Lesson 11 Part A
Тест, 8 вопроса

> 1
>
> Баллы

1.

Which of the following situations would call for a hierarchical model, due to the hierarchical structure of the data?

○ You run an internet connection speed test on your computer each Monday morning for ten consecutive weeks.

○ You survey a random sample of 100 individuals in your neighborhood who report their preferred produce market.

◉ You take blood pressure measurements from each of five individuals on ten consecutive days.

○ You measure an individual's mood with a questionnaire for ten consecutive days.

> 1
>
> Баллы

2.

In hierarchical models, the observations are still conditionally independent, given their respective parameters. How then does such a model capture correlation among grouped observations?

○ Grouped observations require an extra parameter for each pair of observations which estimates the correlation among them.

○ Grouped observations always exhibit correlation, even in an independent model.

◉ Grouped observations share common parameters, which themselves share a common distribution across groups.

○ Observation pairs from different groups require an extra parameter which induces negative correlation among them.

> 1
>
> Баллы

Lesson 11 Part A

In previous lessons, we fit models to data representing percent growth in personnel for companies in two industries. Below are attached additional data from the original two industries (with 10 and six companies respectively), as well as three additional industries. Percent growth is reported for a total of 53 companies.

> pctgrowth.csv

As usual, you can read the data into R with

```
1   dat = read.csv(file="pctgrowth.csv", header=TRUE)
```

where you will need to set R's working directory to where you have saved the file, or include the complete path to the file.

Rather than fit five separate models, one for each industry, we can fit a hierarchical model. As before, we assume a normal likelihood and common variance across all observations. Each industry will have it's own mean growth, and each of these means will come from a common distribution, from which we will estimate the overall mean and variability across industries.

Let i index the individual companies, and g_i indicate the industry (**grp** variable in **pctgrowth.csv**) for company i. Which of the following hierarchical models is the one described above?

○ $$y_i \mid \theta_{g_i}, \sigma^2 \overset{\text{ind}}{\sim} N(\theta_{g_i}, \sigma^2), \quad i = 1, \ldots, 53, \quad g_i \in \{1, \ldots, 5\},$$
$$\theta_g \overset{\text{iid}}{\sim} N(0, 1), \quad g = 1, \ldots, 5,$$
$$\sigma^2 \sim IG(2/2, 2 \cdot 1/2)$$

◉ $$y_i \mid \theta_{g_i}, \sigma^2 \overset{\text{ind}}{\sim} N(\theta_{g_i}, \sigma^2), \quad i = 1, \ldots, 53, \quad g_i \in \{1, \ldots, 5\},$$
$$\theta_g \mid \mu, \tau^2 \overset{\text{iid}}{\sim} N(\mu, \tau^2), \quad g = 1, \ldots, 5,$$
$$\mu \sim N(0, 1e6),$$
$$\tau^2 \sim IG(1/2, 1 \cdot 3/2),$$
$$\sigma^2 \sim IG(2/2, 2 \cdot 1/2)$$

○ $$y_i \mid \theta_{g_i}, \sigma^2 \overset{\text{ind}}{\sim} N(\theta_{g_i}, \sigma^2), \quad i = 1, \ldots, 53, \quad g_i \in \{1, \ldots, 5\},$$
$$\theta_g \mid \mu \overset{\text{iid}}{\sim} N(\mu, 1), \quad g = 1, \ldots, 5,$$
$$\mu \sim N(0, 1e6),$$
$$\sigma^2 \sim IG(2/2, 2 \cdot 1/2)$$

○ $$y_i \mid \theta_{g_i}, \sigma^2 \overset{\text{ind}}{\sim} N(\theta_{g_i}, \sigma^2), \quad i = 1, \ldots, 53, \quad g_i \in \{1, \ldots, 5\},$$
$$\theta_g \mid \tau^2 \overset{\text{iid}}{\sim} N(0, \tau^2), \quad g = 1, \ldots, 5,$$
$$\tau^2 \sim IG(1/2, 1 \cdot 3/2),$$
$$\sigma^2 \sim IG(2/2, 2 \cdot 1/2)$$

Lesson 11 Part A

Тест, 8 вопроса

4.

Fit the hierarchical model from Question 3 in JAGS and obtain posterior mean estimates for each industry's mean growth (posterior mean for each θ_g).

