Bayesian Logistic Regression Model for Sub-Areas

Lu Chen* Balgobin Nandram†

Abstract

Many population-based surveys have binary responses from a large number of individuals in each household within small areas. An example is the Nepal Living Standards Survey (NLSS II), in which health status binary data (good versus poor) for each individual from sampled households (sub-areas) are available in sampled wards (small areas). To make inference for the finite population proportion of individuals in each household, we use the sub-area logistic regression model with reliable auxiliary information. The contribution of this model is twofold. First, we extend an area-level model to a sub-area level model. Second, because there are numerous sub-areas, standard Markov chain Monte Carlo (MCMC) methods to find the joint posterior density are very time consuming. Therefore, we provide a sampling-based method, the integrated nested normal approximation (INNA), which permits fast computation. Our main goal is to describe this hierarchical Bayesian logistic regression model and to show that the computation is much faster than the exact MCMC method and also reasonably accurate. The performance of our method is studied by using NLSS II data. Our model can borrow strength from both areas and sub-areas to obtain more efficient and precise estimates. The hierarchical structure of our model captures the variation in the binary data reasonably well.

Key Words: Hierarchical Bayesian Model, Integrated nested normal approximation, MCMC, Metropolis sampler, Numerical integration

1. Introduction

The Nepal Living Standard Survey (NLSS) II is a two-stage stratified sampling. A random sample of wards (areas) were selected from six strata and 12 household (sub-areas) were selected from each sampled ward. All individual in sampled household were interviewed. One interest is health status, a binary variable. To make smooth estimates of the finite population proportion of individuals with good health in each household, we focus on hierarchical Bayesian (HB) models with sub-area random effects to obtain reliable "indirect" estimates for numerous small areas or sub-areas. Most of the sample surveys are designed to provide reliable "direct" estimates of interests for large areas or domains (e.g. state level, national level). However, direct estimates are not reliable for areas or domains for which only small samples or no sample are available.

*Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA, 01609
†Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA, 01609

Due to the NLSS II data hierarchical structure, we are particular interested in small area models that can capture such property. Although the one-fold basic models are very popular and in common use in producing reliable estimates, the hierarchical structure of the data and the consistency between the estimates for different levels may not hold. In particular, the sampling design of many population-based survey are two-stage stratified sampling as NLSS II. But if we use one-fold unit level model to fit the data, sub-area level effects would be ignored. Yan and Sedransk (2007) studied the case that the data follow a normal model with a two-stage (three-stage) hierarchical structure while the fitted model has a one-stage (two-stage) hierarchical structure by using posterior predictive p-values. Yan and Sedransk (2010) discussed the ability to detect a three-stage model when a two-stage model is actually fitted.

Two-fold models are an important extension of basic small area models. Many authors have considered the problems and proposed such kinds of models. But most of them are for continuous data. Fuller and Goyenche (1998) proposed a sub-area level model which provides model-based estimates that account for the hierarchical structure of data. Two-fold sub-area level models was studied by Torabi and Rao (2014), Rao (2015). This is an area-level model which extend the Fay-Harriot model to sub-area level. Two-fold nested error regression models was considered by Stukel and Rao (1997,1999).

Bayesian logistic regression models with random effects are suitable to handle binary data with covariates. Nandram (1989) discussed discrimination between the logit and the complementary log-log link functions by using logistic regression model. Roberts, Rao and Kumar (1987) discussed logistic regression for sample survey data (not small area estimation). Nandram and Chen (1996) show how to accelerate the Gibbs sampler for a model with latent variables introduced earlier by Albert and Chib (1993) for Bayesian probit analysis. Farrell, MacGibbon and Tomberlin (1997) discussed logistic regression model by using empirical Bayesian approach. Nandram and Erhardt (2005) showed how to analyze binary data with covariates to maintain conjugacy for both logistic and Poisson regression model. The analysis of binary data with covariates under nonignorable nonresponse was discussed by Nandram and Choi (2010). Nandram, Chen, Shu and Binod (2018) proposed a hierarchical Bayesian logistic regression model for binary data in small area estimation. Such model is a unit level model without sub-area effect. Our two-fold sub-area model is an extension of this logistic regression model. We add sub-area level random effect into the model which can capture the hierarchical structure of the sampled data. In the same time, we add more hyper-parameters into the model which make the inference more complicated. However, we propose an approximation method called the integrated nested normal approximation (INNA) which solved the difficulties.

The other side of our application is that there are numerous small areas (households and individuals) and MCMC methods cannot handle them efficiently which involve complicated integrals. Scott et al. (2013) defined big data as data that are too big to comfortably process on a single machine. They considered consensus Monte Carlo methods that split the data to several machines. They proposed algorithms that perform distributed approximate Bayesian analyses in order to minimize the communication between computers. The parallel MCMC methods for non-Gaussian posterior distributions was discussed by Miroshnikov and Colon (2015). Fortunately, in survey sampling the design generally uses stratification which is not artificial, and in this case, consensus Monte Carlo may not be needed; it will be a good idea for a large stratum.

The integrations involved in Bayesian inference are usually intractable which is true for our logistic regression model. The approximation techniques are desired. The procedure we use to approximate the posterior density of the parameters of logistic regression sub-area model, INNA, is similar to the integrated nested Laplace approximation (INLA) originally proposed by Rue, Martino and Chopin (2009), but they are different actually. INLA is a quite popular algorithm and an alternative to MCMC for big data analysis if the joint posterior density is very complicated. It requires posterior modes and, for numerous small areas, computation of modes becomes time-consuming and challenging for logistic regression model or any generalized linear mixed models. Yet INLA has found many useful applications, such as on Poisson regression by Fong, Rue and Wakefield (2010), and on spatial point pattern data by Illian, Srbye and Rue (2012). We note that INLA can be problematic especially for logistic and Poisson hierarchical regression models, even if the modes can be computed. Ferkingstad and Rue (2015), attempting to improve INLA, used a copula-based correction which adds complexity to INLA. Our approximation method, INNA, which does not require to find posterior modes, uses a sampling-based procedure accommodated by the multiplication rule of probability. Instead of finding the posterior modes, INNA finds approximate modes in closed form, facilitated by the empirical logistic transform (Cox and Snell 1972) and the second-order Taylor series approximation.

On the other hand, two-fold models can capture the heterogeneity between samples within not only areas but also sub-areas. Many model-based estimation techniques for the sampling variances have been considered in the literature, but most of them for the area-level model: see Wang and Fuller (2003), You and Chapman (2006) and Erciulescu and Berg (2014). Nandram and Chen (2016) studied a Bayesian model under heterogeneous sampling variance as log-linear structure, which is preferable than homogeneous model.

In section 2, we give a full description of a sub-area HB logistic regression model. In particular, we describe the integrated nested normal approximation (INNA) and some theoretical results are provided. We put the exact MCMC method discussion in appendix. In section 3, we apply our model to the NLSS II data to provide smoothed estimates of the household proportions of members in good health for both sampled and nonsampled households. Finally, in section 4, we make concluding marks and future work for my remaining research.

2. Sub-Area Hierarchical Logistic Regression Model

In this section, we assume that reliable auxiliary information are available at unit level. The model and method we proposed for many small areas and sub-areas is not only for our application NLSS II. It can be applied to other population-based survey with binary response. In our application, we have binary data (good health versus poor health) for each individual within a household, and these households are within wards.

We have a finite population of L small areas (wards) and within the i^{th} area, there are N_i sub-areas (households). Within the j^{th} sub-area there are M_{ij} individuals. We assume that $\ell(< L)$ areas are sampled and a simple random sample of $n_i(< N_i)$ households is taken from the i^{th} area. All individuals in sampled household are sampled. Here we assume the survey weights are the same within all households in each area. Let $y_{ijk},\ k = 1,\ldots,m_{ij}, j = 1,\ldots,n_i, i = 1,\ldots\ldots,\ell$ denote the binary responses. Let $\underset{\sim}{y} = (y_{ijk},\ k = 1,\ldots,m_{ij}, j = 1,\ldots,n_i, i = 1,\ldots\ldots,\ell)'$. Let $y_{ij} = \sum_{k=1}^{m_{ij}} y_{ijk}$ be the number with response 1 and m_{ij} is the total number of people who responsed. Let $\underset{\sim}{x}_{ijk} = (1, x_{ijk1},\ldots,x_{ijkp})'$ be the $(p+1)$ vector with p covariates for individuals and an intercept.

We use P to represent the population proportion and p be the sample proportion. Let p_{ij} be the corresponding sample probability of $y_{ij},\ j = 1,...,n_i,\ i = 1,\ldots,\ell$.

The primary interests are the finite population proportions of households, which are $P_{ij} = \frac{1}{M_{ij}} \sum_{k=1}^{M_{ij}} y_{ijk},\ j = 1,\ldots,N_i,\ i = 1,\ldots\ldots,\ell$ and the finite population proportions of areas which are $P_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \sum_{k=1}^{M_{ij}} y_{ijk},\ i = 1,\ldots\ldots,\ell$.

In the content of logistic regression model, the two-fold hierarchical Bayesian logistic regression model for the sub-area means μ_{ij}:

$$y_{ijk}|\underset{\sim}{\beta}, \nu_i, \mu_{ij} \overset{ind}{\sim} \text{Bernoulli}\left\{\frac{e^{\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}+\nu_i+\mu_{ij}}}{1+e^{\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}+\nu_i+\mu_{ij}}}\right\}, k = 1,\ldots,m_{ij},$$

$$\mu_{ij}|\sigma^2 \overset{iid}{\sim} \text{Normal}(0,\sigma^2), j = 1,\ldots,n_i,$$

$$\nu_i|\delta^2 \overset{iid}{\sim} \text{Normal}(0,\delta^2), i = 1,\ldots,\ell,$$

$$\pi(\underset{\sim}{\beta}, \delta^2, \sigma^2) \propto \frac{1}{(1+\delta^2)^2}\frac{1}{(1+\sigma^2)^2}, \delta^2 > 0,\ \sigma^2 > 0.$$

Here, $\mu_{ij},\ \ i = 1,\ldots,\ell; j = 1,\ldots,n_i$ are the sub-area level random effect, which is not in the area-level model in Nandram and Chen (2018). $\nu_i,\ i = 1,\ldots,\ell$ are the area random effects and $\underset{\sim}{\beta} = (\beta_0,\beta_1,\ldots,\beta_p)'$ are the regression coefficients with σ^2, δ^2, the variance of the random effects, respectively.

In order to apply our approximation method and make inference for posterior distribution, we use an equivalent model. First, we seperate $\underset{\sim}{\beta}$ into β_0 and $\underset{\sim}{\beta}_{(0)}$, where $\underset{\sim}{\beta}_{(0)} = (\beta_1,\beta_2,...,\beta_p)^T$. We set β_0 as the mean of $\underset{\sim}{\nu}$, and then we can omit intercept term from the covariate $\underset{\sim}{x}_{ijk}$. Second, we introduce a new parameter $w_{ij} = \nu_i + \mu_{ij}$ in order to set ν_i and μ_{ij} independent and then easy to make inference on both of them. We have

$$y_{ijk}|\underset{\sim}{\beta}_{(0)}, w_{ij} \overset{ind}{\sim} \text{Bernoulli}\left\{\frac{e^{\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}_{(0)}+w_{ij}}}{1+e^{\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}_{(0)}+w_{ij}}}\right\}, k = 1,\ldots,m_{ij},$$

$$w_{ij}|\nu_i,\sigma^2 \overset{ind}{\sim} \text{Normal}(\nu_i,\sigma^2), j = 1,\ldots,n_i,$$

$$\nu_i|\beta_0,\delta^2 \overset{iid}{\sim} \text{Normal}(\beta_0,\delta^2), i=1,\ldots,\ell,$$
$$\pi(\underset{\sim}{\beta},\delta^2,\sigma^2) \propto \frac{1}{(1+\delta^2)^2}\frac{1}{(1+\sigma^2)^2}, \delta^2>0,\ \sigma^2>0.$$

The joint posterior density for the parameters is

$$\pi(\underset{\sim}{\nu},\underset{\sim}{w},\underset{\sim}{\beta},\sigma^2,\delta^2|\underset{\sim}{y}) \propto \prod_{i=1}^{\ell}\prod_{j=1}^{n_i}\prod_{k=1}^{m_{ij}}\left[\frac{e^{(\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}_{(0)}+w_{ij})y_{ijk}}}{1+e^{\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}_{(0)}+w_{ij}}}\right]\times\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{-\sum_{i=1}^{l}\sum_{j=1}^{ni}\frac{(w_{ij}-\nu_i)^2}{2\sigma^2}\right\}$$
$$\times\left(\frac{1}{\sqrt{2\pi\delta^2}}\right)^l \exp\left\{-\sum_{i=1}^{l}\frac{(\nu_i-\beta_0)^2}{2\delta^2}\right\}\frac{1}{(1+\sigma^2)^2}\frac{1}{(1+\delta^2)^2}.$$

The posterior density is a non-standard density, and there are difficulties in fitting it using MCMC methods, more so when $n_i,\ m_{ij}$ are large. This motivates our approximate methods.

