Stan Korea

StanKorea 2019W Study1

Bayesian Data Analysis Ch.3
Introduction to Multiparameter Models

발표자: 임정욱

발표일: 2019.01.07 (Mon)
발표장: NextOpt/StanKorea Office

3.1 Averaging over 'nuisance parameters'

Nuisance parameter is any parameter which is not of immediate interest but which must be accounted for in the analysis of those parameters which are of inter

Ex) $y|\mu,\sigma^2 \sim \mathrm{N}(\mu,\sigma^2),$

Nuisance parameter

$$p(\theta_1|y) = \int p(\theta_1,\theta_2|y)d\theta_2.$$

$$p(\theta_1|y) = \int p(\theta_1|\theta_2,y)p(\theta_2|y)d\theta_2, \qquad (3.1)$$

3.2 Normal data with a noninformative prior distribution

1. A noninformative prior distribution

$$p(\mu, \sigma^2) \propto (\sigma^2)^{-1}.$$

2. The joint posterior distribution, $p(\mu, \sigma^2 | y)$

$$\begin{aligned} p(\mu, \sigma^2|y) &\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right) \\ &= \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(y_i-\overline{y})^2 + n(\overline{y}-\mu)^2\right]\right) \\ &= \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\overline{y}-\mu)^2]\right), \end{aligned} \qquad (3.2)$$

3. The conditional posterior distribution, $p(\mu|\sigma^2, y)$

$$\mu|\sigma^2, y \sim \mathrm{N}(\overline{y}, \sigma^2/n). \qquad (3.3)$$

Stan Korea

3.2 Normal data with a noninformative prior distribution

4. The marginal posterior distribution, $p(\sigma^2|y)$

$$p(\sigma^2|y) \propto \int \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\overline{y}-\mu)^2]\right) d\mu.$$

5. Sampling from the joint posterior distribution

$$\begin{aligned} p(\sigma^2|y) &\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}(n-1)s^2\right) \sqrt{2\pi\sigma^2/n} \\ &\propto (\sigma^2)^{-(n+1)/2} \exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right), \end{aligned}$$

$$\sigma^2|y \sim \text{Inv-}\chi^2(n-1, s^2). \tag{3.5}$$

3.2 Normal data with a noninformative prior distribution

6. Analytic form of the marginal posterior distribution of μ

$$p(\mu|y) = \int_0^\infty p(\mu, \sigma^2|y) d\sigma^2.$$

$$z = \frac{A}{2\sigma^2}, \quad \text{where } A = (n-1)s^2 + n(\mu - \overline{y})^2,$$

$$\begin{aligned} p(\mu|y) &\propto A^{-n/2} \int_0^\infty z^{(n-2)/2} \exp(-z) dz \\ &\propto [(n-1)s^2 + n(\mu - \overline{y})^2]^{-n/2} \\ &\propto \left[1 + \frac{n(\mu - \overline{y})^2}{(n-1)s^2}\right]^{-n/2}. \end{aligned}$$

$$\longrightarrow t_{n-1}(\overline{y}, s^2/n) \longrightarrow \left.\frac{\mu - \overline{y}}{s/\sqrt{n}}\right| y \sim t_{n-1},$$

7. Posterior predictive distribution for a future observation

$$p(\tilde{y}|y) = \int\int p(\tilde{y}|\mu, \sigma^2, y) p(\mu, \sigma^2|y) d\mu d\sigma^2.$$

$$p(\tilde{y}|\sigma^2, y) = \int p(\tilde{y}|\mu, \sigma^2, y) p(\mu|\sigma^2, y) d\mu \longrightarrow p(\tilde{y}|\sigma^2, y) = \mathrm{N}(\tilde{y}|\overline{y}, (1 + \tfrac{1}{n})\sigma^2),$$