We are interested in comparing these estimates to those obtained from a model that assumes no hierarchy (the ANOVA cell means model). We can approximate the posterior estimates for the five industry means under a noninformative prior by simply calculating the sample mean growth for the five industries. You can do this in R with:

```
1   means_anova = tapply(dat$y, INDEX=dat$grp, FUN=mean)
2     ## dat is the data read from pctgrowth.csv
```

How do these compare with the estimates from the hierarchical model?

Hint: It might help to plot them with:

```
1   plot(means_anova)
2   points(means_theta, col="red") ## where means_theta are the posterior point
        estimates for the industry means.
```

- ◉ The estimates from the hierarchical model have **less** variability than those from the ANOVA model, tending toward **smaller** magnitudes.

- ○ The estimates from the hierarchical model have **greater** variability than those from the ANOVA model, tending toward **larger** magnitudes.

- ○ The estimates from the hierarchical model have **greater** variability than those from the ANOVA model, tending toward **smaller** magnitudes.

- ○ The estimates from the hierarchical model have **less** variability than those from the ANOVA model, tending toward **larger** magnitudes.

5.

In our hierarchical model for personnel growth, we assumed that the variability between companies within an industry was constant across industries (σ^2 was the same for all industries). Each of the following, except one, presents a reasonable approach to checking this model assumption. Which approach would be less informative than the others?

- ○

Lesson 11 Part A

Fit another model with separate σ^2 parameters for each industry, and calculate the posterior probability that they differ from each other by some specified amount.

Тест, 8 вопроса

○ Calculate the sample variance of growth within the industry for each of the five industries. Check if these variances are similar to each other.

◉ Calculate the posterior probability that $\sigma^2/\tau^2 > 1$ in the original model. If this probability exceeds a pre-determined amount, use a model with separate variance parameters.

○ Fit another model with separate σ^2 parameters for each industry, and compare the DIC values between the new and old models.

| 1 |
| Баллы |

6.

Which of the following would yield the correct calculation of the observation level residual for Company i in the hierarchical model for personnel growth?

○ $y_i - (\theta_{g_i} - \mu)$

○ $\theta_{g_i} - \mu$

◉ $y_i - \theta_{g_i}$

○ $y_i - \mu$

| 1 |
| Баллы |

7.

Suppose we are interested in making a prediction for the growth of an 11th (new) company from Industry 1 (`grp = 1`). Which of the following steps should we follow to simulate from the posterior predictive distribution for this new company's growth? Call it y^*.

○ For each iteration of MCMC:

Draw a sample y^* from a normal distribution with mean μ and variance τ^2 using the current iteration of these parameters.

○ For each iteration of MCMC:

Draw a sample θ_1^* from a normal distribution with mean μ and variance τ^2. Then sample y^* from a normal distribution with mean θ_1^* and variance σ^2 using the current iteration of all parameters.

◉

For each iteration of MCMC:

Lesson 11 Part A

Тест, 8 вопроса Draw a sample y^* from a normal distribution with mean θ_1 and variance σ^2 using the current iteration of these parameters.

○   For each iteration of MCMC:

   Draw a sample y^* from a normal distribution with mean μ and variance $\tau^2 + \sigma^2$ using the current iteration of these parameters.

| 1 |
| Баллы |

8.

Suppose we are interested in making a prediction for the growth of a new company from a new industry which so far has not been observed. Which of the following steps should we follow to simulate from the posterior predictive distribution for this new company's growth? Call it y^*.

◉   For each iteration of MCMC:

   Draw a sample θ^* from a normal distribution with mean μ and variance τ^2. Then sample y^* from a normal distribution with mean θ^* and variance σ^2, using the current iteration of all parameters.

○   For each iteration of MCMC:

   Draw a sample y^* from a normal distribution with mean μ and variance σ^2 using the current iteration of these parameters.

○   For each iteration of MCMC:

   Draw a sample y^* from a normal distribution with mean μ and variance τ^2 using the current iteration of these parameters.

○   For each iteration of MCMC:

   Draw a sample y^* from a normal distribution with mean θ_1 and variance σ^2 using the current iteration of these parameters.

| Оплатить курс |

Lesson 11 Part A

Тест, 8 вопроса