3. Integrated Nested Normal Approximation Method

In this section we discuss our approximation method INNA and construct the approximate joint posterior density $\pi_a(\underset{\sim}{\nu},\underset{\sim}{w},\underset{\sim}{\beta},\sigma^2,\delta^2|\underset{\sim}{y})$. INNA method is not required to find the posterior modes. Due to the large amount of subareas, it would be time consuming to find all posterior modes so that is why we did not choose the popular INLA method here.

Notice that the joint posterior density of $\pi(\underset{\sim}{\nu},\underset{\sim}{w},\underset{\sim}{\beta},\sigma^2,\delta^2|\underset{\sim}{y})$ is very complicated and it is the logit expit part $\prod_{i=1}^{\ell}\prod_{j=1}^{n_i}\prod_{k=1}^{m_{ij}}\left[\frac{e^{(\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}_{(0)}+w_{ij})y_{ijk}}}{1+e^{\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}_{(0)}+w_{ij}}}\right]$ that causes the difficulties. Therefore, we find a method to approximate this term to normal density functions by using Laplace approximation, the second-order multivariate Taylor-series approximation and the empirical logistic transform (ELT).

Let $f(\underset{\sim}{\tau})=e^{h(\underset{\sim}{\tau})}$ denote the density of a vector of parameters $\underset{\sim}{\tau}$. Let $\underset{\sim}{g}$ denote the gradient vector and H the Hessian matrix at some point $\underset{\sim}{\tau}^*$.

Lemma 3.1. Let $h(\underset{\sim}{\tau})$ be a logconcave density function with the parameter $\underset{\sim}{\tau}$. Then, $\underset{\sim}{\tau}$ approximately has a multivariate normal distribution,

$$\underset{\sim}{\tau} \sim \text{Normal}(\underset{\sim}{\tau}^*-H^{-1}\underset{\sim}{g},-H^{-1}).$$

Proof. Simply apply the second-order multivariate Taylor series of $h(\underset{\sim}{\tau})$ at $\underset{\sim}{\tau}^*$ is

$$f(\underset{\sim}{\tau}) \approx f(\underset{\sim}{\tau}^*)+(\underset{\sim}{\tau}-\underset{\sim}{\tau}^*)'\underset{\sim}{g}+\frac{1}{2}(\underset{\sim}{\tau}-\underset{\sim}{\tau}^*)'H(\underset{\sim}{\tau}-\underset{\sim}{\tau}^*).$$

Note that due to the logconcavity of $h(\underset{\sim}{\tau})$, its Hessian Matrix $-H$ is positive defenite, which can be the covariance matrix. Here we also use certain point $\underset{\sim}{\tau}^*$ rather than the mode of $h(\underset{\sim}{\tau})$. So we do not need to find the solution of the gradient vector $\underset{\sim}{g} = 0$. And the term $-H^{-1}\underset{\sim}{g}$ is a correction to $\underset{\sim}{\tau}^*$. □

Starting with a flat prior $\underset{\sim}{\beta}_{(0)}$ and the $\underset{\sim}{w}$, the model is

$$y_{ijk}|w_{ij},\underset{\sim}{\beta}_{(0)} \overset{ind}{\sim} \text{Bernoulli}\left\{\frac{e^{\underset{\sim}{x}'\underset{\sim}{\beta}_{(0)}+w_{ij}}}{1+e^{\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}_{(0)}+w_{ij}}}\right\}, j=1,\ldots,n_i, i=1,\ldots,\ell,$$

$$p(\underset{\sim}{w},\underset{\sim}{\beta}_{(0)}) = 1.$$

The joint posterior density is

$$\pi(\underset{\sim}{w},\underset{\sim}{\beta}_{(0)}|\underset{\sim}{y}) \propto \prod_{i=1}^{\ell}\prod_{j=1}^{n_i}\prod_{k=1}^{m_{ij}}\left\{\frac{e^{(\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}_{(0)}+w_{ij})y_{ijk}}}{1+e^{\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}_{(0)}+w_{ij}}}\right\}. \tag{1}$$

The logarithm of the joint posterior density (or log-likelihood) is

$$\Delta = h(\underset{\sim}{\tau}) = \sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\sum_{k=1}^{m_{ij}}\left\{(\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}_{(0)}+w_{ij})y_{ijk} - \log(1+e^{\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}_{(0)}+w_{ij}})\right\}.$$

Let $\underset{\sim}{\tau}' = (\underset{\sim}{\mu}', \underset{\sim}{\beta}'_{(0)})$. In our method, we find a convenient point to expand the log-likelihood in a second-order multivariate Taylor's series expansion.

To begin with, let $\bar{y}_{ij} = \frac{1}{m_{ij}}\sum_{k=1}^{m_{ij}} y_{ijk}$. We use the empirical logistic transform z_{ij} to get an estimate of w_{ij}, where

$$\hat{w}^*_{ij} = z_{ij} = \log\left\{\frac{\bar{y}_{ij}+\frac{1}{2m_{ij}}}{1-\bar{y}_{ij}+\frac{1}{2m_{ij}}}\right\}, i=1,\ldots,\ell; j=1,\ldots,n_i.$$

First, we discuss how to find the quasi mode of $\underset{\sim}{\beta}_{(0)}$. We plug $\hat{w}^*_{ij}$ into the log likelihood function Δ and consider it as a function of $\underset{\sim}{\beta}_{(0)}$ only as $q(\underset{\sim}{\beta}_{(0)})$, we get

$$q(\underset{\sim}{\beta}_{(0)}) = \sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\sum_{k=1}^{m_{ij}}\left[(\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}_{(0)}+\hat{w}^*_{ij})y_{ijk} - \log(1+e^{\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}_{(0)}+\hat{w}^*_{ij}})\right]$$

The frist derivative of $q(\underset{\sim}{\beta}_{(0)})$ is

$$q'(\underset{\sim}{\beta}_{(0)}) = \sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\sum_{k=1}^{m_{ij}}\left\{\underset{\sim}{x}_{ijk}y_{ijk} - \frac{\underset{\sim}{x}_{ijk}e^{(\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}_{(0)}+\hat{w}^*_{ij})}}{1+e^{\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}_{(0)}+\hat{w}^*_{ij}}}\right\}$$
$$= \sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\sum_{k=1}^{m_{ij}}\left\{\underset{\sim}{x}_{ijk}y_{ijk} - \underset{\sim}{x}_{ijk}\left[1+e^{-(\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}_{(0)}+\hat{w}^*_{ij})}\right]^{-1}\right\}$$

Usually we should set $q'(\underset{\sim}{\beta}_{(0)})$ equal to zero and find the modes as the maximum likelihood estimator(MLE) of $\underset{\sim}{\beta}_{(0)}$. But here it is not easy to solve the equation due to the complexity of $q'(\underset{\sim}{\beta}_{(0)})$. We use the first-order Taylor's series to appoximate and then simplfy $q'(\underset{\sim}{\beta}_{(0)})$ so that we can get quasi modes of $\underset{\sim}{\beta}_{(0)}$.
Since the first-order Talor's expansion of $(1+e^{\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}_{(0)}+\hat{w}^*_{ij}})^{-1}$ equals $(1-e^{-(\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}_{(0)}+\hat{w}^*_{ij})})$. Notice that by Taylor seriers, $e^{-(\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}_{(0)}+\hat{w}^*_{ij})} \approx 1-(\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}_{(0)}+\hat{w}^*_{ij})$. Then we can get

$$q'(\underset{\sim}{\beta}_{(0)}) \approx \sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\sum_{k=1}^{m_{ij}}\left\{\underset{\sim}{x}_{ijk}y_{ijk} - \underset{\sim}{x}_{ijk}\left[(1-e^{\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}_{(0)}+\hat{w}^*_{ij}})\right]\right\}$$
$$\approx \sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\sum_{k=1}^{m_{ij}}\left\{\underset{\sim}{x}_{ijk}y_{ijk} - \underset{\sim}{x}_{ijk}\left[(1-1+(\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}_{(0)}+\hat{w}^*_{ij}))\right]\right\}$$
$$= \sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\sum_{k=1}^{m_{ij}}\left\{\underset{\sim}{x}_{ijk}(y_{ijk}-\hat{w}^*_{ij}) - \underset{\sim}{x}_{ijk}\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}_{(0)}\right\}$$

Then solve for $q'(\underset{\sim}{\beta}_{(0)}) = 0$, we can easily get the quasi modes of $\underset{\sim}{\beta}_{(0)}$

$$\underset{\sim}{\beta}^*_{(0)} = [\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\sum_{k=1}^{m_{ij}}\underset{\sim}{x}_{ijk}\underset{\sim}{x}'_{ijk}]^{-1}[\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\sum_{k=1}^{m_{ij}}\underset{\sim}{x}_{ijk}(y_{ijk}-\hat{w}^*_{ij})].$$

Second, we obtain quasi modes for the w_{ij}, a refinement of the z_i. Plug $\underset{\sim}{\beta}^*_{(0)}$ in the likelihood function Δ and consider it as function w_{ij} only,

$$g(w_{ij}) = \sum_{k=1}^{m_{ij}}\left[(\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}^*_{(0)}+w_{ij})y_{ijk} - \log(1+e^{\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}_{(0)^*}+w_{ij}})\right].$$

Similarly we apply Taylor expansion, we get the approximate first derivative of $g(w_{ij})$

$$
\begin{aligned}
g'(w_{ij}) &= \sum_{k=1}^{m_{ij}} \left\{ y_{ijk} - \left[1 + e^{-(\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}^*_{(0)} + w_{ij})}\right]^{-1} \right\} \\
&\approx \sum_{k=1}^{m_{ij}} \left\{ y_{ijk} - \left(1 - e^{-w_{ij}} e^{-\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}^*_{(0)}}\right) \right\}
\end{aligned}
$$

Solve for $g'(w_{ij}) = 0$, we can obtain the approximate posterior mode of w_{ij},

$$
w^*_{ij} = \log \left\{ \frac{\sum_{k=1}^{m_{ij}} e^{-\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}^*_{(0)}}}{m_{ij}(1 - \bar{y}_{ij})} \right\}
$$

Notice that the term $1 - \bar{y}_{ij}$ in denominator may cause trouble if $\bar{y}_{ij} = 1$ for some is and js. Here we borrow the idea from ELT and make a small adjustment in order to avoid zero denominator. That is,

$$
w^*_{ij} \approx \log \left\{ \frac{\sum_{k=1}^{m_{ij}} e^{-\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}^*_{(0)}}}{m_{ij}(1 - \bar{y}_{ij} + \frac{1}{2m_{ij}})} \right\} i = 1, \ldots, \ell,\ j = 1, \ldots, n_i.
$$

Let $\underset{\sim}{\tau}^{*\prime} = (\underset{\sim}{\mu}^{*\prime}, \underset{\sim}{\beta}^{*\prime}_{(0)})$. Next, we evaluate $\underset{\sim}{g}$ and H at the quasi modes $\underset{\sim}{\tau} = \underset{\sim}{\tau}^*$ can also be obtained as

$$
\underset{\sim}{g} = \begin{pmatrix} \frac{\partial \Delta}{\partial w_{11}} & \cdots & \frac{\partial \Delta}{\partial w_{\ell n_\ell}} & \frac{\partial \Delta}{\partial \underset{\sim}{\beta}_{(0)}} \end{pmatrix}^T_{\underset{\sim}{w} = \underset{\sim}{w}^*,\ \underset{\sim}{\beta}_{(0)} = \underset{\sim}{\beta}^*_{(0)}},
$$

$$
H = \begin{pmatrix}
\frac{\partial^2 \Delta}{\partial w_{11}^2} & \cdots & \frac{\partial^2 \Delta}{\partial w_{11} \partial w_{\ell n_\ell}} & \frac{\partial^2 \Delta}{\partial w_{11} \partial \underset{\sim}{\beta}_{(0)}} \\
\vdots & \vdots & \ddots & \vdots \\
0 & \cdots & \frac{\partial^2 \Delta}{\partial w_{\ell n_\ell}^2} & \frac{\partial^2 \Delta}{\partial w_{\ell n_\ell} \partial \underset{\sim}{\beta}_{(0)}} \\
\frac{\partial^2 \Delta}{\partial w_{11} \partial \underset{\sim}{\beta}_{(0)}} & \cdots & \frac{\partial^2 \Delta}{\partial w_{\ell n_\ell} \partial \underset{\sim}{\beta}_{(0)}} & \frac{\partial^2 \Delta}{\partial \underset{\sim}{\beta}^2_{(0)}}
\end{pmatrix}_{\underset{\sim}{w} = \underset{\sim}{w}^*, \underset{\sim}{\beta}_{(0)} = \underset{\sim}{\beta}^*_{(0)}}.
$$

The partial derivatives can be expressed in terms of response y_{ijk} and covariates $\underset{\sim}{x}_{ijk}$ as

$$
\frac{\partial \Delta}{\partial \underset{\sim}{\beta}_{(0)}} = \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \sum_{k=1}^{m_{ij}} \left\{ \underset{\sim}{x}_{ijk} y_{ijk} - \frac{\underset{\sim}{x}_{ijk} e^{\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}^*_{(0)} + w^*_{ij}}}{1 + e^{\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}^*_{(0)} + w^*_{ij}}} \right\},
$$

$$
\frac{\partial \Delta}{\partial w_{ij}} = \sum_{k=1}^{m_{ij}} (y_{ijk} - \frac{e^{\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}^*_{(0)} + w^*_{ij}}}{1 + e^{\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}^*_{(0)} + w^*_{ij}}}),
$$

$$\frac{\partial^2\Delta}{\partial\underset{\sim}{\beta}_{(0)}^2} = -\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\sum_{k=1}^{m_{ij}} \frac{\underset{\sim}{x}_{ijk}\underset{\sim}{x}'_{ijk}e^{\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}^*_{(0)}+w^*_{ij}}}{(1+e^{\underset{\sim}{x}'_{ij}\underset{\sim}{\beta}^*_{(0)}+w^*_{ij}})^2},$$

$$\frac{\partial^2\Delta}{\partial w_{ij}^2} = -\sum_{k=1}^{m_{ij}} \frac{e^{\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}^*_{(0)}+w^*_{ij}}}{(1+e^{\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}^*_{(0)}+w^*_{ij}})^2},$$

$$\frac{\partial^2\Delta}{\partial\mu_i\partial\underset{\sim}{\beta}_{(0)}} = -\sum_{k=1}^{m_{ij}} \frac{\underset{\sim}{x}_{ijk}e^{\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}^*_{(0)}+w^*_{ij}}}{(1+e^{\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}^*_{(0)}+w^*_{ij}})^2},$$

where $i=1,\ldots,\ell,\ j=1,\ldots,n_i$.

