5: Hierarchical models

09/25/19

Hierarchical model

1. choose (hyper)prior $p(\phi)$
2. choose prior $p(\theta \mid \phi)$ — exchangeable regarding θ
3. choose likelihood $p(y \mid \theta) = \prod_{j=1}^{J} p(y_j \mid \theta_j)$

Then

$$p(\theta, \phi \mid y) \propto p(y \mid \theta) p(\theta \mid \phi) p(\phi)$$

Question: how do we draw simulations from the above posterior distribution?

Bayesian analysis of conjugate hierarchical models

Suppose $p(\theta \mid \phi)$ is conjugate to $p(y \mid \theta)$

- Determine $p(\theta \mid y, \phi)$ in closed form
- Determine $p(\phi \mid y)$ (up to normalizing constant)

Drawing simulations from $p(\theta, \phi \mid y)$ according to

Bayesian analysis of conjugate hierarchical models

Suppose $p(\theta \mid \phi)$ is conjugate to $p(y \mid \theta)$

- Determine $p(\theta \mid y, \phi)$ in closed form
- Determine $p(\phi \mid y)$ (up to normalizing constant)

Drawing simulations from $p(\theta, \phi \mid y)$ according to
For $i = 1, \ldots, L$

1. Draw the vector of hyperparameters, ϕ, from $p(\phi \mid y)$, by selecting discretized grid using Chapter 3 or more sophisticated methods in Part III of the book

Bayesian analysis of conjugate hierarchical models

Suppose $p(\theta \mid \phi)$ is conjugate to $p(y \mid \theta)$

- Determine $p(\theta \mid y, \phi)$ in closed form
- Determine $p(\phi \mid y)$ (up to normalizing constant)

Drawing simulations from $p(\theta, \phi \mid y)$ according to
For $i = 1, \ldots, L$

1. Draw the vector of hyperparameters, ϕ, from $p(\phi \mid y)$, by selecting discretized grid using Chapter 3 or more sophisticated methods in Part III of the book
2. Draw the parameter vector θ, from $p(\theta \mid y, \phi)$. If $p(\theta \mid \phi)$ is facterized (conditionally independent of ϕ), then the components of θ can be drawn independently

Bayesian analysis of conjugate hierarchical models

Suppose $p(\theta \mid \phi)$ is conjugate to $p(y \mid \theta)$

- Determine $p(\theta \mid y, \phi)$ in closed form
- Determine $p(\phi \mid y)$ (up to normalizing constant)

Drawing simulations from $p(\theta, \phi \mid y)$ according to
For $i = 1, \ldots, L$

1. Draw the vector of hyperparameters, ϕ, from $p(\phi \mid y)$, by selecting discretized grid using Chapter 3 or more sophisticated methods in Part III of the book
2. Draw the parameter vector θ, from $p(\theta \mid y, \phi)$. If $p(\theta \mid \phi)$ is facterized (conditionally independent of ϕ), then the components of θ can be drawn independently
3. Draw predicted value $\tilde{y}_j$, from $p(y_j \mid \theta_j)$; If $\tilde{y}$ might corresponds to new θ values other than the existing ones, draw $\tilde{\theta}$, from $p(\theta \mid \phi)$, then draw $\tilde{y}$, from $p(y \mid \tilde{\theta})$

Finding $p(\alpha, \beta \mid y)$

We choose the prior $p(\theta_{1:71} \mid \alpha, \beta)p(\alpha, \beta)$. Section 5.3 is mostly interested in the marginal posterior $p(\alpha, \beta \mid y)$.

1. determine the conditional posterior in *closed form* $p(\theta_{1:71} \mid y, \alpha, \beta)$.

Finding $p(\alpha, \beta \mid y)$

We choose the prior $p(\theta_{1:71} \mid \alpha, \beta)p(\alpha, \beta)$. Section 5.3 is mostly interested in the marginal posterior $p(\alpha, \beta \mid y)$.

1. determine the conditional posterior in *closed form* $p(\theta_{1:71} \mid y, \alpha, \beta)$.

2. determine the *unnormalized* version of the marginal posterior using the following formula

$$
\begin{aligned}
p(\alpha, \beta \mid y) &= \frac{p(\theta_{1:71}, \alpha, \beta \mid y)}{p(\theta_{1:71} \mid y, \alpha, \beta)} \\
&\propto \frac{p(y \mid \theta_{1:71})p(\theta_{1:71} \mid \alpha, \beta)p(\alpha, \beta)}{p(\theta_{1:71} \mid y, \alpha, \beta)}
\end{aligned}
$$

Finding $p(\alpha, \beta \mid y)$

1. determine the conditional posterior in *closed form* $p(\theta_{1:71} \mid y, \alpha, \beta)$. We assume $\theta_{1:71} \mid \alpha, \beta \sim \text{Beta}(\alpha, \beta)$. Reminder: when we write $\propto$ we can drop anything that isn't a $\theta_{1:71}$.

$$
\begin{aligned}
p(\theta_{1:71} \mid y, \alpha, \beta) &\propto p(y \mid \theta_{1:71}) p(\theta_{1:71} \mid \alpha, \beta) \\
&\propto \prod_{j=1}^{71} \theta_j^{\alpha-1}(1-\theta_j)^{\beta-1} \\
&\times \prod_{j=1}^{71} \theta_j^{y_j}(1-\theta_j)^{n_j-y_j} \\
&= \prod_{j=1}^{71} \theta_j^{y_j+\alpha-1}(1-\theta_j)^{n_j-y_j+\beta-1}
\end{aligned}
$$

