MA556, Applied Bayesian Statistics   Spring 2018

Mini-Project   March 13, 2018

This is a mini-project of a very interesting application on the analysis of data obtained from indirect questioning or randomized response. The project is worth 15% of the course.

The project is about how to elicit answers to sensitive questions and how to analyze the data using Bayesian methods. Three students in our class have already collected the data that we are going to analyze. The last time I taught this course, I asked the students to collect the data at WPI; so it is a relief that you do not have to collect the data. So, specifically, you will only be involved in the data analysis using the Bayesian paradigm.

When people are asked sensitive questions, there is a tendency for them not to respond or to tell lies if they do. One way to reduce these effects is to use the techniques of randomized response or indirect questioning, and one possible design is to ask an unrelated nonsensitive question in addition to the sensitive question. The respondents are asked to give a honest answer to one of the two questions selected according to a random mechanism, the essential features of the random mechanism being known to the investigator. For example, one tosses a die and if one or six comes up, the respondent must give a honest answer to the sensitive question, and if two, three, four or five comes up, the respondent must give a honest answer to the nonsensitive question. In this way the respondents should be more comfortable to answer the question because the investigator can never know which question the respondents are answering.

An additional twist in this project is to give the respondents an option. Some respondents may think that the sensitive question is not sensitive, and so they can answer it directly (i.e., without using a random mechanism).

In this project, the two possible questions are stated below.

```
                         Sensitive Question

    Have you ever driven a car without a drivers license in the US?

              Nonsensitive Question (unrelated question)

                 Do you like to go to the fitness center?
```

Circle your response. [Yes, No]

The three students (interviewers) were asked to go around campus to find other students (respondents) to answer the questions using a questionnaire. They were asked to take three samples. In the first sample, 50 students were asked the questions, and they were asked to toss a die. If it comes up on 1 or 2, they will answer the sensitive question and if it comes up 2, 3, 4, 5, or 6, they will answer the nonsensitive question. In the second sample, 59 students were asked the questions, and they were asked to toss a die. If it comes up on 1 or 2, they will answer the nonsensitive question and if it comes up 3, 4, 5, or 6, they will answer the sensitive question. In the third sample, 50 students were asked only the nonsensitive question. The students noted the sex, age and driving experience of the respondents, and they were asked to note the answer the question (yes or no) and whether the respondents are answering the sensitive question only.

Data Analysis

The data are presented in domains formed by crossing sex, age and experience, each at two levels. So there are eight domains (e.g., one domain is males younger than 24 years with less than three years experience). In the first two samples the random mechanism is used, but not in the last sample in which only the nonsensitive question is asked. Let $n_s, s = 1, 2, 3$ denote the three sample sizes.

Let π_{i1} and π_{i2} denote respectively the probabilities of a 'yes' of the sensitive question and the nonsensitive question for the i^{th} domain (sex by age by experience). Let γ_i denote the probability that a respondent selects the random mechanism. These parameters are the same for the three samples.

Use the Gibbs sampler to fit this model to the data you have collected. Please show your starting values and explain what to do if the starts are out of range. Note that in the first stage of a hierarchical Bayesian model, there are three independent samples that you need to combine them into a single likelihood. At the second stage, we have

$$\pi_{i1} \mid \mu_1, \tau \overset{ind}{\sim} \text{Beta}\{\mu_1\tau, (1-\mu_1)\tau\}, i = 1, \ldots, \ell,$$

$$\pi_{i2} \mid \mu_1, \tau \overset{ind}{\sim} \text{Beta}\{\mu_2\tau, (1-\mu_2)\tau\}, i = 1, \ldots, \ell,$$

$$\gamma_i \mid \mu_3, \tau \overset{ind}{\sim} \text{Beta}\{\mu_3\tau, (1-\mu_3)\tau\}, i = 1, \ldots, \ell.$$

The π_{i1}, π_{i2} and γ_i are all independent. The joint prior distribution for the hyperparameters is

$$\pi(\mu_1, \mu_2, \mu_3, \tau) = \frac{1}{(1+\tau)^2}, 0 < \mu_1, \mu_2, \mu_3 < 1, \tau > 0.$$

Note that this assumes independence also.

a. Obtain the likelihood function, assuming (a) you do not know which respondents answered the sensitive question without randomization, and (b) you know which respondents answered the sensitive question without randomization (this information is in the data).

b. Use the Gibbs sampler to obtain samples from the joint posterior density. Be sure to use all appropriate MCMC diagnostics.

c. Obtain Rao-Blackwellized kernel density estimators (graphs) for the posterior densities of π_{i1}, π_{i2} and γ_i. Obtain the posterior means, posterior standard deviations and 95% HPD intervals for π_{i1}, π_{i2} and γ_i. [Please make tables.]

Model Assessment

Use a Bayesian cross validation analysis (e.g., the conditional predictive ordinate (CPO)) and posterior predictive p-value to assess model fit. Decide whether the model under fits, over fits or is just right.

Project Report

Write up a neat report that should be no more than five typed pages, including tables and figures, showing the four sections below.

a. Clearly present the objectives of the project. This should include your own views to explain what you think you are doing.

b. Present the technical details. Be sure to motivate why you have done these technical things. Please present all pertinent graphs and tables. Present details about the performance of the Gibbs sampler.

c. Present a detailed set of conclusions showing technical and subject matter results (think about the reasons for the survey).

d. Please explain how this project changes your statistical thinking, and how it might affect your future career.

Good Luck!