For the convenience of computation, denote $\underset{\sim}{g} = \begin{pmatrix} \underset{\sim}{g}_1 \\ \underset{\sim}{g}_2 \end{pmatrix}$ and $H = -\begin{pmatrix} D & C' \\ C & B \end{pmatrix}$, where

$$\underset{\sim}{g}_1 = \begin{pmatrix} \frac{\partial\Delta}{\partial w_{11}} \cdots \frac{\partial\Delta}{\partial w_{\ell n_\ell}} \end{pmatrix}^T, \underset{\sim}{g}_2 = \frac{\partial\Delta}{\partial\underset{\sim}{\beta}_{(0)}},$$

$$B = -\frac{\partial^2\Delta}{\partial\underset{\sim}{\beta}_{(0)}^2}, C = -\begin{pmatrix} \frac{\partial^2\Delta}{\partial w_{11}\partial\underset{\sim}{\beta}_{(0)}} & \cdots & \frac{\partial^2\Delta}{\partial w_{\ell n_\ell}\partial\underset{\sim}{\beta}_{(0)}} \end{pmatrix}, D = -\begin{pmatrix} \frac{\partial^2\Delta}{\partial w_{11}^2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\partial^2\Delta}{\partial w_{\ell n_\ell}^2} \end{pmatrix}.$$

Note that D and I (identity) are diagonal matrices.

$$\text{Let } -H^{-1} = \begin{pmatrix} D & C' \\ C & B \end{pmatrix}^{-1} = \begin{pmatrix} E & F' \\ F & G \end{pmatrix}, \text{ where}$$

$$E = D^{-1} + D^{-1}C'(B-CD^{-1}C')^{-1}CD^{-1}, F = -(B-CD^{-1}C')^{-1}CD^{-1}, G = (B-CD^{-1}C')^{-1}.$$

Lemma 3.2. Assuming that the design matrix is full-rank and $0 < \sum_{k=1}^{m_{ij}} y_{ijk} < m_{ij},\ j=1,\ldots,n_i;\ i=1,\ldots,\ell$, the posterior density, $\underset{\sim}{\tau}|\underset{\sim}{y}$ in (1), is logconcave.

Proof. If $0 < \sum_{k=1}^{m_{ij}} y_{ijk} < m_{ij}, i=1,\ldots,\ell, j=1,\ldots,n_i$, there are solutions to the gradient vector set to zero.

Let $p_{ijk} = \frac{e^{\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}_{(0)}+w_{ij}}}{1+e^{\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}_{(0)}+w_{ij}}}, k=1,\ldots,m_ij, j=1,\ldots,n_i, i=1,\ldots,\ell$. Then, A, B and C of the negative Hessian matrix can be written as,

$$B = -\frac{\partial^2\Delta}{\partial\underset{\sim}{\beta}_{(0)}^2} = \sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\sum_{k=1}^{m_{ij}} p_{ijk}(1-p_{ijk})\underset{\sim}{x}_{ijk}\underset{\sim}{x}'_{ijk},$$

$$D = \text{diagonal}(d_{ij}), \quad d_{ij} = \frac{\partial^2 \Delta}{\partial w_{ij}^2} = \sum_{k=1}^{m_{ij}} p_{ijk}(1-p_{ijk}),$$

$$C = (\underset{\sim}{c}_{ij}), \quad \underset{\sim}{c}_{ij} = \frac{\partial^2 \Delta}{\partial w_{ij} \partial \underset{\sim}{\beta}_{(0)}} = \sum_{k=1}^{m_{ij}} p_{ijk}(1-p_{ijk}) \underset{\sim}{x}_{ijk},$$

where $j = 1, \dots, n_i, i = 1, \dots, \ell$.

It is obvious that D is positive definite. Thus, to show that $-H$ is positive definite, we need to show that its Schur complement of D, $S = B - CD^{-1}C'$, is positive definite (e.g., see Boyd and Vandenberghe 2004). Let $\omega_{ijk} = p_{ijk}(1-p_{ijk}) / \sum_{k=1}^{m_{ij}} p_{ijk}(1-p_{ijk}), k = 1, \dots, m_{ij}, j = 1, \dots, n_i, i = 1, \dots, \ell$. The Schur complement is

$$S = \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \sum_{k=1}^{m_{ij}} p_{ijk}(1-p_{ijk}) \sum_{k=1}^{m_{ij}} \omega_{ijk} \underset{\sim}{x}_{ijk} \underset{\sim}{x}'_{ijk} - \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \sum_{k=1}^{m_{ij}} p_{ijk}(1-p_{ijk}) \sum_{k=1}^{m_{ij}} \omega_{ijk} \underset{\sim}{x}_{ijk} \sum_{k=1}^{m_{ij}} \omega_{ijk} \underset{\sim}{x}'_{ijk}.$$

It is now easy to show that

$$S = \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \sum_{k=1}^{m_{ij}} \omega_{ijk} (\underset{\sim}{x}_{ijk} - \sum_{k=1}^{m_{ij}} \omega_{ijk} \underset{\sim}{x}_{ijk})(\underset{\sim}{x}_{ijk} - \sum_{k=1}^{m_{ij}} \omega_{ijk} \underset{\sim}{x}_{ijk})'.$$

Therefore, $-H$ is positive definite, and $\underset{\sim}{\tau}|\underset{\sim}{y}$ is logconcave. □

Therefore, according to the Lemma 3.1 and Lemma 3.2, we can establish the approximation Theorem.

Theorem 3.1. Assuming that the design matrix is full-rank and $0 < \sum_{k=1}^{m_{ij}} y_{ijk} < m_{ij},\ j = 1, \dots, n_i,\ i = 1, \dots, \ell$, the posterior density, $\underset{\sim}{\tau}|\underset{\sim}{y}$ in (1) is approximately a multivariate normal density, and the conditional posterior density of $\underset{\sim}{w}|\underset{\sim}{\beta}_{(0)}, \underset{\sim}{y}$ and $\underset{\sim}{\beta}_{(0)}|\underset{\sim}{y}$ can also be approximated by multivariate normal distributions.

Proof. By Lemma 3.2, the posterior density is logconcave. Then according to Lemma 3.1, the posterior distribution $\underset{\sim}{\tau}|\underset{\sim}{y}$ is approximately a multivariate normal distribution.

By Lemma 3.1, evaluating all quantities at $\underset{\sim}{\tau}^*$, the mean is

$$\begin{pmatrix} \underset{\sim}{\mu}_w \\ \underset{\sim}{\mu}_\beta \end{pmatrix} = \underset{\sim}{\tau}^* - H^{-1} \underset{\sim}{g} = \begin{pmatrix} \underset{\sim}{w}^* \\ \underset{\sim}{\beta}^*_{(0)} \end{pmatrix} + \begin{pmatrix} E & F' \\ F & G \end{pmatrix} \begin{pmatrix} \underset{\sim}{g}_1 \\ \underset{\sim}{g}_2 \end{pmatrix} = \begin{pmatrix} \underset{\sim}{w}^* + E\underset{\sim}{g}_1 + F'\underset{\sim}{g}_2 \\ \underset{\sim}{\beta}^*_{(0)} + F\underset{\sim}{g}_1 + G\underset{\sim}{g}_2 \end{pmatrix}.$$

Also, the covariance matrix is

$$-H^{-1} = \begin{pmatrix} D & C' \\ C & B \end{pmatrix}^{-1} = \begin{pmatrix} E & F' \\ F & G \end{pmatrix}.$$

Therefore, by Lemma 3.1, the approximate joint posterior density of $\underset{\sim}{w}, \underset{\sim}{\beta}_{(0)} | \underset{\sim}{y}$ is

$$\begin{pmatrix} \underset{\sim}{w} \\ \underset{\sim}{\beta}_{(0)} \end{pmatrix} | \underset{\sim}{y} \sim \text{Normal}\left\{ \begin{pmatrix} \underset{\sim}{\mu}_w \\ \underset{\sim}{\mu}_\beta \end{pmatrix}, \begin{pmatrix} E & F' \\ F & G \end{pmatrix} \right\}.$$

Finally, using the property of the multivariate normal density, the conditional posterior density of $\underset{\sim}{w}|\underset{\sim}{\beta}_{(0)}, \underset{\sim}{y}$ and $\underset{\sim}{\beta}_{(0)}|\underset{\sim}{y}$ can also be approximated by multivariate normal distributions,

$$\underset{\sim}{w}|\underset{\sim}{\beta}_{(0)}, \underset{\sim}{y} \sim \text{Normal}\{\underset{\sim}{\mu}_w - D^{-1}C'(\underset{\sim}{\beta}_{(0)} - \underset{\sim}{\mu}_\beta), D^{-1}\} \text{ and } \underset{\sim}{\beta}_{(0)}|\underset{\sim}{y} \sim \text{Normal}\{\underset{\sim}{\mu}_\beta, G\},$$

where

$$\underset{\sim}{\mu}_w = \underset{\sim}{w}^* + E\underset{\sim}{g}_1 + F'\underset{\sim}{g}_2 \text{ and } \underset{\sim}{\mu}_\beta = \underset{\sim}{\beta}^*_{(0)} + F\underset{\sim}{g}_1 + G\underset{\sim}{g}_2.$$

□

Therefore, we can approximate that logit expit term $\prod_{i=1}^{\ell}\prod_{j=1}^{n_i}\prod_{k=1}^{m_{ij}}\left[\frac{e^{(\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}_{(0)}+w_{ij})y_{ijk}}}{1+e^{\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}_{(0)}+w_{ij}}}\right]$ into two multivariate density by Theorem 3.1. And then we can get our approximate two-fold Bayesian logistic regression model.

Recall the posterior density of our two-fold logistic model is

$$\pi(\underset{\sim}{w}, \underset{\sim}{v}, \underset{\sim}{\beta}, \sigma^2, \delta^2 \mid \underset{\sim}{y}) \propto \pi(\underset{\sim}{y}|\underset{\sim}{w}, \underset{\sim}{\beta}_{(0)})\pi(\underset{\sim}{w} \mid \underset{\sim}{v}, \sigma^2)\pi(\underset{\sim}{v} \mid \beta_0, \delta^2)\pi(\underset{\sim}{\beta}_{(0)}, \beta_0, \sigma^2, \delta^2)$$

The likelihood function $\pi(\underset{\sim}{y}|\underset{\sim}{w}, \underset{\sim}{\beta}_{(0)})$ can be approximated by the multivariate normal distribution by Theorem 3.1. Combine the prior of $\underset{\sim}{w}$ and $\underset{\sim}{v}$ given by our Bayesian Logistic model and the results in Theroem 3.1, we can obtain our INNA model

$$\underset{\sim}{w}|\underset{\sim}{\beta}_{(0)}, \underset{\sim}{y} \sim \text{Normal}\{\underset{\sim}{\mu}_w - D^{-1}C'(\underset{\sim}{\beta}_{(0)} - \underset{\sim}{\mu}_\beta), D^{-1}\}$$

$$\underset{\sim}{\beta}_{(0)}|\underset{\sim}{y} \sim \text{Normal}\{\underset{\sim}{\mu}_\beta, G\}$$

$$\underset{\sim}{w}|\underset{\sim}{v}, \sigma^2 \overset{ind}{\sim} \text{Normal}(\underset{\sim}{\mu}_v, \sigma^2 I),$$

$$\underset{\sim}{v}|\beta_0, \delta^2 \overset{iid}{\sim} \text{Normal}(\beta_0 \underset{\sim}{j}, \delta^2 I),$$

$$\pi(\underset{\sim}{\beta}_{(0)}, \beta_0, \delta^2, \sigma^2) \propto \frac{1}{(1+\delta^2)^2}\frac{1}{(1+\sigma^2)^2}, \delta^2 > 0,\ \sigma^2 > 0,$$

where $\underset{\sim}{\mu}'_\nu = (\underbrace{\nu_1,\ldots,\nu_1}_{n_1}\cdots\underbrace{\nu_\ell,\ldots,\nu_\ell}_{n_\ell})'$ and $\underset{\sim}{j}$ is a vector of ones.

