Design consistent small area estimates using Gibbs algorithm for logistic models

Babubhai Shah, Beth Barnwell, Ralph Folsom, Akhil Vaish, Research Triangle Institute
Babubhai Shah, RTI, P. O. Box 12194, Research Triangle Park, NC 27709.

Keywords: small area estimates ; Pseudo Bayes intervals ; Mixed models in sample surveys

ABSTRACT

Noting that the full Hierarchical Bayes solution for logistic mixed model has demonstrated good frequentist properties for small samples, we have derived a survey weighted version of the associated Gibbs algorithm. Survey weights are incorporated in conditional posterior distributions to achieve design consistent asymptotic mean vector and covariance matrices. Proper inverse Wishart priors are used for the covariance matrices. The bias of the estimated fixed effects and covariance components is examined for a three stage cluster sample design. The method has been applied to the National Household Survey on Drug Abuse (NHSDA) to produce age specific small area estimates and associated pseudo-Bayes posterior intervals for the fifty states and the District of Columbia.

1. Introduction

Growing out of the demand for state level drug use, treatment, and treatment need statistics, SAMHSA first contracted with RTI to produce model based statistics combining data from the 1991 through 1993 NHSDA surveys. For the analysis of 1991-1993 data, we used penalized quasi-likelihood (PQL) to estimate fixed and random effects in the logistic mixed model. Some limitations of the PQL approach were noted in the Methodology Report produced by Folsom and Judkins (1997). In the current project a hierarchical Bayes model (Malec et. al. 1993) is used and the fixed and random effects are estimated by Gibbs algorithm. The details of theory and mathematics for applying the full hierarchical Bayes (FHB) model to survey data has been presented by Folsom et. al. (1999) The methodology has been tested with data from 1994-1996 survey. The work in applying it to 1999 data with some improvements is in progress.

In the present paper, the objective is to describe the software. The theory will be presented in brief. The software consists of two procedures GIBBS and GSTAT. The procedure GIBBS simulates the posterior distributions of the fixed and random effects using Gibbs algorithm. The resulting fixed as well as random effects are saved for each cycle on data set. The procedure GSTAT estimates the prevalence rate for each block group for each type of individual within the block group, for each cycle of the simulated fixed and random effects. The weighted averages of the block group estimates provide county or state level estimates of prevalence rates. The estimates over cycles represent the posterior distribution of the prevalence rates; the mean, the variance, and the percentiles of this distribution can be evaluated.

2. Theory

The development has been restricted to binary outcomes with the following basic model:

$$\text{logit}\ [\text{Prob}(y_{aijk}=1\,|\,\eta_i,\nu_{ij})] = x_{aijk}\beta_a + \eta_{ai} + \nu_{aij}$$

is the logistic mixed model for the probability that age group-a member-k of PSU-j in State-i has $y_{aijk}=1$. To specify our GIBBS algorithm we assume that the State and PSU level random effect vectors η_i and v_{ij} with age group specific elements are four variate normal with null mean vectors and 4x4 covariance matrices D_η and D_v respectively. We assign an improper uniform prior to the age specific fixed coefficient vectors β_a, and use proper inverse Wishart priors for D_η and D_v.

To keep the description of the algorithm simple, we present formulae for the unweighted case. The software is general and allows for weights. The weights were incorporated in the kernel density to provide design consistent estimates. The details regarding weights are not necessary to understand the algorithm and are omitted in the following sections. This software does implement formulae with appropriate weight adjustments as discussed in Folsom et. al. (1999).

3. Mathematical Formulae

The basic mixed model is

$$\ln(P_{ijk})-\ln(1-P_{ijk}) = X_{ijk}\beta + Z_{1ijk}\eta_{1i} + Z_{2ijk}\eta_{2ij} \quad (1)$$

where $i = 1, 2, ..., n$ indexes state, $j = 1, 2, ..., r_i$ denotes PSUs within state-i, k=1, 2, ..., m_{ij} depicts responding persons. Other symbols are:

P_{ijk} = The probability that the response y_{ijk} is equal to 1.

