Hierarchical Model

Gwangsu Kim

JBNU

Second semester of 2024

Constructing a parameterized prior distribution I

- Example: Estimating the risk of tumor in a group of rats.
  1. Suppose the immediate aim is to estimate θ, the probability of tumor in a population of female laboratory rats of type 'F344' that receive a zero dose of the drug (a control group). The data show that 4 out of 14 rats developed endometrial stromal polyps (a kind of tumor).
  2. Consider the following:

$$y|\theta \sim Bin(n,\theta), \theta \sim \mathcal{B}eta(\alpha,\beta).$$

Constructing a parameterized prior distribution II

3 If we have 70 observations from the previous experiments, $y_j/n_j,\ j = 1, \ldots, 70$ (sample mean 0.136, sample standard deviation 0.103), then we use this information to decide the α and β (1.4, 8.6). Note that we consider the models of $y_j \sim Bin(n_j, \theta_j)$.

Constructing a parameterized prior distribution III

4. Usage of this information to the inferences of θ_j $(j = 1, \ldots, 70)$ is not appropriate.

Constructing a parameterized prior distribution IV

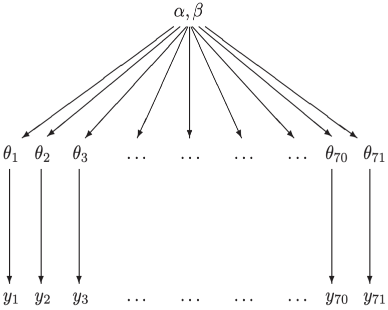

Figure 5.1: *Structure of the hierarchical model for the rat tumor example.*

Constructing a parameterized prior distribution V

- Logic of combination information
  1. All θ_j have common facts, and should be dependent in posterior.
  2. In this case, we consider the hierarchies.

Exchangeability and setting up hierarchical models I

- Exchangeability
  1. Joint distribution of $p(\theta_1, \dots, \theta_J)$ is invariant to permutations of the index $(1, \dots, J)$.
  2. Example: i.i.d., $\pi(\theta_1, \dots, \theta_J) = \int \left[\prod_{j=1}^J \pi(\theta_j|\phi)\right] \pi(\phi) d\phi$.
  3. Note that $\pi(\theta) = \int \left[\prod_{j=1}^J \pi(\theta_j|\phi)\right] \pi(\phi) d\phi$. De Finetti's theorem is related to this exchangeabilty.

Exchangeability and setting up hierarchical models II

- Exchangeability when additional information is available on the units

$$\pi(\theta_1, \dots, \theta_J | x_1, \dots, x_J) = \int \left[\prod_{j=1}^{J} \pi(\theta_j | x_j, \phi) \right] \pi(\phi | x_1, \dots, x_J) d\phi.$$

- Smaller n_j can occurs smaller θ_j.

Exchangeability and setting up hierarchical models III

- The full Bayesian treatment of the hierarchical model

$$\begin{aligned} \pi(\theta, \phi|y) &\propto \pi(\theta|\phi)\pi(\phi)p(y|\phi, \theta) \\ &= \pi(\theta|\phi)\pi(\phi)p(y|\theta). \end{aligned}$$

- Priors for ϕ, non-informative or others are considered.

Exchangeability and setting up hierarchical models IV

- Posterior predictive distribution

$$\int p(\tilde{y}|y,\theta,\phi)\pi(\theta|\phi,y)\pi(\phi|y)d\theta d\phi.$$

Computation with hierarchical model I

- Analytic derivation of conditional and marginal distributions
  1. Write down $p(y|\theta, \phi), \pi(\theta|\phi)$ and $\pi(\phi)$.
  2. Determine the conditional posterior of $\pi(\theta|\phi, y)$.
  3. Marginal posterior of $\pi(\phi|y)$ can be obtained from joint posterior.
- $\pi(\phi|y)$ can be obtained directly in some models.

Computation with hierarchical model II

- Example of Rat tumors
  1. Priors and models

$$y_j \sim B(n_j, \theta_j),\ \theta_j \sim \mathcal{B}eta(\alpha, \beta).$$

  2. Joint, conditional posteriors

$$\pi(\theta, \alpha, \beta) \propto \pi(\alpha, \beta) \prod_{j=1}^{J} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_j^{\alpha-1}(1-\theta_j)^{\beta-1}$$

Computation with hierarchical model III

$$
\begin{aligned}
\pi(\theta \mid \alpha, \beta, y) &= \prod_{j=1}^{J} \frac{\Gamma(\alpha+\beta+n_j)}{\Gamma(\alpha+y_j)\Gamma(\beta+n_j-y_j)} \theta_j^{\alpha+y_j-1}(1-\theta_j)^{\beta+n_j-y_j-1}, \\
\pi(\alpha, \beta \mid y) &\propto \pi(\alpha, \beta) \prod_{j=1}^{J} \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+y_j)\Gamma(\beta+n_j-y_j)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta+n_j)}.
\end{aligned}
$$

Computation with hierarchical model IV

3 Prior for ϕ

$$\pi(\alpha, \beta) \propto (\alpha + \beta)^{-5/2}.$$

It is equal to the uniform prior on the scale of $(\frac{\alpha}{\alpha+\beta}, (\alpha+\beta)^{-1/2})$ (related to mean and variance).

