Applied Bayesian and Numerical Methods

Hierarchical model

Mohib Ullah

October 27, 2016

Exercise 5.13 (Hierarchical binomial model)

(a). Set up a model for the data in Table 3.3 so that, for $j = 1, ..., 10$, the observed number of bicycles at location j is binomial with unknown probability θ_j and sample size equal to the total number of vehicles (bicycles included) in that block. The parameter θ_j can be interpreted as the underlying or 'true' proportion of traffic at location j that is bicycles. Assign a beta population distribution for the parameters θ_j and a non-informative hyper-prior distribution. Write down the joint posterior distribution i.e $p(\theta, \alpha, \beta | y)$.

Joint Posterior Distribution

$$p(\theta, \alpha, \beta | y) \propto \underbrace{\boxed{p(y|\theta, \alpha, \beta)}}_{\text{Likelihood}} \; \underbrace{\boxed{p(\theta|\alpha, \beta)p(\alpha, \beta)}}_{\text{joint prior distribution } p(\theta,\,\alpha,\,\beta)}$$

Joint Posterior Distribution

$$p(\theta, \alpha, \beta | y) \propto \underbrace{\boxed{p(y|\theta, \alpha, \beta)}}_{\text{Likelihood}} \underbrace{\boxed{p(\theta|\alpha, \beta)p(\alpha, \beta)}}_{\text{joint prior distribution } p(\theta, \alpha, \beta)}$$

Let us consider $y_1, y_2, ..., y_{10}$ are the number of bicycles at ten different locations, and $n_1, n_2..., n_{10}$ are the totel vehicles including bicycles at location 1 to 10, According to the statement, y_j can be considered to be generated from a binomial process with known θ_j;

$$y_j|\theta_j \sim Bin(n_j, \theta_j), j = 1, 2, ..., 10$$

Joint Posterior Distribution

$$p(\theta, \alpha, \beta | y) \propto \underbrace{\boxed{p(y|\theta, \alpha, \beta)}}_{\text{Likelihood}} \; \underbrace{\boxed{p(\theta|\alpha, \beta) p(\alpha, \beta)}}_{\text{joint prior distribution } p(\theta, \alpha, \beta)}$$

Let us consider $y_1, y_2, ..., y_{10}$ are the number of bicycles at ten different locations, and $n_1, n_2..., n_{10}$ are the totel vehicles including bicycles at location $1 - 10$, According to the statement, y_j can be considered to be generated from a binomial process with known θ_j;

$$y_j|\theta_j \sim Bin(n_j, \theta_j), j = 1, 2, ..., 10$$

$$p(y|\theta, \alpha, \beta) = \prod_{j=1}^{10} \binom{n_j}{y_j} \theta_j^{y_j} (1 - \theta_j)^{n_j - y_j}$$

Joint Posterior Distribution

$$p(\theta, \alpha, \beta|y) \propto \boxed{p(y|\theta, \alpha, \beta)} \; \boxed{p(\theta|\alpha, \beta)p(\alpha, \beta)}$$

Likelihood　　joint prior distribution $p(\theta, \alpha, \beta)$

Let us consider $y_1, y_2, ..., y_{10}$ are the number of bicycles at ten different locations, and $n_1, n_2..., n_{10}$ are the totel vehicles including bicycles at location $1 - 10$, According to the statement, y_j can be considered to be generated from a binomial process with known θ_j;

$$y_j|\theta_j \sim Bin(n_j, \theta_j), j = 1, 2, ..., 10$$

$$p(y|\theta, \alpha, \beta) = \prod_{j=1}^{10} \binom{n_j}{y_j} \theta_j^{y_j}(1-\theta_j)^{n_j - y_j}$$

Similarly, θ_j has *Beta* distribution with hyper-parameter α and β,

$$\theta_j \sim Beta(\alpha, \beta)$$

Joint Posterior Distribution

$$p(\theta|\alpha,\beta) = \prod_{j=1}^{10} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_j^{\alpha-1}(1-\theta_j)^{\beta-1}$$

One of the possible non-informative prior distribution for the hyper parameters is;

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

Joint Posterior Distribution

$$p(\theta|\alpha,\beta) = \prod_{j=1}^{10} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_j^{\alpha-1}(1-\theta_j)^{\beta-1}$$

One of the possible non-informative prior distribution for the hyper parameters is;

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

By putting things together,

$$p(\theta,\alpha,\beta|y) \propto p(\alpha,\beta) \prod_{j=1}^{10} \theta_j^{y_j}(1-\theta_j)^{n_j-y_j} \prod_{j=1}^{10} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_j^{\alpha-1}(1-\theta_j)^{\beta-1}$$

Joint Posterior Distribution

$$p(\theta|\alpha,\beta) = \prod_{j=1}^{10} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_j^{\alpha-1}(1-\theta_j)^{\beta-1}$$

One of the possible non-informative prior distribution for the hyper parameters is;