X_{ijk} = A row vector of observed predictors for the person-ijk.

β = The vector of unknown fixed effect parameters to be estimated.

Z_{1ijk} = The vector of one zero indicator variables for the state level random effects.

Z_{2ijk} = The vector of one zero indicator variables for the PSU level random effects.

η_{1i} = The vector of state level random effects for the state-i.

η_{2ij} = The random effect for PSU-ij.

We assume that η_{1i} and η_{2ij} are identically distributed multivariate normal variables with the mean vector zero and variance covariance matrices $W_1=T_1^{-1}$ and $W_2=T_2^{-1}$ respectively.

With assumptions outlined above, the joint likelihood of y_{ijk}, η_{1i}, and η_{2ij} can be written as follows:

$$L_0(Y,\eta_1,\eta_2,\beta,W_1,W_2) = \prod_{i=1}^{n}\prod_{j=1}^{r_i}\prod_{k=1}^{m_{ij}}\Big[P(y_{ijk},\beta|\eta_1,\eta_2)\, \Phi(\eta_{1i}|W_1)\, \Phi(\eta_{2ij}|W_2) \Big] \quad (2)$$

where

$$P(y_{ijk}=1,\beta|\eta_1,\eta_2) = \{1+\exp(X_{ijk}\beta + Z_{1ijk}\eta_{1i} + Z_{2ijk}\eta_{2ij})\}^{-1},$$

$$P(y_{ijk}=0,\beta|\eta_1,\eta_2) = 1 - P(y_{ijk}=1,\beta|\eta_1,\eta_2),$$

$$\Phi(\eta|W) = (2\pi)^{-p/2}|W|^{-1/2}\exp(-\frac{1}{2}\eta' W^{-1}\eta),$$

where $|W|$ is the determinant of the matrix W. The marginal likelihood function is obtained by integration over η_{1i} and η_{2ij}

$$L(Y,\beta,W_1,W_2) = \prod_{i=1}^{n}\int\prod_{j=1}^{r_i}\prod_{k=1}^{m_{ij}} [P(y_{ijk},\beta|\eta_1,\eta_2)\,\Phi(\eta_{1i}|W_1)\,\Phi(\eta_{2ij}|W_2)]\,\delta\eta_{1i}\delta\eta_{2ij}. \quad (3)$$

The integral in Equation (3) does not have an analytic solution. Maximum likelihood estimates can be obtained through numerical integration. The integral's dimension is equal to the number of random effects, and a practically useful approach is unavailable. In the next section we describe a Bayesian approach to the above problem.

4. Bayesian Approach.

For a Bayesian approach, we shall assume an improper uniform prior for the age specific fixed coefficient vectors β_a. For $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$, we assume that the p_r by p_r matrix $\boldsymbol{W}_i^{-1}$ has a nonsingular Wishart distribution with degrees of freedom $\nu_i > p_i$ and the expected value of $\nu_i \boldsymbol{D}_i$ is given by the density function

$$f_w(\boldsymbol{W}_r^{-1}, \boldsymbol{D}_r, \nu_r) = K|\boldsymbol{D}_r|^{-\nu_r/2} \, |\boldsymbol{W}_r^{-1}|^{(\nu_r - p_r - 1)/2} \exp\{-\frac{1}{2} tr(\boldsymbol{D}_r^{-1} \boldsymbol{W}_r^{-1})\} \tag{4}$$

where the value of the constant is

$$K^{-1} = 2^{\nu_r p_r/2} \pi^{p_r(p_r-1)/4} \prod_{s=1}^{p_r} \Gamma\{(\nu_r + 1 - s)/2\}.$$