Computation with hierarchical model V

- Computation
  1. Draw the posterior of $(\log(\alpha/\beta), \log(\alpha+\beta))|y$.
  2. Sampling from the above, and transformation to α, β.
  3. Sampling from $\theta|\alpha, \beta, y$ to obtain the credible intervals.

Estimating an exchangeable set of parameters from a normal model I

- Data structure

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

Rewrite through sufficient statistics

$$\bar{y}_{\cdot j}|\theta_j \sim N(\theta_j, \sigma_j^2),\ \sigma_j^2 = \sigma^2/n_j.$$

Estimating an exchangeable set of parameters from a normal model II

- We can consider the pooled estimate of $\theta = \theta_j$,

$$\bar{y}_{\cdot\cdot} = \frac{\sum_{j=1}^{J} \bar{y}_{\cdot j}/\sigma_j^2}{\sum_{j=1}^{J} 1/\sigma_j^2}.$$

- We can consider the separate estimate of θ_j, $\bar{y}_{\cdot j}$.

- If we have the evidence of $\theta_1 = \cdots = \theta_J$, then pooled estimate is preferred.

Estimating an exchangeable set of parameters from a normal model III

- Also we can consider the following:

$$\hat{\theta}_j = \lambda_j \bar{y}_{\cdot j} + (1 - \lambda_j)\bar{y}_{\cdot\cdot}.$$

1. If $\pi(\theta_1, \ldots, \theta_J) = \prod_{j=1}^{J} \pi(\theta_j) \propto 1$, then $\hat{\theta}_j = \bar{y}_{\cdot j}$.
2. If $\pi(\theta = \theta_1 = \cdots = \theta_J) = 1,\ \pi(\theta) \propto 1$, then $\hat{\theta}_j = \bar{y}_{\cdot\cdot}$.
3. If θ_js have priors of i.i.d. normals, then weighted mean is the posterior mean of θ_j.

Estimating an exchangeable set of parameters from a normal model IV

- The hierarchical model

$$
\begin{aligned}
\pi(\theta_1, \ldots, \theta_J | \mu, \eta) &= \prod_{j=1}^{J} \phi(\theta_j; \mu, \eta^2), \\
\pi(\theta_1, \ldots, \theta_J) &= \int \left[\prod_{j=1}^{J} \phi(\theta_j; \mu, \eta^2) \right] \pi(\mu, \eta) d(\mu, \eta).
\end{aligned}
$$

Estimating an exchangeable set of parameters from a normal model V

- The joint posterior distribution

$$\pi(\theta, \mu, \eta \mid y) \propto \pi(\mu, \eta) \prod_{j=1}^{J} \phi(\theta_j \mid \mu, \eta^2) \prod_{j=1}^{J} \phi(\bar{y}_{\cdot j}; \theta_j, \sigma_j^2).$$

Estimating an exchangeable set of parameters from a normal model VI

- Conditional posterior distribution given hyperparameters

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

where

$$\hat{\theta}_j = \frac{\bar{y}_{\cdot j}/\sigma_j^2 + \mu/\eta^2}{1/\sigma_j^2 + 1/\eta^2} \text{ and } V_j = \frac{1}{1/\sigma_j^2 + 1/\eta^2}.$$

Estimating an exchangeable set of parameters from a normal model VII

- The marginal distribution of the hyperparameters

$$\begin{aligned} \pi(\mu, \eta | y) &\propto \pi(\mu, \eta) \prod_{j=1}^{J} \phi(\bar{y}_{\cdot j}; \mu, \sigma_j^2 + \eta^2) \\ \mu | \eta, y &\sim N\left(\hat{\mu}, V_\mu\right), \end{aligned}$$

where $\pi(\mu \mid \eta) \propto 1,\ \hat{\mu} = \frac{\sum_{j=1}^{J} \bar{y}_{\cdot j}/(\sigma_j^2+\eta^2)}{\sum_{j=1}^{J} 1/(\sigma_j^2+\eta^2)}, V_\mu^{-1} = \sum_{j=1}^{J} 1/(\sigma_j^2 + \eta^2).$