3.3 Normal data with a conjugate prior distribution

1. A family of conjugate prior distributions

$$\begin{aligned} \mu|\sigma^2 &\sim \mathrm{N}(\mu_0, \sigma^2/\kappa_0) \\ \sigma^2 &\sim \text{Inv-}\chi^2(\nu_0, \sigma_0^2), \end{aligned} \longrightarrow p(\mu, \sigma^2) \propto \sigma^{-1}(\sigma^2)^{-(\nu_0/2+1)} \exp\left(-\frac{1}{2\sigma^2}[\nu_0\sigma_0^2 + \kappa_0(\mu_0 - \mu)^2]\right). \quad (3.6)$$

2. The joint posterior distribution, $p(\mu, \sigma^2 | y)$

$$\begin{aligned} p(\mu, \sigma^2|y) &\propto \sigma^{-1}(\sigma^2)^{-(\nu_0/2+1)} \exp\left(-\frac{1}{2\sigma^2}[\nu_0\sigma_0^2 + \kappa_0(\mu - \mu_0)^2]\right) \times \\ &\quad \times (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\overline{y} - \mu)^2]\right) \quad (3.7) \end{aligned}$$

$$\begin{aligned} \mu_n &= \frac{\kappa_0}{\kappa_0 + n}\mu_0 + \frac{n}{\kappa_0 + n}\overline{y} \\ \kappa_n &= \kappa_0 + n \\ \nu_n &= \nu_0 + n \\ \nu_n\sigma_n^2 &= \nu_0\sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_0 + n}(\overline{y} - \mu_0)^2. \end{aligned}$$

$$= \text{N-Inv-}\chi^2(\mu_n, \sigma_n^2/\kappa_n; \nu_n, \sigma_n^2),$$

3.3 Normal data with a conjugate prior distribution

3. The conditional posterior distribution, $p(\mu|\sigma^2, y)$

$$\begin{aligned} \mu|\sigma^2, y &\sim \mathrm{N}(\mu_n, \sigma^2/\kappa_n) \\ &= \mathrm{N}\left(\frac{\frac{\kappa_0}{\sigma^2}\mu_0 + \frac{n}{\sigma^2}\overline{y}}{\frac{\kappa_0}{\sigma^2} + \frac{n}{\sigma^2}}, \frac{1}{\frac{\kappa_0}{\sigma^2} + \frac{n}{\sigma^2}}\right), \end{aligned} \tag{3.8}$$

4. The marginal posterior distribution, $p(\sigma^2|y)$

$$\sigma^2|y \sim \text{Inv-}\chi^2(\nu_n, \sigma_n^2). \tag{3.9}$$

5. Sampling from the joint posterior distribution

6. Analytic form of the marginal posterior distribution of μ

$$\begin{aligned} p(\mu|y) &\propto \left(1 + \frac{\kappa_n(\mu - \mu_n)^2}{\nu_n \sigma_n^2}\right)^{-(\nu_n+1)/2} \\ &= t_{\nu_n}(\mu|\mu_n, \sigma_n^2/\kappa_n). \end{aligned}$$

3.4 Multinomial model for categorical data

$$p(y|\theta) \propto \prod_{j=1}^{k} \theta_j^{y_j}, \longleftrightarrow p(\theta|\alpha) \propto \prod_{j=1}^{k} \theta_j^{\alpha_j - 1},$$

Conjugate prior

Example. Pre-election polling

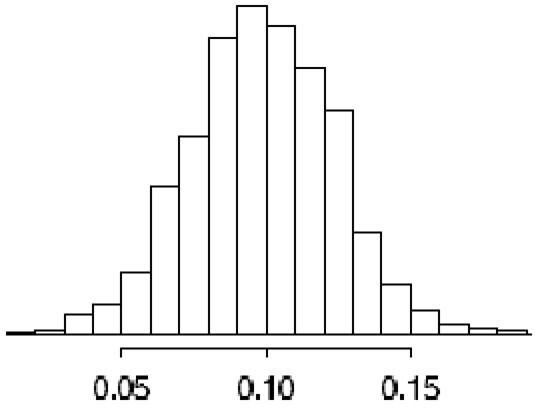

Figure 3.2 *Histogram of values of $(\theta_1 - \theta_2)$ for 1000 simulations from the posterior distribution for the election polling example.*