Assuming that prior distributions of β, $\boldsymbol{W}_1$, and $\boldsymbol{W}_2$ are independent of each other, the posterior density function of β, $\boldsymbol{W}_1$, and $\boldsymbol{W}_2$ conditional on the observed values of y is

$$f(\beta, \boldsymbol{W}_1, \boldsymbol{W}_2) = A \int L_0(\boldsymbol{Y}, \eta_1, \eta_2, \beta, \boldsymbol{W}_1, \boldsymbol{W}_2) \, \Phi(\beta | \beta_0, \alpha \boldsymbol{I}) \, f_w(\boldsymbol{W}_1, \boldsymbol{D}_1) \, f_w(\boldsymbol{W}_2, \boldsymbol{D}_2) \, \delta\eta_{1i} \delta\eta_{2ij} \tag{5}$$

where

$$A^{-1} = \int L_0(\boldsymbol{Y}, \eta_1, \eta_2, \beta, \boldsymbol{W}_1, \boldsymbol{W}_2) \, \Phi(\beta | \beta_0, \boldsymbol{I}) \, f_w(\boldsymbol{W}_1, \boldsymbol{D}_1) \, f_w(\boldsymbol{W}_2, \boldsymbol{D}_2) \, \delta\eta_{1i} \, \delta\eta_{2ij} \, \delta\beta \, \delta\boldsymbol{W}_1 \, \delta\boldsymbol{W}_2. \tag{6}$$

The numerical evaluation of the function f in Equation (5) is intractable. A Monte Carlo approach to study the posterior distribution of β, η_1, η_2, $\boldsymbol{W}_1$, and $\boldsymbol{W}_2$ is a viable alternative.

5. Monte Carlo method

The approach consists of simulating the joint distribution of the five variables β, η_1, η_2, $\boldsymbol{W}_1$, and $\boldsymbol{W}_2$ for given the data for Y, X, $\boldsymbol{Z}_1$, and $\boldsymbol{Z}_2$. A direct simulation of the joint distribution is not feasible. However, it is possible to generate any one of the five random variables conditional on knowing the other four. If we repeat the process of generating each five variables β, η_1, η_2, $\boldsymbol{W}_1$, and $\boldsymbol{W}_2$ conditional on the known values of the other four, then the Gibbs sampler method suggests the random distribution will converge to the joint distribution of the five variables. For more explanation of the Gibbs sampler, one may refer to Casella and George (1992).

In the following sections, we shall consider the generation of a conditional random variable for each of the five variables. To generate the sample, we plan to use the Metropolis algorithm, which requires a conditional density function for each of the variables. From Equations (2) and (5), we can deduce that the joint distribution for all five of the variables is the product of the following five terms except for constant that does not depend on any of the five variables. We present the natural logarithm for each of the terms:

$$T_1 = \sum_{i=1}^{n} \sum_{j=1}^{r_i} \sum_{k=1}^{m_{ij}} \ln\{P(y_{ijk}, \beta, \eta_1, \eta_2)\}, \tag{7}$$

$$T_2 = \frac{1}{2}\{\ln(|\boldsymbol{W}_1^{-1}|) - \sum_{i=1}^{n} (\eta_{1i}' \boldsymbol{W}_1^{-1} \eta_{1i})\}, \tag{8}$$

$$T_3 = \frac{1}{2}\{\sum_{i=1}^{n} r_i \ln(|W_2^{-1}|) - \sum_{i=1}^{n}\sum_{j=1}^{r_i} (\eta_{2ij}' W_2^{-1} \eta_{2ij})\}, \tag{9}$$

$$T_4 = \frac{1}{2}\{(\nu_1 - p_1 - 1)\ln(|W_1|) + \nu_1 \ln(|D_1^{-1}|) - tr(D_1^{-1} W_1^{-1})\}, \tag{10}$$

$$T_5 = \frac{1}{2}\{(\nu_2 - p_2 - 1)\ln(|W_2|) + \nu_2 \ln(|D_2^{-1}|) - tr(D_2^{-1} W_2^{-1})\}, \tag{11}$$

where the prior distribution for W_1 and W_2 is Wishart with ν_1 and ν_2 degrees of freedom and parameters D_1 and D_2, respectively.