By Bayes' Theorem and the multiplication rule, the posterior density $\pi(\underset{\sim}{w}, \underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2, \delta^2 \mid \underset{\sim}{y})$ can be approximated by

$$\begin{aligned}
\pi_a(\underset{\sim}{w}, \underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2, \delta^2 \mid \underset{\sim}{y}) &\propto \pi_a(\underset{\sim}{w} \mid \underset{\sim}{\nu}, \underset{\sim}{\beta}_{(0)}, \sigma^2, \underset{\sim}{y})\pi_a(\underset{\sim}{\nu} \mid \beta_0, \delta^2, \underset{\sim}{y})\pi_a(\underset{\sim}{\beta}_{(0)} \mid \underset{\sim}{y})\pi_a(\underset{\sim}{\beta}, \sigma^2, \delta^2 \mid \underset{\sim}{y}) \\
&= e^{-\frac{1}{2}\left\{\left[\underset{\sim}{w}-\left(\underset{\sim}{\mu}_w - D^{-1}C'(\underset{\sim}{\beta}_{(0)}-\underset{\sim}{\mu}_\beta)\right)\right]' D\left[\underset{\sim}{w}-\left(\underset{\sim}{\mu}_w - D^{-1}C'(\underset{\sim}{\beta}_{(0)}-\underset{\sim}{\mu}_\beta)\right)\right]\right\}} \\
&\times e^{-\frac{1}{2}\left\{\left[\underset{\sim}{w}-\underset{\sim}{\mu}_\nu\right]'(\sigma^2 I)^{-1}\left[\underset{\sim}{w}-\underset{\sim}{\mu}_\nu\right]+\left[\underset{\sim}{\nu}-\beta_0\underset{\sim}{j}\right]'(\delta^2 I)^{-1}\left[\underset{\sim}{\nu}-\beta_0\underset{\sim}{j}\right]+\left[\underset{\sim}{\beta}_{(0)}-\underset{\sim}{\mu}_\beta\right]' G^{-1}\left[\underset{\sim}{\beta}_{(0)}-\underset{\sim}{\mu}_\beta\right]\right\}} \\
&\times \frac{|D|^{1/2}}{|\delta^2 I|^{1/2}|\sigma^2 I|^{1/2}|G|^{1/2}}\frac{1}{(1+\sigma^2)^2}\frac{1}{(1+\delta^2)^2} \qquad (2)
\end{aligned}$$

Therefore, we can get the following key result,

Theorem 3.2. Using the multiplication rule, the joint posterior density, $\pi(\underset{\sim}{w}, \underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2, \delta^2 \mid \underset{\sim}{y})$ in (2), can be approximated by

$$\pi_a(\underset{\sim}{w}, \underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2, \delta^2 \mid \underset{\sim}{y}) \propto \pi_a(\underset{\sim}{w} \mid \underset{\sim}{\nu}, \underset{\sim}{\beta}_{(0)}, \sigma^2, \underset{\sim}{y})\pi_a(\underset{\sim}{\nu} \mid \beta_0, \delta^2, \underset{\sim}{y})\pi_a(\underset{\sim}{\beta}_{(0)} \mid \underset{\sim}{y})\pi_a(\underset{\sim}{\beta}, \sigma^2, \delta^2 \mid \underset{\sim}{y}),$$

where the first three densities on the right-hand side are all multivariate normal densities.

Proof. First, look at the exponent terms containing $\underset{\sim}{w}$ in the above approximate posterior density function

$$\begin{aligned}
&\left[\underset{\sim}{w} - \left(\underset{\sim}{\mu}_w - D^{-1}C'(\underset{\sim}{\beta}_{(0)} - \underset{\sim}{\mu}_\beta)\right)\right]' D\left[\underset{\sim}{w} - \left(\underset{\sim}{\mu}_w - D^{-1}C'(\underset{\sim}{\beta}_{(0)} - \underset{\sim}{\mu}_\beta)\right)\right] + \left[\underset{\sim}{w} - \underset{\sim}{\mu}_\nu\right]'(\sigma^2 I)^{-1}\left[\underset{\sim}{w} - \underset{\sim}{\mu}_\nu\right] \\
&= \left[\underset{\sim}{w} - (D + \frac{1}{\sigma^2}I)^{-1}\left(D\underset{\sim}{\mu}_w - C'(\underset{\sim}{\beta}_{(0)} - \underset{\sim}{\mu}_\beta) + \frac{1}{\sigma^2}\underset{\sim}{\mu}_\nu\right)\right]'(D + \frac{1}{\sigma^2}I)\left[\underset{\sim}{w} - (D + \frac{1}{\sigma^2}I)^{-1}\left(D\underset{\sim}{\mu}_w - C'(\underset{\sim}{\beta}_{(0)} - \underset{\sim}{\mu}_\beta) + \frac{1}{\sigma^2}\underset{\sim}{\mu}_\nu\right)\right] \\
&\quad + \left[\underset{\sim}{\mu}_w - D^{-1}C'\left(\underset{\sim}{\beta}_{(0)} - \underset{\sim}{\mu}_\beta\right) - \underset{\sim}{\mu}_\nu\right]'(D^{-1} + \sigma^2 I)^{-1}\left[\underset{\sim}{\mu}_w - D^{-1}C'\left(\underset{\sim}{\beta}_{(0)} - \underset{\sim}{\mu}_\beta\right) - \underset{\sim}{\mu}_\nu\right]
\end{aligned}$$

Then it can show that

$$\underset{\sim}{w}|\underset{\sim}{\nu}, \underset{\sim}{\beta}_{(0)}, \sigma^2, \underset{\sim}{y} \overset{\text{app}}{\sim} \text{Normal}\left\{(D + \frac{1}{\sigma^2}I)^{-1}\left(D\underset{\sim}{\mu}_w - C'(\underset{\sim}{\beta}_{(0)} - \underset{\sim}{\mu}_\beta) + \frac{1}{\sigma^2}\underset{\sim}{\mu}_\nu\right), (D + \frac{1}{\sigma^2}I)^{-1}\right\}$$

which is the $\pi_a(\underset{\sim}{w} \mid \underset{\sim}{\nu}, \underset{\sim}{\beta}_{(0)}, \sigma^2, \underset{\sim}{y})$.

Notice that $(D + \frac{1}{\sigma^2}I)$ is diagonal matrix. Then given $\underset{\sim}{\nu}, \underset{\sim}{\beta}_{(0)}, \sigma^2, \underset{\sim}{y}$, all w_{ij}s are independent.

This is an important result because parallel computation can be done for w_{ij}, which accommodates time-consuming and massive storage challenges in big data analysis. This result holds for the exact conditional posterior density of the μ_{ij}. Since $\underset{\sim}{w}$ has a multivariate normal distribution, we can integrate out $\underset{\sim}{w}$ from the joint approximate posterior density $\pi_a(\underset{\sim}{w},\underset{\sim}{v},\underset{\sim}{\beta},\sigma^2,\delta^2 \mid \underset{\sim}{y})$, and obtain the joint posterior density of $\underset{\sim}{v},\underset{\sim}{\beta},\sigma^2$ and δ^2

$$\begin{aligned}
\pi_a(\underset{\sim}{v},\underset{\sim}{\beta},\sigma^2,\delta^2 \mid \underset{\sim}{y}) \propto\; & e^{-\frac{1}{2}\left\{\left[\underset{\sim}{\mu}_v - D^{-1}C'\left(\underset{\sim}{\beta}_{(0)}-\underset{\sim}{\mu}_\beta\right)-\underset{\sim}{\mu}_w\right]'(D^{-1}+\sigma^2 I)^{-1}\left[\underset{\sim}{\mu}_v - D^{-1}C'\left(\underset{\sim}{\beta}_{(0)}-\underset{\sim}{\mu}_\beta\right)-\underset{\sim}{\mu}_w\right]\right\}} \\
& \times e^{-\frac{1}{2}\left\{\left[\underset{\sim}{v}-\beta_0\underset{\sim}{j}\right]'(\delta^2 I)^{-1}\left[\underset{\sim}{v}-\beta_0\underset{\sim}{j}\right]+\left[\underset{\sim}{\beta}_{(0)}-\underset{\sim}{\mu}_\beta\right]'G^{-1}\left[\underset{\sim}{\beta}_{(0)}-\underset{\sim}{\mu}_\beta\right]\right\}} \\
& \times \frac{|D|^{1/2}}{|\delta^2 I|^{1/2}|D+\frac{1}{\sigma^2}I|^{1/2}|G|^{1/2}}\frac{1}{(1+\sigma^2)^2}\frac{1}{(1+\delta^2)^2}
\end{aligned}$$

Next, we will show that the approximate conditional posterior density of v_i is also normal distribution and all v_is are independent as well. Here we consider each v_i. Let $\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} = n$, $(\sigma_{ij}^2)_{n\times n} = D^{-1}+\sigma^2 I$, $(\underset{\sim}{t}_{ij})_{n\times 1} = D^{-1}C$ and $(\mu_{w_{ij}})_{n\times 1} = \underset{\sim}{\mu}_w$.

Look at the exponent only containing $v_i, i=1,\ldots,\ell$. in the $\pi_a(\underset{\sim}{v},\underset{\sim}{\beta},\sigma^2,\delta^2 \mid \underset{\sim}{y})$

$$\begin{aligned}
&\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\frac{1}{\sigma_{ij}^2}\left[v_i-\mu_{w_{ij}}+\underset{\sim}{t}_{ij}'(\underset{\sim}{\beta}_{(0)}-\underset{\sim}{\mu}_\beta)\right]^2+\frac{1}{\delta^2}\sum_{i=1}^{\ell}(v_i-\beta_0)^2 \\
&=\sum_{i=1}^{\ell}\Big(\frac{1}{\sum_{j=1}^{n_i}\sigma_{ij}^2}+\frac{1}{\delta^2}\Big)^{-1}\left\{v_i-\frac{\left(\frac{1}{\sum_{j=1}^{n_i}\sigma_{ij}^2}\right)\left[\bar{\mu}_{w_i}-\bar{\underset{\sim}{t}}_i'(\underset{\sim}{\beta}_{(0)}-\underset{\sim}{\mu}_\beta)\right]+\frac{1}{\delta^2}\beta_0}{\frac{1}{\sum_{j=1}^{n_i}\sigma_{ij}^2}+\frac{1}{\delta^2}}\right\}^2 \\
&\quad+\sum_{i=1}^{\ell}\left(\frac{1}{1/\sum_{j=1}^{n_i}\sigma_{ij}^2}+\delta^2\right)^{-1}\left\{\bar{\mu}_{w_i}-\bar{\underset{\sim}{t}}_i'(\underset{\sim}{\beta}_{(0)}-\underset{\sim}{\mu}_\beta)-\beta_0\right\}^2 \\
&\quad+\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\frac{1}{\sigma_{ij}^2}\left\{(\bar{\underset{\sim}{t}}_i-\underset{\sim}{t}_{ij})'(\underset{\sim}{\beta}_{(0)}-\underset{\sim}{\mu}_\beta)-(\bar{\mu}_{w_i}-\mu_{w_{ij}})\right\}^2
\end{aligned}$$

,where $\bar{\mu}_{w_i} = \frac{1}{n_i}\sum_{j=1}^{n_i}\mu_{w_{ij}}$ and $\bar{\underset{\sim}{t}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\underset{\sim}{t}_{ij}$

Then it is easy to see that

$$v_i|\beta_0,\sigma^2,\delta^2,\underset{\sim}{y} \overset{\text{app}}{\sim} \text{Normal}\left\{\frac{\left(\frac{1}{\sum_{j=1}^{n_i}\sigma_{ij}^2}\right)\left[\bar{\mu}_{w_i}-\bar{\underset{\sim}{t}}_i'(\underset{\sim}{\beta}_{(0)}-\underset{\sim}{\mu}_\beta)\right]+\frac{1}{\delta^2}\beta_0}{\frac{1}{\sum_{j=1}^{n_i}\sigma_{ij}^2}+\frac{1}{\delta^2}},\frac{1}{\sum_{j=1}^{n_i}\sigma_{ij}^2}+\frac{1}{\delta^2}\right\}$$