3.5 Multivariate normal model with known variance

1. Multivariate normal likelihood

$$y|\mu, \Sigma \sim \mathrm{N}(\mu, \Sigma), \qquad (3.10)$$

$$p(y|\mu, \Sigma) \propto |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(y-\mu)^T \Sigma^{-1} (y-\mu)\right),$$

$$\begin{aligned} p(y_1, \ldots, y_n|\mu, \Sigma) &\propto |\Sigma|^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(y_i-\mu)^T \Sigma^{-1}(y_i-\mu)\right) \\ &= |\Sigma|^{-n/2} \exp\left(-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}S_0)\right), \qquad (3.11) \end{aligned}$$

2. Conjugate analysis

3. Posterior conditional and marginal distributions of subvectors of μ with known Σ

$$p(\mu|y, \Sigma) \propto \exp\left(-\frac{1}{2}\left((\mu-\mu_0)^T \Lambda_0^{-1}(\mu-\mu_0) + \sum_{i=1}^{n}(y_i-\mu)^T \Sigma^{-1}(y_i-\mu)\right)\right),$$

$$\begin{aligned} p(\mu|y, \Sigma) &\propto \exp\left(-\frac{1}{2}(\mu-\mu_n)^T \Lambda_n^{-1}(\mu-\mu_n)\right) \\ &= \mathrm{N}(\mu|\mu_n, \Lambda_n), \end{aligned}$$

$$\begin{aligned} \mu_n &= (\Lambda_0^{-1} + n\Sigma^{-1})^{-1}(\Lambda_0^{-1}\mu_0 + n\Sigma^{-1}\overline{y}) \\ \Lambda_n^{-1} &= \Lambda_0^{-1} + n\Sigma^{-1}. \end{aligned}$$

3.5 Multivariate normal model with known variance

4. Posterior predictive distribution for new data

$$\mu^{(1)}|\mu^{(2)}, y \sim \mathrm{N}\left(\mu_n^{(1)} + \beta^{1|2}(\mu^{(2)} - \mu_n^{(2)}), \Lambda^{1|2}\right), \qquad (3.14)$$

$$\begin{aligned} \beta^{1|2} &= \Lambda_n^{(12)}\left(\Lambda_n^{(22)}\right)^{-1} \\ \Lambda^{1|2} &= \Lambda_n^{(11)} - \Lambda_n^{(12)}\left(\Lambda_n^{(22)}\right)^{-1}\Lambda_n^{(21)}. \end{aligned}$$

5. Posterior predictive distribution for new data

$$\begin{aligned} \mathrm{E}(\tilde{y}|y) &= \mathrm{E}(\mathrm{E}(\tilde{y}|\mu, y)|y) \\ &= \mathrm{E}(\mu|y) = \mu_n, \end{aligned}$$

$$\begin{aligned} \mathrm{var}(\tilde{y}|y) &= \mathrm{E}(\mathrm{var}(\tilde{y}|\mu, y)|y) + \mathrm{var}(\mathrm{E}(\tilde{y}|\mu, y)|y) \\ &= \mathrm{E}(\Sigma|y) + \mathrm{var}(\mu|y) = \Sigma + \Lambda_n. \end{aligned}$$