The estimating equations for β are

$$\sum_{ijk} (y_{ijk} - \pi_{ijk})\, x_{ijk}\, \hat{\beta} = 0, \tag{12}$$

$$[V(\hat{\beta})]^{-1} = \sum_{jk} \pi_{ijk}(\hat{\beta})[(1 - \pi_{ijk}(\hat{\beta})] x_{ijk} x_{ijk}' . \tag{13}$$

Because sample size for estimating $\hat{\beta}$ is large, we assume that the conditional distribution for β is normal with mean $\hat{\beta}$ and variance equal to $V(\hat{\beta})$. For estimating random effects η_{1i} and η_{2ij}, we apply a similar approach and the resulting equations for η_{1i} are:,

$$\sum_k [y_{ijk} - \pi_{ijk}(\hat{\eta}_{1i})]\, x_{ijk}\, \hat{\eta}_{1i} + W_1^{-1} (\hat{\eta}_{1i}) = 0, \tag{14}$$

$$[V(\hat{\eta}_{1i})]^{-1} = \sum_{jk} \pi_{ijk}(\hat{\eta}_{1i})[(1 - \pi_{ijk}(\hat{\eta}_{1i})] x_{ijk} x_{ijk}' + W_1^{-1}. \tag{15}$$

$$L_C(\eta_{1i}) = \sum_k \pi_{ijk}(\hat{\eta}_{1i}) x_{ijk} \eta_{1i} + \sum_k (\ln[1 - \pi_{ijk}(\eta_{1i})] + (\eta_{1i} - \hat{\eta}_{1i})' W_1^{-1} (\eta_{1i} - \hat{\eta}_{1i}), \tag{16}$$

The resulting equations for η_{2ij} are:

$$\sum_k [y_{ijk} - \pi_{ijk}(\hat{\eta}_{2ij})]\, x_{ijk}\, \hat{\eta}_{2ij} + W_2^{-1} (\hat{\eta}_{2ij}) = 0, \tag{17}$$

$$[V(\hat{\eta}_{2ij})]^{-1} = \sum_k \pi_{ijk}(\hat{\eta}_{2ij})[(1 - \pi_{ijk}(\hat{\eta}_{2ij})] x_{ijk} x_{ijk}' + W_2^{-1}. \tag{18}$$

$$L_C(\eta_{2ij}) = \sum_k \pi_{ijk}(\hat{\eta}_{2ij}) x_{ijk} \eta_{2ij} + \sum_k \ln[1 - \pi_{ijk}(\eta_{2ij})] + (\eta_{2ij} - \hat{\eta}_{2ij})' W_2^{-1} (\eta_{2ij} - \hat{\eta}_{2ij}). \tag{19}$$

We assume that the conditional distribution for β to be normal, because the total sample size is large. However, we cannot assume normality for η, because the sample size for the state or PSU is not large. The exact conditional probability is given by equations (16) ans (19), and we apply Metropolis-Hastings algorithm to draw a random observation of η.

To derive the conditional distribution of W_1 and W_2, using the well known relationship between multi variate normal and Wishart distribution, we note that the distribution of

$$S_1 = \sum_{i=1}^{n} \eta_{1i} \eta_{1i}', \tag{20}$$

has a Wishart distribution with degrees of freedom equal to n and parameter W_1 given by

$$f(S_1, n, W_1) = K|W_1|^{-n/2}\, |S_1|^{(n-p_1-1)/2} \exp\{-\frac{1}{2} tr(W_1^{-1} S_1)\}, \tag{21}$$

hence the conditional distribution of W_1^{-1} given S_1 is a Wishart distribution with $n+p_1+1$ degrees of freedom and parameter S_1^{-1}. The posterior distribution of W_1^{-1} is a Wishart distribution with parameters $n+p_1+\nu_1+1$ degrees of freedom and parameter $(S_1+D_1)^{-1}$.

$$S_2 = \sum_{i=1}^{n} \sum_{j=1}^{r_i} \eta_{2ij}\eta_{2ij}^{\prime}, \qquad (22)$$

Similarly, it can be shown that the posterior distribution of S_2 is a Wishart distribution with $[\sum_{i=1}^{n} r_i]$ degrees of freedom. Hence, the posterior distribution of W_2^{-1} is a Wishart distribution with parameters $[\sum_{i=1}^{n} r_i]+p_2+\nu_2+1$ degrees of freedom and parameter $(S_2+D_2)^{-1}$.