Similarly, we can use parallel computing to draw $\nu_i,\ i = 1,\ldots,\ell$ as well since all of them are independent given $\underset{\sim}{\beta}_{(0)},\beta_0,\sigma^2,\delta^2$. Then we can integrate out $\underset{\sim}{\nu}$ from the joint approximate posterior density $\pi_a(\underset{\sim}{\nu},\underset{\sim}{\beta},\sigma^2,\delta^2 \mid \underset{\sim}{y})$ and obtain the joint posterior density of $\underset{\sim}{\beta},\sigma^2$ and δ^2

$$\begin{aligned}
\pi_a(\underset{\sim}{\beta},\sigma^2,\delta^2|\underset{\sim}{y}) \propto & \exp\left\{-\frac{1}{2}\sum_{i=1}^{\ell}\left(\frac{1}{1/\sum_{j=1}^{n_i}\sigma_{ij}^2}+\delta^2\right)^{-1}\left[\bar{\mu}_{w_i}-\underset{\sim}{\bar{t}}_i'(\underset{\sim}{\beta}_{(0)}-\underset{\sim}{\mu}_\beta)-\beta_0\right]^2\right\} \\
& \times \exp\left\{-\frac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\frac{1}{\sigma_{ij}^2}\left[\{(\underset{\sim}{\bar{t}}_i-\underset{\sim}{t}_{ij})'(\underset{\sim}{\beta}_{(0)}-\underset{\sim}{\mu}_\beta)-(\bar{\mu}_{w_i}-\mu_{w_{ij}})\right]^2\right\} \\
& \times \prod_{i=1}^{l}\left(\frac{1}{\sum_{j=1}^{n_i}\sigma_{ij}^2}+\frac{1}{\delta^2}\right)^{\frac{1}{2}}\frac{1}{|\delta^2 I|^{1/2}|D+\frac{1}{\sigma^2}I|^{1/2}}\frac{1}{(1+\sigma^2)^2}\frac{1}{(1+\delta^2)^2} \\
= & e^{-\frac{1}{2}\left(\underset{\sim}{\mu}_w-D^{-1}C'(\underset{\sim}{\beta}_{(0)}-\underset{\sim}{\mu}_\beta)-\beta_0\underset{\sim}{j}\right)'\left(D^{-1}+\sigma^2 I+\delta^2 I\right)^{-1}\left(\underset{\sim}{\mu}_w-D^{-1}C'(\underset{\sim}{\beta}_{(0)}-\underset{\sim}{\mu}_\beta)-\beta_0\underset{\sim}{j}\right)} \\
& \times e^{-\frac{1}{2}\left(\underset{\sim}{\beta}_{(0)}-\underset{\sim}{\mu}_\beta\right)'G^{-1}\left(\underset{\sim}{\beta}_{(0)}-\underset{\sim}{\mu}_\beta\right)} \\
& \times \prod_{i=1}^{l}\left(\frac{1}{\sum_{j=1}^{n_i}\sigma_{ij}^2}+\frac{1}{\delta^2}\right)^{\frac{1}{2}}\frac{1}{|\delta^2 D+\frac{\delta^2}{\sigma^2}I|^{1/2}}\frac{1}{(1+\sigma^2)^2}\frac{1}{(1+\delta^2)^2}.
\end{aligned}$$

Next we assume that the conditional posterior density of $\underset{\sim}{\beta}|\sigma^2,\delta^2,\underset{\sim}{y}$ has an approximate multivariate normal density,

$$\begin{pmatrix}\beta_0\\ \underset{\sim}{\beta}_{(0)}\end{pmatrix}|\sigma^2,\delta^2,\underset{\sim}{y} \sim \text{Normal}\left\{\begin{pmatrix}\omega_0\\ \underset{\sim}{\omega}_{(0)}\end{pmatrix},\begin{pmatrix}\delta_0^2 & \underset{\sim}{\gamma}'\\ \underset{\sim}{\gamma} & \Delta_{(0)}\end{pmatrix}^{-1}\right\},$$

which is denoted by $\pi_a(\underset{\sim}{\beta} \mid \sigma^2,\delta^2,\underset{\sim}{y})$. The density function is

$$\pi_a(\underset{\sim}{\beta} \mid \sigma^2,\delta^2,\underset{\sim}{y}) \propto \left|\begin{pmatrix}\delta_0^2 & \underset{\sim}{\gamma}'\\ \underset{\sim}{\gamma} & \Delta_{(0)}\end{pmatrix}\right|^{\frac{1}{2}} \times e^{-\frac{1}{2}\begin{pmatrix}\beta_0-\omega_0\\ \underset{\sim}{\beta}_{(0)}-\underset{\sim}{\omega}_{(0)}\end{pmatrix}'\begin{pmatrix}\delta_0^2 & \underset{\sim}{\gamma}'\\ \underset{\sim}{\gamma} & \Delta_{(0)}\end{pmatrix}\begin{pmatrix}\beta_0-\omega_0\\ \underset{\sim}{\beta}_{(0)}-\underset{\sim}{\omega}_{(0)}\end{pmatrix}}$$

So the exponent terms are

$$\begin{pmatrix}\beta_0-\omega_0\\ \underset{\sim}{\beta}_{(0)}-\underset{\sim}{\omega}_{(0)}\end{pmatrix}'\begin{pmatrix}\delta_0^2 & \underset{\sim}{\gamma}'\\ \underset{\sim}{\gamma} & \Delta_{(0)}\end{pmatrix}\begin{pmatrix}\beta_0-\omega_0\\ \underset{\sim}{\beta}_{(0)}-\underset{\sim}{\omega}_{(0)}\end{pmatrix}.$$

Consider the exponent terms containing $\underset{\sim}{\beta}_{(0)}$ and β_0

$$
\begin{aligned}
&\left(\underset{\sim}{\mu}_w - D^{-1}C'(\underset{\sim}{\beta}_{(0)} - \underset{\sim}{\mu}_\beta) - \beta_0 \underset{\sim}{j}\right)' \left(D^{-1} + \sigma^2 I + \delta^2 I\right)^{-1} \left(\underset{\sim}{\mu}_w - D^{-1}C'(\underset{\sim}{\beta}_{(0)} - \underset{\sim}{\mu}_\beta) - \beta_0 \underset{\sim}{j}\right) \\
&\quad + \left(\underset{\sim}{\beta}_{(0)} - \underset{\sim}{\mu}_\beta\right)' G^{-1} \left(\underset{\sim}{\beta}_{(0)} - \underset{\sim}{\mu}_\beta\right) \\
&= \underset{\sim}{\beta}'_{(0)} \left[CD^{-1}(D^{-1} + \sigma^2 I + \delta^2 I)^{-1} D^{-1} C' + G^{-1}\right] \underset{\sim}{\beta}_{(0)} + \underset{\sim}{j}'(D^{-1} + \sigma^2 I + \delta^2 I)^{-1} \underset{\sim}{j} \beta_0^2 \\
&\quad - 2\left[(\underset{\sim}{\mu}_w + D^{-1}C'\underset{\sim}{\mu}_\beta)'(D^{-1} + \sigma^2 I + \delta^2 I) D^{-1} C' + \underset{\sim}{\mu}'_\beta G^{-1}\right] \underset{\sim}{\beta}_{(0)} \\
&\quad - 2\left[(\underset{\sim}{\mu}_w + D^{-1}C'\underset{\sim}{\mu}_\beta)'(D^{-1} + \sigma^2 I + \delta^2 I)^{-1} \underset{\sim}{j}\right] \beta_0 + 2CD^{-1}(D^{-1} + \sigma^2 I + \delta^2 I)^{-1} \underset{\sim}{j} \beta_0 \underset{\sim}{\beta}_{(0)} \\
&\quad + (D^{-1}C'\underset{\sim}{\mu}_\beta + \underset{\sim}{\mu}_w)'(D^{-1} + \sigma^2 I + \delta^2 I)^{-1} \underset{\sim}{j} \beta_0 \underset{\sim}{\beta}_{(0)} \\
&\quad (\underset{\sim}{\mu}_w + D^{-1}C'\underset{\sim}{\mu}_\beta)'(D^{-1} + \sigma^2 I + \delta^2 I)^{-1}(\underset{\sim}{\mu}_w + D^{-1}C'\underset{\sim}{\mu}_\beta) + \underset{\sim}{\mu}'_\beta G^{-1} \underset{\sim}{\mu}_\beta
\end{aligned}
$$

We know those two exponent parts are equal, so we have

$$\Delta_{(0)} = CD^{-1}(D^{-1} + \sigma^2 I + \delta^2 I)^{-1} D^{-1} C' + G^{-1},$$

$$\delta_0^2 = \underset{\sim}{j}'(D^{-1} + \sigma^2 I + \delta^2 I)^{-1} \underset{\sim}{j},$$

$$\underset{\sim}{\gamma} = CD^{-1}(D^{-1} + \sigma^2 I + \delta^2 I)^{-1} \underset{\sim}{j},$$

$$\begin{pmatrix} \omega_0 \\ \underset{\sim}{\omega}_{(0)} \end{pmatrix} = \begin{pmatrix} \delta_0^2 & \underset{\sim}{\gamma}' \\ \underset{\sim}{\gamma} & \Delta_{(0)} \end{pmatrix}^{-1} \begin{pmatrix} (\underset{\sim}{\mu}_w + D^{-1}C'\underset{\sim}{\mu}_\beta)'(D^{-1} + \sigma^2 I + \delta^2 I)^{-1} \underset{\sim}{j} \\ (\underset{\sim}{\mu}_w + D^{-1}C'\underset{\sim}{\mu}_\beta)'(D^{-1} + \sigma^2 I + \delta^2 I) D^{-1} C' + \underset{\sim}{\mu}'_\beta G^{-1} \end{pmatrix}.$$

That is, $\underset{\sim}{\beta}|\sigma^2, \delta^2, \underset{\sim}{y}$ approximately follows multivariate normal distribution,

$$\begin{pmatrix} \beta_0 \\ \underset{\sim}{\beta}_{(0)} \end{pmatrix} |\sigma^2, \delta^2, \underset{\sim}{y} \sim \text{Normal}\left\{ \begin{pmatrix} \omega_0 \\ \underset{\sim}{\omega}_{(0)} \end{pmatrix}, \begin{pmatrix} \delta_0^2 & \underset{\sim}{\gamma}' \\ \underset{\sim}{\gamma} & \Delta_{(0)} \end{pmatrix}^{-1} \right\},$$

Then we can easily integrate out $\underset{\sim}{\beta}$ from the joint density of $\underset{\sim}{\beta}, \sigma^2, \delta^2|\underset{\sim}{y}$, and get the posterior density of $\sigma^2, \delta^2|\underset{\sim}{y}$

$$\pi_a(\sigma^2,\delta^2|\underset{\sim}{y}) \propto \begin{vmatrix} \delta_0^2 & \underset{\sim}{\gamma}' \\ \underset{\sim}{\gamma} & \Delta_{(0)} \end{vmatrix}^{-\frac{1}{2}} \times \prod_{i=1}^{l} \left(\frac{1}{\sum_{j=1}^{n_i} \sigma_{ij}^2} + \frac{1}{\delta^2} \right)^{\frac{1}{2}} \frac{1}{|\delta^2 D + \frac{\delta^2}{\sigma^2} I|^{1/2}} \frac{1}{(1+\sigma^2)^2} \frac{1}{(1+\delta^2)^2}$$
$$\times \exp\left\{ -\frac{1}{2} (\underset{\sim}{\mu}_w + D^{-1} C' \underset{\sim}{\mu}_\beta)' (D^{-1} + \sigma^2 I + \delta^2 I)^{-1} (\underset{\sim}{\mu}_w + D^{-1} C' \underset{\sim}{\mu}_\beta) + \underset{\sim}{\mu}_\beta' G^{-1} \underset{\sim}{\mu}_\beta \right\}$$
$$\times \exp\left\{ -\frac{1}{2} \begin{pmatrix} \beta_0 - \omega_0 \\ \underset{\sim}{\beta}_{(0)} - \underset{\sim}{\omega}_{(0)} \end{pmatrix}' \begin{pmatrix} \delta_0^2 & \underset{\sim}{\gamma}' \\ \underset{\sim}{\gamma} & \Delta_{(0)} \end{pmatrix} \begin{pmatrix} \beta_0 - \omega_0 \\ \underset{\sim}{\beta}_{(0)} - \underset{\sim}{\omega}_{(0)} \end{pmatrix} \right\}.$$

□

The INNA is actually a random sampler. First, we draw samples for σ^2, δ^2 from $\pi(\sigma^2, \delta^2|\underset{\sim}{y})$. The posterior distribution of $\sigma^2, \delta^2|\underset{\sim}{y}$ does not have standardized form. Here we use grid method and numerical integration to sample σ^2 and δ^2. Since $0 < \sigma^2 < \infty$ and $0 < \delta^2 < \infty$, we make a transformation to $\phi_1 = \frac{1}{1+\sigma^2}$ and $\phi_2 = \frac{1}{1+\delta^2}$ so that we get $0 < \phi_1 < 1$ and $0 < \phi_2 < 1$. Then the posterior density of $\phi_1, \phi_2|\underset{\sim}{y}$ is

$$\pi_a(\phi_1,\phi_2|\underset{\sim}{y}) \propto \left\{ \begin{vmatrix} \delta_0^2 & \underset{\sim}{\gamma}' \\ \underset{\sim}{\gamma} & \Delta_{(0)} \end{vmatrix}^{-\frac{1}{2}} \times \prod_{i=1}^{l} \left(\frac{1}{\sum_{j=1}^{n_i} \sigma_{ij}^2} + \frac{1}{\delta^2} \right)^{\frac{1}{2}} \frac{1}{|\delta^2 D + \frac{\delta^2}{\sigma^2} I|^{1/2}} \right\}_{\phi_1=\frac{1}{1+\sigma^2},\ \phi_2=\frac{1}{1+\delta^2}}$$
$$\times \exp\left\{ -\frac{1}{2} (\underset{\sim}{\mu}_w + D^{-1} C' \underset{\sim}{\mu}_\beta)' (D^{-1} + \sigma^2 I + \delta^2 I)^{-1} (\underset{\sim}{\mu}_w + D^{-1} C' \underset{\sim}{\mu}_\beta) + \underset{\sim}{\mu}_\beta' G^{-1} \underset{\sim}{\mu}_\beta \right\}$$
$$\times \exp\left\{ -\frac{1}{2} \begin{pmatrix} \beta_0 - \omega_0 \\ \underset{\sim}{\beta}_{(0)} - \underset{\sim}{\omega}_{(0)} \end{pmatrix}' \begin{pmatrix} \delta_0^2 & \underset{\sim}{\gamma}' \\ \underset{\sim}{\gamma} & \Delta_{(0)} \end{pmatrix} \begin{pmatrix} \beta_0 - \omega_0 \\ \underset{\sim}{\beta}_{(0)} - \underset{\sim}{\omega}_{(0)} \end{pmatrix} \right\}_{\phi_1=\frac{1}{1+\sigma^2},\ \phi_2=\frac{1}{1+\delta^2}}.$$

