to consider the group means \bar{y}_i .

Hierarchical normal model (known variance)

- ▶ The joint posterior distribution is

$$\begin{aligned} p(\boldsymbol{\theta}, \mu \mid \mathbf{y}) &\propto p(\boldsymbol{\theta}, \mu, \mathbf{y}) \\ &= \prod_{i=1}^m \text{Normal}(\mathbf{y}_i \mid \theta_i, \sigma^2) \cdot \prod_{i=1}^m \text{Normal}(\theta_i \mid \mu, \tau^2) \cdot p(\mu) \\ &\propto \prod_{i=1}^m \text{Normal}(\bar{y}_i \mid \theta_i, \sigma_i^2) \cdot \prod_{i=1}^m \text{Normal}(\theta_i \mid \mu, \tau^2) \cdot p(\mu) \end{aligned}$$

Example: IQ scores

- ▶ Human IQs have variance 225 and are to be centered at 100
- ▶ Infer the “calibration” μ of a certain IQ test
- ▶ Sample of 20 individuals, each takes the test n_i times.
- ▶ Model:
 - ▶ $y_{ij} \sim \text{Normal}(\theta_i, 64)$ for $i = 1, \dots, m, j = 1, \dots, n_i$
 - ▶ $\theta_i \sim \text{Normal}(\mu, 225)$ for $i = 1, \dots, m$
 - ▶ $p(\mu) = 1$ for all μ :
- ▶ Data: <http://www.ericfrazerlock.com/IQData.csv>

- ▶ The posterior distribution for μ is

$$p(\mu | \mathbf{y}) = \text{Normal}(\hat{\mu}, V_{\mu})$$

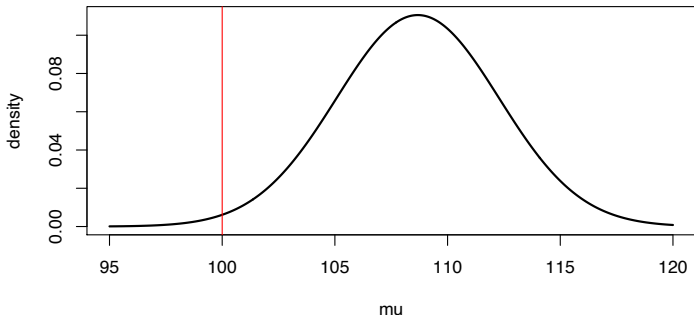
where

$$\hat{\mu} = \frac{\sum_{i=1}^m (\sigma_i^2 + \tau^2)^{-1} \bar{y}_i}{\sum_{i=1}^m (\sigma_i^2 + \tau^2)^{-1}} \quad \text{and} \quad V_{\mu} = \left[\sum_{i=1}^m (\sigma_i^2 + \tau^2)^{-1} \right]^{-1}$$

- ▶ Homework

Example: IQ scores

- Posterior for μ is Normal(108.7, 13.0):



http://www.ericfrazerlock.com/More_on_Hierarchical_Models_Rcode2.r

- $P(\mu \geq 100 | \mathbf{y}) \approx 0.99$: suggests test is too generous