6. PROC GIBBS

The objective of this method is to generate random values from a joint distribution function which may not be known but the conditional distributions of each set of variables given the other sets of variables are known. For the weighted data with the logistic regression model, we shall use equations (12) through (22).

The steps for each iterative cycle are given below. Fixed effects require three steps:

1) Estimate $\hat{\beta}$ using Equation (12).

2) Compute variance of $\hat{\beta}$ from Equation (13).

3) Generate a random β from a normal distribution with mean $\hat{\beta}$ and variance $V(\hat{\beta})$.

Each random effect requires four steps:

4) Estimate $\hat{\eta}$ using Equation (14) or (17).

5) Compute variance of $\hat{\eta}$ from Equation (15) or (18).

6) Generate a random η from a normal distribution with mean $\hat{\eta}$ and variance $V(\hat{\eta})$.

7) Apply rejection method (Metropolis), with $L^*=L_C(\eta)$ of Equation (16) or (19).

After steps 4 through 7 have been performed for each subscript of η.

8) Compute S, and generate random W^{-1} from a Wishart distribution.

The steps 4 through 8 are repeated for each random effect. All of the above steps are repeated for each iterative cycle by using the random values generated in previous cycle. The computer simulation ends only when desired number of cycles are completed.

7. User options and limitations in PROC GIBBS

The current program is limited to fitting only logistic models to binomial outcome variables. The user has an option to specify different models for fixed effects for each level of a categorical level. For example, if the variable for age groups has four levels then user may specify four different model statements, one corresponding to each level of the age group.

The models for random effects are limited to nested design structure, and the user may specify a random model for each of the nest levels. Of course, each model for random or fixed effects may contain any number of independent variables.

The Gibbs sampling requires many cycles for convergence, so the user can specify the number of cycles before starting to save sample results. The user may also specify the rate at which the parameters are updated; for example, the user may update fixed effect coefficients every eighth cycle, may update state level random effect every even cycle, and PSU level random effect on each cycle.

Since, the Gibbs sampler may require many cycles to converge and each cycle takes many seconds, there is an option to save the entire work space and restart the iterative cycles from the last cycle of the previous run.

The final output of PROC GIBBS is a random sample from the posterior distribution of the fixed effects (β), random effects (η), and variance covariance matrices (W).

7. PROC GSTAT

The input to PROC GSTAT is the data for each geographical block group and the output from PROC GIBBS. For each random sample of β, η, *and* W, PROC GSTAT estimates the prevalence rate for each sub group of population (such as white males 25 to 45 year old) in each block group:

$$\hat{P}_{gijkr} = \frac{\exp\left(X_{gijk}\beta_r + Z_{1gijk}\eta_{1ir} + Z_{2gijk}\eta_{2ijr}\right)}{1 + \exp\left(X_{gijk}\beta_r + Z_{1gijk}\eta_{1ir} + Z_{2gijk}\eta_{2ijr}\right)},$$

where $\hat{P}_{gijkr}$ represents the estimated probability for the subgroup-g, in the block-k in the PSU-j of the state-i, for the cycle-r; β_r, η_{1ir}, *and* η_{2ijr} are the simulated values for the cycle-r of PROC GIBBS; and X_{gijk}, Z_{1gijk}, *and* Z_{2gijk} are the corresponding independent variables. For the PSU or state if there is no observations in the sample then random set of η_{1ir}, *or* η_{2ijr} are generated.