We need draw ϕ_1, ϕ_2 together. The joint density can be rewritten as

$$\pi(\phi_1,\phi_2|\underset{\sim}{y}) = \pi(\phi_2|\phi_1)\pi(\phi_1|\underset{\sim}{y}) = \pi(\phi_2|\phi_1) \int_0^1 \pi(\phi_1,\phi_2|\underset{\sim}{y}) d\phi_2$$

We plug each grid of $\phi_1 \in (0,1)$ into $\int_0^1 \pi(\phi_1,\phi_2|\underset{\sim}{y})d\phi_2$ and then use numerical integration to get the density of $(\phi_1|\underset{\sim}{y})$. After we plug all the 100 grids, we can get 100 value of $\pi(\phi_1|\underset{\sim}{y})$ and then draw ϕ_1 from them, i.e. $\phi_1^{(h)}$. Next, we plug $\phi_1^{(h)}$ into $\pi(\phi_2|\phi_1)$ and use grid method to draw $\phi_2^{(h)}$. Repeat those steps 10000 times to get the sample of $(\phi_1^{(h)}, \phi_2^{(h)}), h = 1, \ldots, 10000$. Once we get samples for ϕ_1, ϕ_2, we transform them back to σ^2 and δ^2 respectively. Second, given σ^2, δ^2, simply draw samples of $\underset{\sim}{\beta}$ from the approximate multivariate normal distribution $\pi_a(\underset{\sim}{\beta} \mid \sigma^2, \delta^2, \underset{\sim}{y})$. Third, we can draw samples of ν_i independently given $\underset{\sim}{\beta}, \delta^2$ and data from the approximate normal distribution $\pi_a(\underset{\sim}{v}|\beta_0, \delta^2, \underset{\sim}{y})$. Finally samples of w_{ij} independently given $\underset{\sim}{v}, \underset{\sim}{\beta}, \sigma^2$ can be obtained from the approximate normal distribution $\pi_a(\underset{\sim}{w}|\underset{\sim}{v}, \underset{\sim}{\beta}_{(0)}, \sigma^2, \underset{\sim}{y})$. Notice that the last three step are very simple, just drawing samples from normal densities respectively. In addition, w_{ij} and v_i are all independent so that we can draw them simultaneously. Therefore, those latter steps permit fast computing.

4. Numerical Example

4.1 Nepal Living Standards Survey II

The performance of our method is studied using the Nepal Living Standard Survey (NLSS II), conducted in the years 2003-2004. The main objective of NLSS II is to track changes and progress about national living standards and social indicators of Nepalese population. It is an integrated survey which covers samples from the whole country and runs throughout the year.

The NLSS II gathers information on a variety of aspects. It has collected data on demographics, housing, education, health, fertility, employment, income, agricultural activity, consumption, and various other areas. The sampling design of NLSS II is two-stage stratified sampling. Nepal is stratified into PSUs and within each PSU, there are a number of households (sub-area) are selected. All household members in the sample were interviewed.

In detail, NLSS II has records for 20,263 individuals from 3,912 households (sub-areas) from 326 PSUs (areas) from a population of 60,262 households and about two million Nepalese. A sample of PSUs was selected from strata using the probability proportional to size (PPS) sampling and 12 households were systematically selected from each PSU. The survey is self-weighed and some adjustments were maded after conducting the survey for non-response or missing data. For simplicity, in this paper, we assume all samples have the same weight. Table 1 shows the distribution of all samples by stratum.

Table 1: Distribution of wards and households in the sample

Strata	Mountains	Kathemandu	Urban Hill	Rural Hills	Urban Tarai	Rural Tarai	Total
PSU	32	34	28	96	34	102	326
Households	384	408	336	1,152	408	1,224	3,912
Individuals	1,949	1,954	1,467	5,755	2,104	7,034	20,263

We choose four relevant covariates which can influence health status from the same NLSS II survey for out two-fold logistic regression model. They are age, nativity, sex and religion. We created binary variables nativity (Indigenous = 1, Non-indigenous = 0) and religion ((Hindu = 1, Non-Hindu = 0), sex (Male = 1, Female = 0). Table 2 shows the details of these 4 covariates. In the model fitting, we standardize age covariate. Older age and child age are more vulnerable than younger age. Indigenous people can have different health status from migrated people.

Table 2: The descriptives of 4 covariates

Covariates		Frequency	Percentage
Age	0-14	7,765	38.32
	15-59	10,951	54.04
	60+	1,547	7.64
Gender	Male	9,763	48.18
	Female	10,500	51.82
Nativity	Indigeous	11,903	41.25
	Non-Indigous	8,360	58.75
Religion	Hingdu	16,378	80.83
	Non-Hingdu	3,385	19.17

According to the 2001 census data, only about 0.091% of households and only 0.904% of PSU were sampled. NLSS II was designed to provide reliable estimates only at stratum level or even larger areas than stratum. It cannot give estimates in small area (PSU or household level) since the sample sizes are too small. Therefore, we need to use statistical models to fit the available data and find reliable estimates in small areas. In our study, we choose the binary variable, health status, from the health section of the questionnaire.

4.2 Numerical Comparison

We use data from NLSS II to illustrate our sub-area logistic regression model. We predict the household proportions of members in good health for 18,924 households (sampled and non-sampled). This analysis is based on 1,224 sample households from 102 wards (PSUs) in strata 6. Our primary purposes are to show that our model can provide good estimates and to compare the approximate method with the exact method when there are random effects at the household level.

In Figure 1, 2, 3, we compare respectively the posterior means (PMs), posterior standard deviations (PSDs) and posterior coefficient of variations (CVs) in the household level as our primary purpose.

Figure 1: Comparison of the posterior means (PM) of the household proportions by the approximate method and the exact method

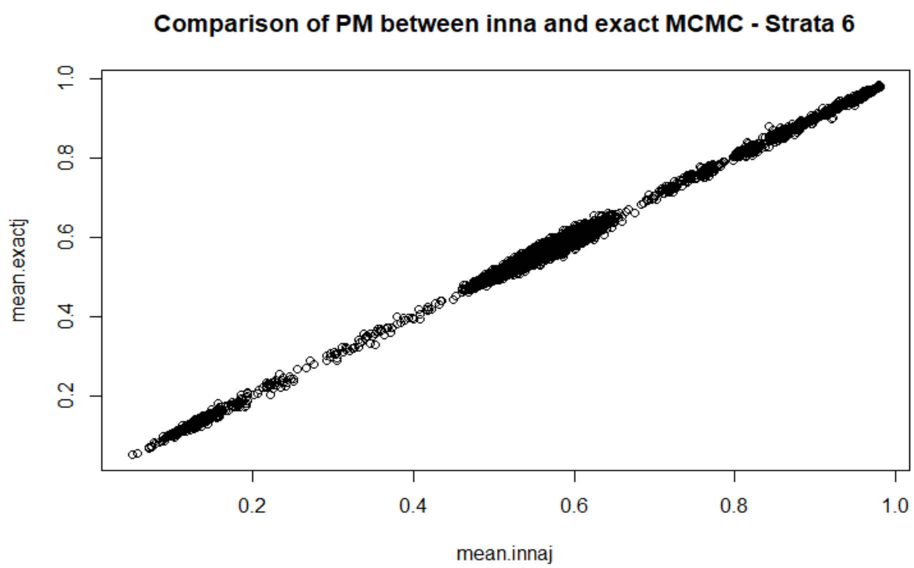


Figure 2: Comparison of the posterior standard deviations (PSD) of the household proportions by the approximate method and the exact method

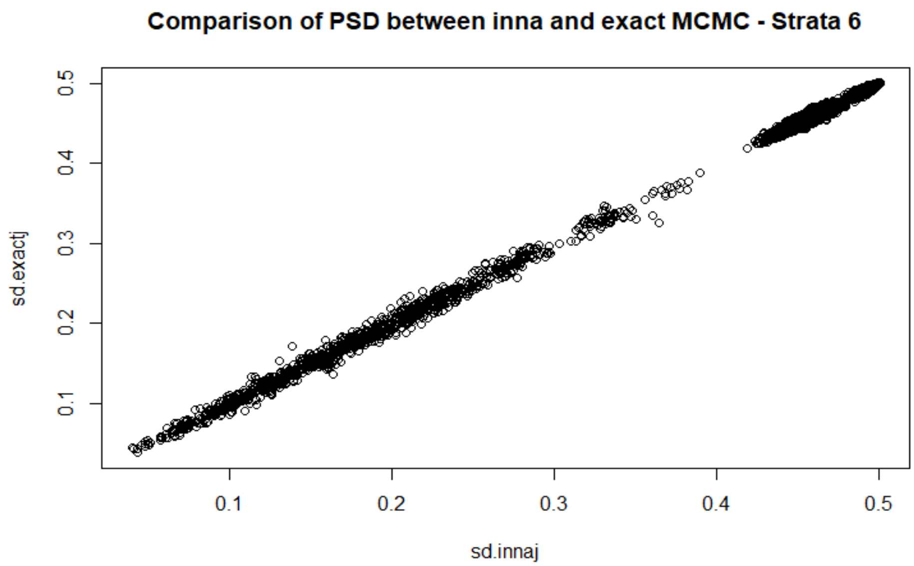
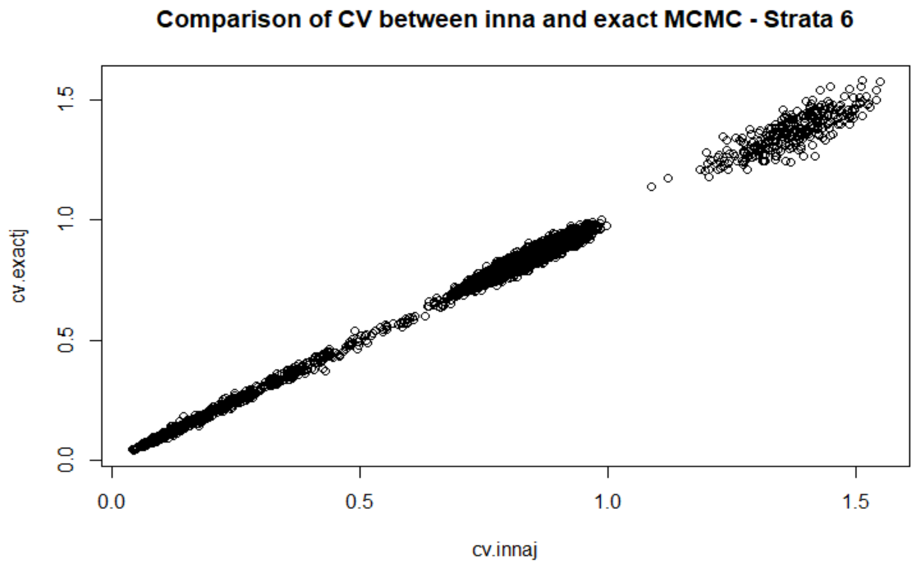


Figure 3: Comparison of the posterior coefficient of variations (CV) of the household proportions by the approximate method and the exact method



We can see that the PMs are very close, nearly lying on the 45 degree line through the origin. PSDs a little bit spread out and thicker but all points are still lies on the 45 degree line and so as the CVs. Overall, these approximations are acceptable in data analysis. In Figure 4, 5, 6, we compare respectively the posterior means (PMs), posterior standard deviations (PSDs) and posterior coefficient of variations (CVs) in the household level as our primary purpose. The plots of PM are still very good. Notice that other two plots of PSDs and CVs are more spread out that those in household level. But again the approximate method and the exact method are reasonably closed.

