A Hierarchical Bayesian Model with Nonresponse

Geunshik Han

Hanshin University, Div of Computer Science and Information Telecommunications
411 Yangsan-dong
Osan, Korea
gshan@hucc.hanshin.ac.kr

1. Introduction

We use a hierarchical Bayesian model to analyze nonresponse data from the National Family Income and Expenditure Survey (NFIES). Since Bayesian approach can incorporate prior information about nonrespondents, the Bayesian method is appropriate for the analysis of nonignorable nonresponse problems. In this study we obtain a full Bayesian approach for multinomial nonignorable nonresponse data

2. National Family Income and Expenditure Survey

The National Family Income and Expenditure Survey (NFIES) has been conducted every five year since 1991 by National Statistical Office to measure an aspect of income and expenditure structure of Korean household. One of the variables of interest in the NFIES is a doctor visit of households. In this survey the average response rate was about 80%. The nonresponses of sample households were imputed assuming the distribution of the respondents and nonrespondents are the same within the same imputation class. However this imputation method may be unreliable, thus there is a need to consider the adjustment by a method other than random imputation. The Bayesian method is discussed as a possible alternative to impute the NFIES nonresponses.

3. A Hierarchical full Bayesian Model.

For each household-urbanization type, an individual k in area i belongs to one of J categories. Category j is defined. $\underline{x}_{ik} = (x_{i1k,}...,x_{ijk,}...,x_{iJk})', i = 1,...,c; j = 1,...J; k = 1,...,n_i.$

For each group, at the first stage, we have $\underline{x}_{ik} \mid \underline{p}_i : \ multinomial(1, \underline{p}_i)$. where

$x_{ijk} = 1$ or 0, for each j =1,...J.

At the second stage, for each group, $\underline{p}_i \mid \underline{m}_3, \boldsymbol{t}_3 : \ Dirichlet(\underline{m}_3 \boldsymbol{t}_3)$.

To make a full Bayesian analysis, we take uniform distribution for each $\underline{m}$. For each t, we take gamma distribution.

Using Bayes theorem, the joint posterior density of all the parameters are constructed for each nonresponse model. We use Markov chain Monte Carlo algorithm to obtain the posterior distribution of $\underline{p}_i$. For the full conditional density (Dirichlet and Beta), we use Gibbs kernal, while joint posterior density we use Metropolis-Hastings algorithm of Nandram (1988). We obtain Metropolis-Hastings samples from the joint posterior densities and use these samples to make posterior inferences about $\underline{p}_i$.

4. Conclusion

We have discussed the problem of nonresponse for the estimation of the doctor's visit proportions. Our model was applied to the NFIES data. We observed 95% credible intervals for each parameters. We found that the nonresponse was reasonably addressed by our model. The MCMC method allowed us to assess the complex structure of the multinomial nonresponse estimation. Our empirical analysis indicate good performance of the model for this data.

REFERENCE

Little R. J. A. and Rubin D. B. (1987), Statistical Analysis with missing data, New York: Wiley.

Nandram, B (1988). A Bayesian Analysis of the three stage Hierarchical Multinomial model. Journal of Statistical Computation and Simulation, 61, 97-126

National Statistical Office of Korea (1996). National Survey of Family Income and Expenditure. Yearly Income, Savings & Liabilities and Durable Goods, Vol 2.

Stasny, E. A. (1991). Hierarchical Models for the Probabilities of a Survey Classification and Nonresponse: An Example from the National Crime Survey . Journal of the American Statistical Association 86, 296-303.