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

By putting things together,

$$p(\theta,\alpha,\beta|y) \propto p(\alpha,\beta) \prod_{j=1}^{10} \theta_j^{y_j}(1-\theta_j)^{n_j-y_j} \prod_{j=1}^{10} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_j^{\alpha-1}(1-\theta_j)^{\beta-1}$$

$$p(\theta,\alpha,\beta|y) \propto (\alpha+\beta)^{-5/2} \left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right)^{10} \prod_{j=1}^{10} \theta_j^{y_j}(1-\theta_j)^{n_j-y_j} \theta_j^{\alpha-1}(1-\theta_j)^{\beta-1}$$

Joint Posterior Distribution

$$p(\theta, \alpha, \beta | y) \propto (\alpha + \beta)^{-5/2} \left(\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \right)^{10} \prod_{j=1}^{10} \theta_j^{\alpha + y_j - 1} (1 - \theta_j)^{\beta + n_j - y_j - 1}$$

Marginal Distribution of the Hyperparameters

(b). Compute the marginal posterior density of the hyperparameters and draw simulations from the joint posterior distribution of the parameters and hyperparameters.

The marginal distribution of the hyperparameters can be computed through the following algebraic expression,

$$p(\alpha, \beta|y) = \frac{p(\theta, \alpha, \beta|y)}{p(\theta|\alpha, \beta, y)}$$

Marginal Distribution of the Hyperparameters

$$p(\theta, \alpha, \beta|y) \propto (\alpha+\beta)^{-5/2} \left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right)^{10} \prod_{j=1}^{10} \theta_j^{\alpha+y_j-1}(1-\theta_j)^{\beta+n_j-y_j-1}$$

$$p(\theta|\alpha, \beta, y) = \prod_{j=1}^{10} \frac{\Gamma(\alpha+\beta+n_j)}{\Gamma(\alpha+y_j)\Gamma(\beta+n_j-y_j)} \theta_j^{\alpha+y_j-1}(1-\theta_j)^{\beta+n_j-y_j-1}$$

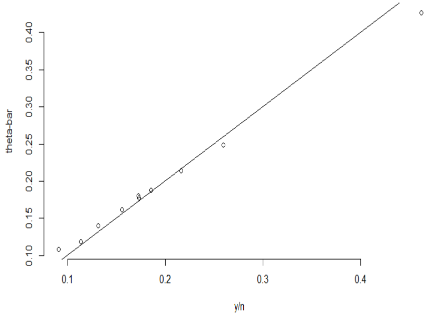
$$p(\alpha, \beta|y) \propto (\alpha+\beta)^{-5/2} \left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right)^{10} \prod_{j=1}^{10} \frac{\Gamma(\alpha+y_j)\Gamma(\beta+n_j-y_j)}{\Gamma(\alpha+\beta+n_j)}$$

Simulation from the posterior distribution

- Usually Gibbs simpling is used to simulate from a distribution. But it works only if we know the normalized form of the distribution. In this case $p(\alpha, \beta|y)$ is not available in normalized form. In order to sample from unnormalized distribution, different techniques like acceptance-rejection methods, Metropolis-Hastings algorithm can be used.

- For the given values of α and β, we can simulate $p(\theta_j|\alpha, \beta, y)$ because its normalized form is available.

(c). Compare the posterior distributions of the parameters θ_j to the raw proportions. (number of bicycles / total number of vehicles) in location j.



95% posterior interval

(d). Give a 95% posterior interval for the average underlying proportion of traffic that is bicycles.

- $p < -alpha/(alpha + beta)$
- $quantile(p, c(0.025, 0.975))$
- 2.5% 97.5%
- 0.1465 0.2938

It shows that 14.6% to 29.4% of the vehicles are bicycles.

Predictive distribution

(e). A new city block is sampled at random and is a residential street with a bike route. In an hour of observation, 100 vehicles of all kinds go by. Give a 95% posterior interval for the number of those vehicles that are bicycles.

- Essentially, we are looking for posterior predictive distribution $p(\tilde{y}|y)$.
- In order to get a sample from it, we need to generate new values of $\tilde{\theta}$ from α,β samples.

$$\tilde{\theta} \sim Beta(\alpha, \beta)$$

- Now using these values, we can sample from binomial distribution with $n = 100$ i.e.

$$\tilde{y}|\tilde{\theta} \sim Bin(100, \tilde{\theta})$$

Predictive distribution

- $theta.new <- rbeta(S, alpha, beta)$
- $y.new <- rbinom(S, 100, theta.new)$
- $quantile(y.new, c(0.025, 0.975))$
- 2.5%   97.5%
- 3   49

which means 3 to 49 bicycles will pass from the residential street with bike route (newly sampled) with the probability of 95%.

(f). Was the beta distribution for the θ_j's reasonable ?

The plot of θ vs $\frac{y}{n}$ fits well which indicates the values of θ are reasonable.

Thanks