Figure 4: Comparison of the posterior means (PM) of the ward proportions by the approximate method and the exact method

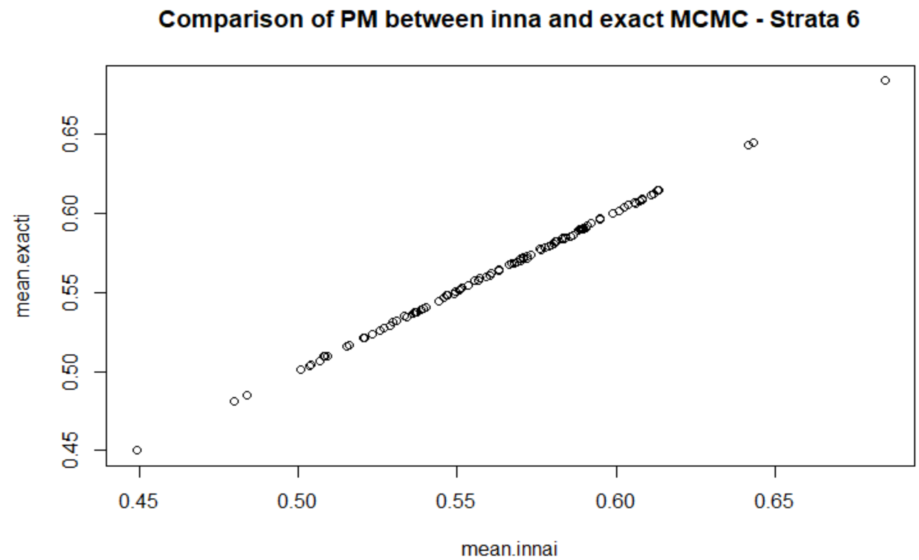


Figure 5: Comparison of the posterior standard deviations (PSD) of the ward proportions by the approximate method and the exact method

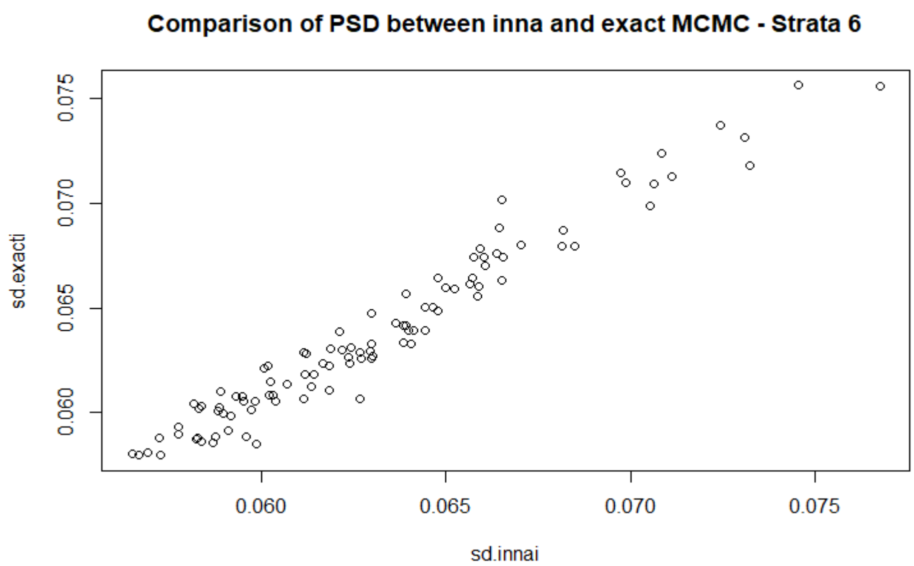
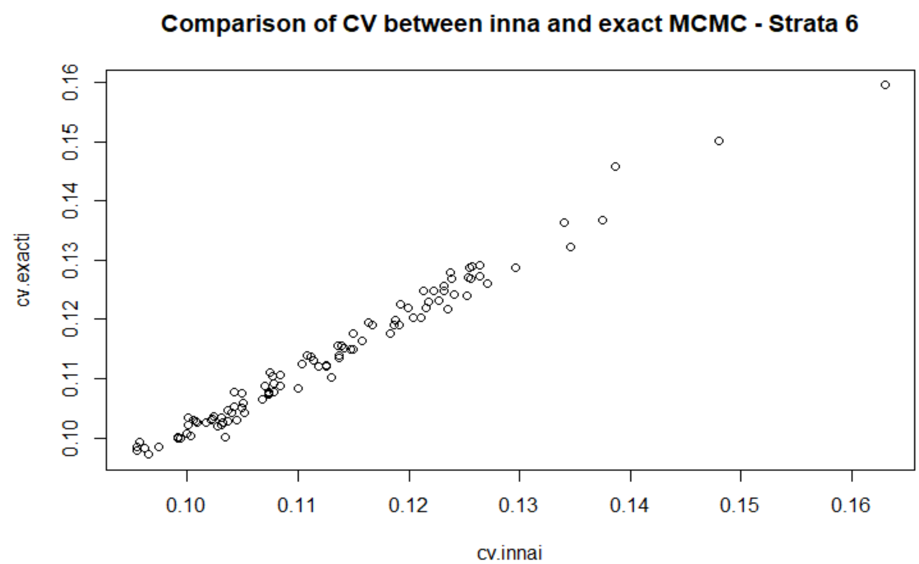


Figure 6: Comparison of the posterior coefficient of variations (CV) of the ward proportions by the approximate method and the exact method



5. Conclusion and Future Work

Sub-area HB logistic regression model can be applied to analyze binary response variable. This model is an extension of the HB logistic regression area-level model, which ignores the actual hierarchical structure of data. We propose an approximation method, INNA, to fit the model. For large dataset, it is very unrealistic to use MCMC method to fit the model. We propose the approximation method, INNA, which save time significantly becasue there is no need to compute numerous modes. In the numerical example, we can show that INNA can provide reliable estimates as well. An illustrative example of NLSS II is presented in order to compare the approximation method and the exact method. It shows that when there are large number of areas and subareas, the approximation method can be efficient and it can also provide reasonable estimates.

There are many future works on two-fold small areas model. First, in this paper, we assume equal survey weights since NLSS II is a self-weighted sampling. However, after data were collected, the sampling weights are usually adjusted for various characteristics or based on nonresponse as well. Incorporating those survey weights into the model is also very important. Generally, we need to consider these weights into the model. NLSS II is a national wide and population based survey. We should rescale sample weights to sum to an equivalent sample size. That is, we consider adjusted weight as $w^*_{ijk} = \hat{n}(\frac{w_{ijk}}{\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\sum_{k=1}^{m_{ij}} w_{ijk}})$, where $\hat{n} = \frac{(\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\sum_{k=1}^{m_{ij}} w_{ijk})^2}{\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\sum_{k=1}^{m_{ij}} w_{ijk}^2}$ as an equivalent sample. Introducing the sampling weights, we can obtain a updated normalized likelihood function. Based the updated likelihood function and same prior in the two-fold model, we can have full Bayesian analysis on the updated model and then project the finite population proportion of family members with good health in each household.

Second, we focus on the binary data. Actually there are 4 options in the health status questionnaire. Multinomial-Dirichlet model can be an extension for polychotomous data. Third, the two fold sub-area level models can also be extended to three-fold models if the data have additional hierarchical structure, actually the NLSS II has this structure (households within wards, wards within districts). Forth, in our models, we are considering parametric priors. Introducing Dirichlet process as prior might be able to make our method more robust to its specifications.

A. Exact Method for Sub-Area Logistic Regression Model

Recall that the joint posterior distribution of our two-fold logistic regression model is The joint posterior density for the parameters is

$$\pi(\underset{\sim}{\nu}, \underset{\sim}{w}, \underset{\sim}{\beta}, \sigma^2, \delta^2 | \underset{\sim}{y}) \propto \prod_{i=1}^{\ell}\prod_{j=1}^{n_i}\prod_{k=1}^{m_{ij}} \left[\frac{e^{(\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}_{(0)}+w_{ij})y_{ijk}}}{1+e^{\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}_{(0)}+w_{ij}}}\right] \times \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{-\sum_{i=1}^{l}\sum_{j=1}^{ni}\frac{(w_{ij}-\nu_i)^2}{2\sigma^2}\right\}$$
$$\times\left(\frac{1}{\sqrt{2\pi\delta^2}}\right)^l \exp\left\{-\sum_{i=1}^{l}\frac{(\nu_i-\beta_0)^2}{2\delta^2}\right\}\frac{1}{(1+\sigma^2)^2}\frac{1}{(1+\delta^2)^2}.$$

We can see that the form of the joint posterior density is very complicated. It is very time consuming to draw all the posterior samples if applying the exact MCMC method. But the exact method will provide reliable estimates of all parameters, so in order to test the performance of our approximation method, we need to apply MCMC method on our model and then compare the performance of two methods. We use Metropolis-Hastings sampler to draw samples for $\underset{\sim}{\beta}, \sigma^2, \delta^2$ together and then draw $\underset{\sim}{\nu}$ given $\underset{\sim}{\beta}, \sigma^2, \delta^2$ samples. At last, we use MH method to draw $\underset{\sim}{w}$ given $\underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2, \delta^2$ samples.

In order to draw samples for $\underset{\sim}{\beta}, \sigma^2, \delta^2$ together, we need to integrate out $\underset{\sim}{w}$ and $\underset{\sim}{\nu}$. First, we integrate out $\underset{\sim}{\nu}$ from the joint posterior density $\pi(\underset{\sim}{\nu}, \underset{\sim}{w}, \underset{\sim}{\beta}, \sigma^2, \delta^2 | \underset{\sim}{y})$ to get

$$\pi(\underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2, \delta^2 | \underset{\sim}{y}) \propto \int_{\Omega}\prod_{i=1}^{\ell}\left\{\prod_{j=1}^{n_i}\left[\prod_{k=1}^{m_{ij}}\frac{e^{(\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}_{(0)}+\mu_i+w_{ij})y_{ij}}}{1+e^{\underset{\sim}{x}'_{ijk}\underset{\sim}{\beta}_{(0)}+\mu_i+w_{ij}}}\right]\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(w_{ij}-\nu_i)^2}{2\sigma^2}}\right\}d\underset{\sim}{w}$$
$$\times\left(\frac{1}{\sqrt{2\pi\delta^2}}\right)^l \exp\left\{-\sum_{i=1}^{l}\frac{(\nu_i-\beta_0)^2}{2\delta^2}\right\}\frac{1}{(1+\sigma^2)^2}\frac{1}{(1+\delta^2)^2}.$$

Notice that the integrant is not any simple distribution function, so we use Monte Carlo numberical integration to approximate the integrals. Let $z_{ij}^w = \frac{w_{ij}-\nu_i}{\sigma}$. Notice that z_{ij}^w follows standard normal distribution. For standard normal density, 99.7% of data will fall within 3 standard deviations of the mean, which corresponds to the interval $[-3,3]$. Therefore, we bounded the integration domain to $[-3,3]$ and divide the interval to M equal subintervals $[p_{a-1}, p_a], a = 1, \ldots, M$. Then we can get an approximate but very accurate joint density

$$\pi(\underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2, \delta^2 | \underset{\sim}{y}) \propto \prod_{i=1}^{\ell} \prod_{j=1}^{n_i} \left\{ \sum_{a=1}^{M} \int_{p_{a-1}}^{p_a} \frac{e^{\sum_{k=1}^{m_{ij}} (\underset{\sim}{x}'_{ijk} \underset{\sim}{\beta}_{(0)} + w_{ij}) y_{ijk}}}{\prod_{k=1}^{m_{ij}} \left[1 + e^{\underset{\sim}{x}'_{ijk} \underset{\sim}{\beta}_{(0)} + w_{ij}}\right]} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(w_{ij}-\nu_i)^2}{2\sigma^2}} dw_{ij} \right\}$$
$$\times \left(\frac{1}{\sqrt{2\pi\delta^2}}\right)^l \exp\left\{-\sum_{i=1}^{l} \frac{(\nu_i - \beta_0)^2}{2\delta^2}\right\} \frac{1}{(1+\sigma^2)^2} \frac{1}{(1+\delta^2)^2}$$
$$\propto \prod_{i=1}^{\ell} \prod_{j=1}^{n_i} \left\{ \sum_{a=1}^{M} \int_{p_{a-1}}^{p_a} \frac{e^{\sum_{k=1}^{m_{ij}} (\underset{\sim}{x}'_{ijk} \underset{\sim}{\beta}_{(0)} + \sigma z_{ij}^w + \nu_i) y_{ijk}}}{\prod_{k=1}^{m_{ij}} \left[1 + e^{\underset{\sim}{x}'_{ijk} \underset{\sim}{\beta}_{(0)} + \sigma z_{ij}^w + \nu_i}\right]} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z_{ij}^w)^2}{2}} dz_{ij}^w \right\}$$
$$\times \left(\frac{1}{\sqrt{2\pi\delta^2}}\right)^l \exp\left\{-\sum_{i=1}^{l} \frac{(\nu_i - \beta_0)^2}{2\delta^2}\right\} \frac{1}{(1+\sigma^2)^2} \frac{1}{(1+\delta^2)^2}$$

Let $\bar{z}_a^w = \frac{p_a - p_{a-1}}{2}$, which is the midpoint of each interval $[p_{a-1}, p_a], a = 1, \ldots, M$. We use midpoint rule to approximate the definite integrals. We divide the interval $[-3,3]$ into 100 subintervals, and so we use 100 midpoints to get the approximate joint posterior distribution

$$\pi(\underset{\sim}{\nu}, \underset{\sim}{\beta}, \sigma^2, \delta^2 | \underset{\sim}{y}) \approx \prod_{i=1}^{\ell} \prod_{j=1}^{n_i} \left\{ \sum_{a=1}^{100} \frac{e^{\sum_{k=1}^{m_{ij}} (\underset{\sim}{x}'_{ijk} \underset{\sim}{\beta}_{(0)} + \sigma \bar{z}_a^w + \nu_i) y_{ijk}}}{\prod_{k=1}^{m_{ij}} \left[1 + e^{\underset{\sim}{x}'_{ijk} \underset{\sim}{\beta}_{(0)} + \sigma \bar{z}_a^w + \nu_i}\right]} \Big(\Phi(a) - \Phi(a-1)\Big) \right\}$$
$$\times \left(\frac{1}{\sqrt{2\pi\delta^2}}\right)^l \exp\left\{-\sum_{i=1}^{l} \frac{(\nu_i - \beta_0)^2}{2\delta^2}\right\} \frac{1}{(1+\sigma^2)^2} \frac{1}{(1+\delta^2)^2}$$