6. Noninformative prior density for μ

3.6 Multivariate normal with unknown mean and variance

1. Conjugate inverse-Wishart family of prior distributions

$$\begin{aligned}\Sigma &\sim \text{Inv-Wishart}_{\nu_0}(\Lambda_0^{-1})\\ \mu|\Sigma &\sim \text{N}(\mu_0, \Sigma/\kappa_0),\end{aligned}$$

$$p(\mu, \Sigma) \propto |\Sigma|^{-((\nu_0+d)/2+1)} \exp\left(-\frac{1}{2}\text{tr}(\Lambda_0\Sigma^{-1}) - \frac{\kappa_0}{2}(\mu-\mu_0)^T\Sigma^{-1}(\mu-\mu_0)\right).$$

$$\begin{aligned}\mu_n &= \frac{\kappa_0}{\kappa_0+n}\mu_0 + \frac{n}{\kappa_0+n}\overline{y}\\ \kappa_n &= \kappa_0+n\\ \nu_n &= \nu_0+n\\ \Lambda_n &= \Lambda_0 + S + \frac{\kappa_0 n}{\kappa_0+n}(\overline{y}-\mu_0)(\overline{y}-\mu_0)^T,\end{aligned}$$

2. Different noninformative prior distributions

3.6 Multivariate normal with unknown mean and variance

3. Inverse-Wishart with d + 1 degrees of freedom

4. Inverse-Wishart with d−1 degrees of freedom

$$p(\mu, \Sigma) \propto |\Sigma|^{-(d+1)/2},$$

$$\begin{aligned} \Sigma|y &\sim \text{Inv-Wishart}_{n-1}(S^{-1}) \\ \mu|\Sigma, y &\sim \text{N}(\overline{y}, \Sigma/n). \end{aligned}$$

5. Scaled inverse-Wishart model

$$\Sigma = \text{Diag}(\xi)\Sigma_\eta\text{Diag}(\xi),$$

3.7 Example: analysis of a bioassay experiment

1. The scientific problem and the data

$(x_i, n_i, y_i);\ i = 1, \dots, k,$

Ex)

Dose, x_i (log g/ml)	Number of animals, n_i	Number of deaths, y_i
−0.86	5	0
−0.30	5	1
−0.05	5	3
0.73	5	5

Table 3.1: *Bioassay data from Racine et al. (1986).*

2. Modeling the dose–response relation

$y_i|\theta_i \sim \text{Bin}(n_i, \theta_i),$

$\text{logit}(\theta_i) = \alpha + \beta x_i,$ Logistic regression

3. The likelihood

$$p(y_i|\alpha, \beta, n_i, x_i) \propto [\text{logit}^{-1}(\alpha + \beta x_i)]^{y_i}[1 - \text{logit}^{-1}(\alpha + \beta x_i)]^{n_i - y_i}.$$

$$\begin{aligned} p(\alpha, \beta|y, n, x) &\propto p(\alpha, \beta|n, x)p(y|\alpha, \beta, n, x) \qquad (3.16) \\ &\propto p(\alpha, \beta)\prod_{i=1}^{k} p(y_i|\alpha, \beta, n_i, x_i). \end{aligned}$$

4. The prior distribution

3.7 Example: analysis of a bioassay experiment

5. A rough estimate of the parameters

6. Obtaining a contour plot of the joint posterior density

7. Sampling from the joint posterior distribution
   1. Compute the marginal posterior distribution of α by numerically summing over β in the discrete distribution computed on the grid of Figure 3.3a.
   2. For $s = 1, \ldots, 1000$:
      (a) Draw α^s from the discretely computed $p(\alpha|y)$; this can be viewed as a discrete version of the inverse cdf method described in Section 1.9.
      (b) Draw β^s from the discrete conditional distribution, $p(\beta|\alpha, y)$, given the just-sampled value of α.
      (c) For each of the sampled α and β, add a uniform random jitter centered at zero with a width equal to the spacing of the sampling grid. This gives the simulation draws a continuous distribution.

8. The posterior distribution of the LD50

9. Difficulties with the LD50 parameterization if the drug is beneficial

3.8 Summary of elementary modeling and computation

Lack of multiparameter models is not a major handicap because of

1. simple simulation methods,
2. a hierarchical or conditional manner,
3. applying a normal approximation & conjugate structure

can play an important role in practice.

Thank You.