So $p(\theta_{1:71} \mid y, \alpha, \beta) = \prod_{j=1}^{71} \text{Beta}(\alpha + y_j, \beta + n_j - y - j)$

Finding $p(\alpha, \beta \mid y)$

2. determine the *unnormalized* version of the marginal posterior using the following formula. When we write $\propto$, we can drop anything that isn't involving α, β.

$$
\begin{aligned}
p(\alpha, \beta \mid y) &\propto \frac{p(y \mid \theta_{1:71}) p(\theta_{1:71} \mid \alpha, \beta) p(\alpha, \beta)}{p(\theta_{1:71} \mid y, \alpha, \beta)} \qquad \text{(earlier slides)} \\
&\propto \frac{p(\theta_{1:71} \mid \alpha, \beta) p(\alpha, \beta)}{\prod_{j=1}^{71} \frac{\Gamma(n_j+\alpha+\beta)}{\Gamma(y_j+\alpha)\Gamma(n_j-y_j+\beta)} \theta_j^{y_j+\alpha-1} (1-\theta_j)^{n_j-y_j+\beta-1}} \\
&= \frac{p(\alpha, \beta) \prod_{j=1}^{71} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_j^{\alpha-1} (1-\theta_j)^{\beta-1}}{\prod_{j=1}^{71} \frac{\Gamma(n_j+\alpha+\beta)}{\Gamma(y_j+\alpha)\Gamma(n_j-y_j+\beta)} \theta_j^{y_j+\alpha-1} (1-\theta_j)^{n_j-y_j+\beta-1}} \\
&\propto p(\alpha, \beta) \prod_{j=1}^{71} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \Big/ \frac{\Gamma(n_j+\alpha+\beta)}{\Gamma(y_j+\alpha)\Gamma(n_j-y_j+\beta)}
\end{aligned}
$$

Finding $p(\alpha, \beta \mid y)$

$$p(\alpha, \beta \mid y) \propto (\alpha + \beta)^{-5/2} \prod_{j=1}^{71} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \Big/ \frac{\Gamma(n_j + \alpha + \beta)}{\Gamma(y_j + \alpha)\Gamma(n_j - y_j + \beta)}$$

so

$$\log p(\alpha, \beta \mid y)$$
$$= c + \sum_{j=1}^{71} \left\{ \log \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} - \log \frac{\Gamma(n_j + \alpha + \beta)}{\Gamma(y_j + \alpha)\Gamma(n_j - y_j + \beta)} \right\} - \frac{5}{2} \log(\alpha + \beta)$$

Finding $p(\alpha, \beta \mid y)$

Missing $-\frac{5}{2}\log(\alpha+\beta)$?

```
A <- seq(0.5, 6, length.out = 100)
B <- seq(3, 33, length.out = 100)
cA <- rep(A, each = length(B))
cB <- rep(B, length(A))
lpfun <- function(a, b, y, n) log(a+b)*(-5/2) +
  sum(lgamma(a+b)-lgamma(a)-lgamma(b)
  +lgamma(a+y)+lgamma(b+n-y)-lgamma(a+b+n))
lp <- mapply(lpfun, cA, cB, MoreArgs = list(y, n))
```

http://avehtari.github.io/BDA_R_demos/demos_ch5/demo5_1.html

Finding $p(\alpha, \beta \mid y)$

So then we exponentiate. But watch out:

```
> head(lp)
[1] -747.6954 -747.6320 -747.8540 -748.3062 -748.9466 -749.742
> head(exp(lp))
[1] 0 0 0 0 0 0
```

Finding $p(\alpha, \beta \mid y)$

So then we exponentiate. But watch out:

```
> head(lp)
[1] -747.6954 -747.6320 -747.8540 -748.3062 -748.9466 -749.742
> head(exp(lp))
[1] 0 0 0 0 0 0
```

This is **numerical underflow**. Solution:

$$p(\alpha, \beta \mid y) \propto \exp[\log p(\alpha, \beta \mid y) + m]$$

`m` is any "big" number. Careful not to set it too large, because then you will get **overflow**. The author uses a good data-dependent solution: set `m` to be equal to be negative of the maximum of these log-values, which is $\log p(\alpha, \beta \mid y)$.

Prior $p(\alpha, \beta)$

A noninformative hyperprior is desired. Consider "uniform" distribution on the reparameterized hyperparameters

- $(\log(\alpha/\beta), \log(\alpha+\beta)) \in (-\infty, \infty)^2$, which correspond to logit mean and the logarithm of the "sample size" in the Beta prior
- $(\alpha/(\alpha+\beta), (\alpha+\beta)^{-1/2}) \in (0,1) \times (0, \infty)$, which correspond to mean and standard deviation () in the Beta prior

The first choice leads to improper posterior distribution because $p(\alpha, \beta \mid y)$ is unbounded when $(\alpha+\beta) \to \infty$. Setting a weakly informative prior, uniform$[-10^{10}, 10^{10}] \times [-10^{10}, 10^{10}]$ would not be acceptable!

The second choice leads to $p(\alpha, \beta) \propto (\alpha+\beta)^{-5/2}$ or $p(\log(\alpha/\beta), \log(\alpha+\beta)) \propto (\alpha+\beta)^{-5/2}\alpha\beta$.

Results

- Behavior of the posterior mean and credible interval of θ_j
- Comparison of full Bayesian and empirical Bayesian