Similarly, let $z_i^\nu = \frac{\nu_i - \beta_0}{\delta}$ and $\bar{z}_b^\nu = \frac{p_b - p_{b-1}}{2}$, $b = 1, \ldots, 100$. We use the midpoint rule to approximate the definite integral with respect to $\underset{\sim}{\nu}$ and then get the posterior density of $\underset{\sim}{\beta}, \sigma^2, \delta^2 | \underset{\sim}{y}$

$$\pi(\underset{\sim}{\beta}, \sigma^2, \delta^2 | \underset{\sim}{y}) \approx \prod_{i=1}^{\ell} \left\{ \sum_{b=1}^{100} \left[\prod_{j=1}^{n_i} \left(\sum_{a=1}^{100} \frac{e^{\sum_{k=1}^{m_{ij}} (\underset{\sim}{x}'_{ijk} \underset{\sim}{\beta}_{(0)} + \beta_0 + \bar{z}_a^w \delta + \bar{z}_b^v \sigma) y_{ijk}}}{\prod_{k=1}^{m_{ij}} \left[1 + e^{\underset{\sim}{x}'_{ijk} \underset{\sim}{\beta}_{(0)} + \beta_0 + \bar{z}_a^w \delta + \bar{z}_b^v \sigma} \right]} \Delta(\Phi(p_a)) \right) \right] \Delta(\Phi(p_b)) \right\}$$
$$\times \frac{1}{(1+\sigma^2)^2} \frac{1}{(1+\delta^2)^2}.$$

We propose to draw samples from $\underset{\sim}{\beta}, \sigma^2, \delta^2$ jointly by applying M-H sampler. Target function is $\pi(\underset{\sim}{\beta}, \sigma^2, \delta^2 | \underset{\sim}{y})$. We set the proposal function as

$$\begin{pmatrix} \underset{\sim}{\beta} \\ \log \sigma^2 \\ \log \delta^2 \end{pmatrix} | \underset{\sim}{y} \sim \text{Normal} \left\{ \begin{pmatrix} \bar{\underset{\sim}{\beta}}_a \\ \log \bar{\sigma}_a^2 \\ \log \bar{\delta}_a^2 \end{pmatrix}, \sigma_t^2 \Sigma_a \right\},$$

where $\frac{t}{\sigma_t^2} \sim \chi_t^2$, Chi-square on t degree of freedom, i.e. $\begin{pmatrix} \underset{\sim}{\beta} \\ \log \sigma^2 \\ \log \delta^2 \end{pmatrix} | \underset{\sim}{y} \sim$ Student's t. Here t is tuning constant.

We also use M-H sampler draw samples for $\underset{\sim}{v}$ and $\underset{\sim}{w}$ respectively. Proposal functions are $\pi_a(\underset{\sim}{v} \mid \underset{\sim}{\beta}_{(0)}, \beta_0, \sigma^2, \underset{\sim}{y})$ and $\pi_a(\underset{\sim}{w} \mid \underset{\sim}{v}, \underset{\sim}{\beta}_{(0)}, \sigma^2, \underset{\sim}{y})$ respectively from the INNA method.
The target function to draw $\underset{\sim}{v}$ is

$$\pi(\underset{\sim}{v} | \underset{\sim}{\beta}, \sigma^2, \delta^2, \underset{\sim}{y}) \propto \prod_{i=1}^{\ell} \prod_{j=1}^{n_i} \left\{ \sum_{a=1}^{100} \frac{e^{\sum_{k=1}^{m_{ij}} (\underset{\sim}{x}'_{ijk} \underset{\sim}{\beta}_{(0)} + \sigma \bar{z}_a^w + v_i) y_{ijk}}}{\prod_{k=1}^{m_{ij}} \left[1 + e^{\underset{\sim}{x}'_{ijk} \underset{\sim}{\beta}_{(0)} + \sigma \bar{z}_a^w + v_i} \right]} \Big(\Phi(a) - \Phi(a-1) \Big) \right\}$$
$$\times \left(\frac{1}{\sqrt{2\pi\delta^2}} \right)^l \exp \left\{ - \sum_{i=1}^{l} \frac{(v_i - \beta_0)^2}{2\delta^2} \right\}$$

After we get samples for $\underset{\sim}{v}$, we can use M-H sampler to draw

$$\pi(\underset{\sim}{w} | \underset{\sim}{v}, \underset{\sim}{\beta}_{(0)}, \sigma^2, \underset{\sim}{y}) \propto \prod_{i=1}^{\ell} \prod_{j=1}^{n_i} \prod_{k=1}^{m_{ij}} \left[\frac{e^{(\underset{\sim}{x}'_{ijk} \underset{\sim}{\beta}_{(0)} + w_{ij}) y_{ijk}}}{1 + e^{\underset{\sim}{x}'_{ijk} \underset{\sim}{\beta}_{(0)} + w_{ij}}} \right] \times \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp \left\{ - \sum_{i=1}^{l} \sum_{j=1}^{n_i} \frac{(w_{ij} - v_i)^2}{2\sigma^2} \right\}.$$

REFERENCES

Albert, J. H. and Chib, S. (1993). Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association*, **88**, 669-679.

Battese, G. E., Harter, R., and Fuller, W. A. (1988). An Error Components Model for Prediction of County. Crop Areas Using Survey and Satellite Data. *Journal of the American Statistical Association*, **83(401)**, 28-36.

Beaumont, M. A., Zhang, W. and Balding, D. J. (2002). Approximate Bayesian Computation in Population Genetics. *Genetics*, **162**, 2025-2035.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press, Cambridge, United Kingdom.

Central Bureau of Statistics Thapathali. Nepal living Standards Survey 2003/04. *Statistical Report Volume 1*, Kathmandu, Nepal.

Cox, D. R. and Snell, E. J. (1972). *Analysis of Binary Data, Second Edition*. Chapman Hall. **2**, 31-32.

Dharmadhikari, S. W. and Joag-Dev, K. (1988). *Unimodality, Convexity and Applications*. Academic Press, New York.

Erciulescu, A. L. and Berg, E. (2014). Small Area Estimates for the Conservation Effects Assessment Project. *Frontiers of Hierarchical Modeling in Observational Studies, Complex Surveys and Big Data: A Conference Honoring Professor Malay Ghosh, College Park, MD; Women in Statistics Conference, Cary, NC*.

Farrell, P.J., MacGibbon, B. and Tomberlin, T.J. (1997). Empirical Bayes Small Area Estimation Using Logistic Regression Models and Summary Statistics. *Journal of Business and Economic Statistics*, **15**, 101-108.

Fay, R.E. and Herriot, R.A. (1979). Estimates of Income for Small Places: an Application of James-Stein Procedures to Census Data. *Journal of the American Statistical Association*, **74**, 269-277.

Fearnhead, P. and Prangle, D. (2012). Constructing Summary Statistics for Approximate Bayesian computation: Semi-automatic ABC. *Journal of the Royal Statistical Society*, **B(74)**, 419-474.

Ferkingstad, E. and Rue, H. (2015). Improving the INLA approach for approximate Bayesian inference for latent Gaussian models. *Electronic Journal of Statistics*, **9**, 2706-2731.

Fong, Y., Rue, H. and Wakefield, J. (2010). Bayesian inference for generalized linear mixed models. *Biostatistics*, **11(3)**, 397-412.

Fuller, W.A. and Goyeneche, J.J. (1998). Estimation of the State Variance Component (unpublished).

Ghosh, M., Natarajan, K., Stroud, T.W.F. and Carlin, B.P. (1998). Generalized Linear Models for Small-Area Estimation. *Journal of the American Statistical Association*, **93(441)**, 273-282.

Held, L., Schrödle, B. and Rue, H. (2010). Posterior and Cross-Validatory Predictive Checks: A Comparison of MCMC and INLA. *Statistical Modeling and Regression Structures*, Physica-Verig HD, 91-110.

Holmes, C. C. and Held, L. (2006). Bayesian Auxiliary Variable Models for Binary and Multinomial Regression. *Bayesian Analysis*, **1(1)**, 145-168.

Illian, J. B., Sørbye, S. H. and Rue, H. (2012). A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). *The Annals of Applied Statistics*, **6(4)**, 1499-1530.

MacGibbon, B.R.E.N.D.A. and Tomerlin, T.J. (1989). Small Area Estimation of Proportions Via Empirical Bayes Techniques. *Survey Methodology*, **15**, 237-252.

Miroshnikov, A. and Conlon, E. M. (2015). Parallel Markov Chain Monte Carlo for Non-Gaussian Posterior Distributions. *Department of Mathematics and Statistics, University of Massachusetts, Amherst, Massachusetts*, pp. 1-32.

Nandram, B. (2000). Bayesian Generalized Linear Models for Inference about Small Areas. *In Generalized linear* *models: A Bayesian Perspective*, Eds. D. K. Dey, S.K. Ghosh and B.K. Mallick, Marcel Dekker, Chapter 6, 91-114.

Nandram, B. (1989). Discrimination between Complementary Log-log and Logistic Model for Ordinal Data. *Communications in Statistics, Theory and Methods*, **18**, 2155-2164.

Nandram, B. and Chen, M-H. (1996). Reparameterizing the Generalized Linear Model to Accelerate Gibbs Sampler Convergence. *Journal of Statistical Computation and simulation*, **54**, 129-144.

Nandram, B., Chen, L., Fu, S-T. and Manandhar, B. (2018). Bayesian Logistic Regression for Small Areas with Numerous Households. *Statistics and Application*, **1**, 171-205.

Nandram, B. and Choi, J. W. (2010). A Bayesian Analysis of Body Mass Index Data from Small Domains Under Nonignorable Nonresponse and Selection. *Journal of the American Statistical Association*, **105**, 120-135.

Nandram, B. and Erhardt, E. (2005). Fitting Bayesian Two-Stage Generalized Linear Models Using Random Samples via the SIR Algorithm. *Sankhya*, **66**, 733-755.

Nandram, B. and Sedransk, J. (1993). Bayesian Predictive Inference for a Finite Population Proportion: Two-Stage Cluster Sampling. *Journal of the Royal Statistical Society. Series B (Methodological)*, 399-408.

Potthoff, R.F., Woodbury, M.A. and Manton, K.G. (1992). Equivalent Sample Size and Equivalent Degrees of Freedom Refinements for Inference Using Survey Weights under Superpopulation Models. *Journal of the American Statistical Association*, **87**, 383-396.

Rao, J.N. (2015). Small Area Estimation. *John Wiley & Sons*.

Roberts, G., Rao, J. N. K. and Kumar, S. (1987). Logistic Regression Analysis of Sample Survey Data. *Biometrika*, **74(1)**, 1-12.

Rubin, D. B. (1981). The Bayesian Bootstrap. *The Annals of Statistics*, **9(1)**, 130-134.

Rue, H., Martino, S. and Chopin, IN. (2009). Approximate Bayesian Inference for Latent Gaussian Models Using Integrated Nested Laplace Approximations. *Journal of the Royal Statistical Society* Series B, **71(2)**, 319-392.

Scott, S. L., Blocker, A. W., Bonassi1, F. V., Chipman, H. A., George, E. I. and McCulloch, R. E. (2013). Bayes and Big Data: The Consensus Monte Carlo Algorithm. *Technical Report*, Google, Inc., 1-22.

Stukel, D.M. and Rao, J.N.K. (1997). Estimation of Regression Models with Nested Error Structure and Unequal Error Variances Under Two and Three Stage Cluster Sampling. *Statistics & Probability Letters*, **35(4)**, 401-407.

Stukel, D.M. and Rao, J.N.K. (1997). On Small-Area Estimation under Two-Fold Nested Error Regression Models. *Journal of Statistical Planning and Inference*, **78(1-2)**, 131-147.

Torabi, M. and Rao, J. N. K. (2014). On Small Area Estimation under a Sub-Area Level Model. *Journal of Multivariate Analysis*,**127**, 36-55.

Wang, J. and Fuller, W.A. (2003). The Mean Squared Error of Small Area Predictors Constructed With Estimated Area Variances. *Journal of the American Statistical Association*,**98(463)**, 716-723.

Yan, G. and Sedransk, J. (2007). Bayesian Diagnostic Techniques for Detecting Hierarchical Structure. *Bayesian Analysis*,**2(4)**, 735-760.

Yan, G. and Sedransk, J. (2010). A Note on Bayesian Residuals as A Hierarchical Model Diagnostic Technique. *Statistical Papers*,**51(1)**, 1.

You, Y. and Chapman, B. (2006). Small Area Estimation Using Area Level Models and Estimated Sampling Variances. *Survey Methodology*,**32**, 97-103.