Estimating an exchangeable set of parameters from a normal model VIII

- If we let $\mu = \hat{\mu}$ for simplicity, then

$$\begin{aligned}\pi(\eta|y) &\propto \frac{\pi(\mu,\eta|y)}{\pi(\mu|\eta,y)} = \frac{\pi(\eta)\prod_{j=1}^{J}\phi(\bar{y}_{\cdot j};\hat{\mu},\sigma_j^2+\eta^2)}{\phi\left(\hat{\mu};\hat{\mu},V_\mu\right)} \\ &\propto \pi(\eta)V_\mu^{1/2}\prod_{j=1}^{J}(\sigma_j^2+\eta^2)^{-1/2}\exp\left(-\frac{(\bar{y}_{\cdot j}-\hat{\mu})^2}{2(\sigma_j^2+\eta^2)}\right).\end{aligned}$$

- Prior of η: uniform on the scale of η or $\log\eta$.
- Note that frequentist estimates of η^2 can be negative.

Example: Combining information from educational testing experiments in eight schools I

- Data

School	Estimated treatment effect, y_j	Standard error of effect estimate, σ_j
A	28	15
B	8	10
C	−3	16
D	7	11
E	−1	9
F	1	11
G	18	10
H	12	18

Table 5.2 *Observed effects of special preparation on SAT-V scores in eight randomized experiments. Estimates are based on separate analyses for the eight experiments.*

Example: Combining information from educational testing experiments in eight schools II

- Separated estimates

  Credible intervals of θ_js are overlapped.

- A pooled estimate

  Posterior interval of θ, $(-0.3, 16.0)$. Can the value of school A be explained?

Example: Combining information from educational testing experiments in eight schools III

- In the case of school A, separated estimates give the Bayes estimate of 28.4 with a standard error of 14.9. Pooled estimate give the the Bayes estimate of 7.9 with a standard error of 4.2.
- Hierarchical model can be more reasonable.

Example: Combining information from educational testing experiments in eight schools IV

- Results of hierarchical model

School	Posterior quantiles				
	2.5%	25%	median	75%	97.5%
A	−2	7	10	16	31
B	−5	3	8	12	23
C	−11	2	7	11	19
D	−7	4	8	11	21
E	−9	1	5	10	18
F	−7	2	6	10	28
G	−1	7	10	15	26
H	−6	3	8	13	33

Table 5.3: *Summary of 200 simulations of the treatment effects in the eight schools.*

Example: Combining information from educational testing experiments in eight schools V

- Plot of the posterior mean for each η ($\eta = \tau$)

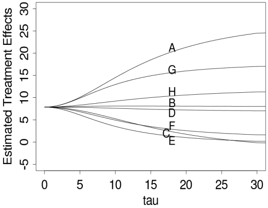

Figure 5.6 *Conditional posterior means of treatment effects, $E(\theta_j|\tau, y)$, as functions of the between-school standard deviation τ, for the educational testing example. The line for school* C *crosses the lines for* E *and* F *because* C *has a higher measurement error (see Table 5.2) and its estimate is therefore shrunk more strongly toward the overall mean in the Bayesian analysis.*

Example: Combining information from educational testing experiments in eight schools VI

- Plot of the posterior s.d. for each η ($\eta = \tau$)

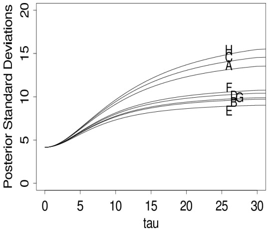

Figure 5.7 *Conditional posterior standard deviations of treatment effects,* $\mathrm{sd}(\theta_j|\tau, y)$*, as functions of the between-school standard deviation* τ*, for the educational testing example.*

Example: Combining information from educational testing experiments in eight schools VII

- Marginal posterior of η ($\eta = \tau$)

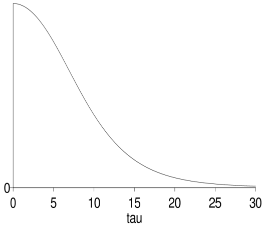

Figure 5.5 *Marginal posterior density, $p(\tau|y)$, for standard deviation of the population of school effects θ_j in the educational testing example.*

Example: Combining information from educational testing experiments in eight schools VIII

- Notes
  1. As decreasing η, results are similar to that of $\theta = \theta_1, \ldots = \theta_J$.
  2. $P(\eta > 25|y) \approx 0$.
  3. $P(\theta_1 > 28|y) \leq 0.1$ where separate estimates give the the probability of 0.5.
  4. $P(\max_j \theta_j > 28.4|y) \approx \frac{22}{200}$ and $P(\theta_1 > \theta_3|y) \approx \frac{141}{200} = 0.705$.
- Appropriate shrinkage to common θ.