The state level estimates are obtained by taking the weighted average over the block groups:

$$\hat{P}_{gir} = \frac{\sum_j \sum_k w_{gijkr}\, \hat{P}_{gijkr}}{\sum_j \sum_k w_{gijkr}}.$$

The estimated $\hat{P}_{gir}$ for various values of r represents a random sample from the posterior distribution of the prevalence rate π_{gi}. From this sample, PROC GSTAT computes mean, variance, percentiles, etc.

PROC GSTAT permits options to select various subsets of cycles to analyze. A user can specify the number of cycles to skip from sample results. The user may also specify the sampling rate and sample size, to be used in the analysis. This will permit user to compare statistics over different groups of cycles such as results from first thousand as compared those from the last thousand from a data containing 10,000 records.

8. Magnitude of the task

RTI's GIBBS procedure has been designed to handle large data sets as well as models with many parameters. PROC GIBBS was able to fit four age group specific models simultaneously, on 70,000 observations with 30 to 40 independent variables in each model totaling over 100 fixed effects and more than 350 random effects. PROC GIBBS completed over 10,000 Gibbs cycles in 10 hours, using Win95 operating system on a personal computer with 400 MHZ Pentium II (see Table 1) . We were unable to get BUGS or MLwiN to complete even a single cycle on such a large model.

Since we used the option to update fixed parameter after every eighth cycle, we only considered every eighth cycle from the 10,000 cycles or 1,250 cycles as representing the random sample from the posterior distribution of the parameters. For estimating the prevalence rates in each of the small areas, we considered 32 age-race-gender specific profiles for each of the 226,000 block groups, estimated the probability based on the logistic model based the parameters from each of the 1,250 cycles. The frequency or the estimated population in each of the 32 age-race-gender specific group with in a block was used as weight. The weighted average of the these estimates over the block groups within a state, produced a sample 1,250 observations from the posterior distribution of prevalence rates for the state. We then estimated mean, median and various percentiles and other properties of the posterior distribution.

Table 1. Magnitude of tasks

GIBBS

Sample size	70,000
States	51
PSUs	300
No. of models	4
Fixed effects	120
Random effects	1,400
Cycles	10,000
Time in hours[1]	10 to 12

GSTAT

Block Groups	226,000
Age-Race-Gender groups	32
States	51
Fixed effects	120
Random effects	1,400
Cycles	1,250
No. of P-values	9 Billion
Time in hours[2]	10 to 12

[1]Time is for computers with 300 MHz Pentium with 256M memory

For large states with many observations, PROC GIBBS and PROC GSTAT produced final results that were close to the design weighted consistent estimates obtained from PROC DESCRIPT of SUDAN software. The initial results have been promising and we plan to make further improvements.

References

Breslow, N. E. and Clayton, D. C. (1993). "Approximate Inference in Generalized Linear Mixed Models." Journal of the American Statistical Association, 88, 9-25.

Folsom, R.E. and Judkins, D.R. (1997). Substance Abuse in States and Metropolitan Areas: Model Based Estimates from the 1991-1993 National Household Survey on Drug Abuse--Methodology Report. Prepared for the Substance Abuse and Mental Health Services Administration.

Folsom, R.E., Shah, B.V., and Vaish, A. K. (1999). "Substance abuse in states: A methodological report on model based estimates from the 1994-1996 national household surveys on drug abuse," Proceedings of the section on survey research methods, American statistical Association, 371-375.

Malec, D., Sedransk, J., and Tompkins, L. (1993). "Bayesian Predictive Inference for Small Areas for Binary Variables in the National Health Interviews Survey," in Case Studies in Bayesian Statistics: eds. C. Gatsonis, J.S. Hodges, R.E. Kass, and N.D. Singpurualla, New York: Springer-Verlag, 377-389.

Casella, G., and George, E. I., (1992). "explaining Gibbs Sampler," The American Statistician 46, 167-174