

Article

Bayesian Logistic Regression Model for Sub-Areas

Lu Chen ^{1,*}  and Balgobin Nandram ² ¹ National Institute of Statistical Sciences, 1750 K Street NW Suite 1100, Washington, DC 20006, USA² Worcester Polytechnic Institute, Stratton Hall 103, Worcester, MA 01609, USA

* Correspondence: lchen@niss.org

Abstract: Many population-based surveys have binary responses from a large number of individuals in each household within small areas. One 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 the sampled wards (small areas). To make an 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.

Keywords: hierarchical Bayesian model; integrated nested normal approximation; MCMC; metropolis sampler; numerical integration; parallel computing



Citation: Chen, L.; Nandram, B. Bayesian Logistic Regression Model for Sub-Areas. *Stats* **2023**, *6*, 209–231. <https://doi.org/10.3390/stats6010013>

Academic Editor: Wei Zhu

Received: 27 October 2022

Revised: 4 January 2023

Accepted: 17 January 2023

Published: 29 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The Nepal Living Standard Survey (NLSS) II (see [1]) is a two-stage stratified sampling. A random sample of wards (areas) were selected from six strata and 12 households (sub-areas) were selected from each sampled ward. All individuals in each sampled household were interviewed. One interest is on health status, a binary variable. To make smooth estimates of the finite population proportion of the 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 samples are available—see [2].

In many applications, some areas, e.g., states and wards, are sampled; in each sampled area, a sample of sub-areas, e.g., counties and households, is further selected. Ref. [3] proposed a one-fold hierarchical Bayesian logistic regression model and applied the model to NLSS II data. The main objective is to make an inference for the finite population proportion of individuals with a specific character for each area. However, the one-fold model ignores the sub-area level structure in the data. As an extension of [3], we are particularly interested in small area models that can capture the hierarchical structure of the NLSS II data in this paper. 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 designs of many population-based surveys were two-stage stratified sampling as

NLSS II. But if we use a one-fold unit level model to fit the data, the sub-area level effects will have been ignored. Ref. [4] 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 using posterior predictive p -values. Ref. [5] 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 these kinds of models. Much of the literature focuses on continuous data. Ref. [6] 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 were studied by [2,7–9], and many others. This type of model is an area-level model which extends the Fay-Harriot model (see [10]) to the sub-area level. Two-fold nested error regression models were considered by [11,12]. On the other hand, some literature focus on the categorical data. Ref. [13] described a HB model to make an inference about the finite population proportion under two-stage clustering sampling. Ref. [14] extended the Beta-Binomial model to the two-fold model and used Gibbs sampling to obtain the posterior estimates. Ref. [15] showed that the two-fold Beta-Binomial model is preferable over the one-fold one if the data have a hierarchical structure. Ref. [16] extended [15] to accommodate heterogeneous correlations. They used a HB model to make a posterior inference about the finite population proportion of each area, accounting for intracluster correlations. Ref. [17] discussed the sub-area Beta-Binomial model and applied the model to estimate the finite population proportion of healthy individuals in each household covered by the NLSS II, assuming no covariate was available.

Bayesian logistic regression models with random effects are suitable for handling binary data with covariates. Ref. [18] discussed discrimination between the logit and the complementary log-log link functions by using the logistic regression model. Roberts, Ref. [19], discussed logistic regression for the sample survey data (not small area estimation). Ref. [20] showed how to accelerate the Gibbs sampler for a model with latent variables introduced earlier by [21] for Bayesian probit analysis. Ref. [22], discussed the logistic regression model by using the empirical Bayesian approach. Ref. [23] showed how to analyze binary data with covariates to maintain conjugacy for both the logistic and Poisson regression models. The analysis of binary data with covariates under nonignorable nonresponse was discussed by [24]. Ref. [3] proposed a hierarchical Bayesian logistic regression model for binary data in a small area estimation. This model is a unit level model without a sub-area effect. Our two-fold sub-area model is an extension of this logistic regression model. We add the sub-area level random effect into the model which can capture the hierarchical structure of the sampled data. At 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 solves the difficulties.

The other side of our application is that there are numerous small areas (households and individuals) and MCMC methods, which involve complicated integrals, and cannot handle them efficiently. “Big data” are defined as data that are too big to comfortably process on a single machine [25]. The researchers considered consensus Monte Carlo methods that split the data across 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 were discussed by [26]. Fortunately, in survey sampling, the design generally uses a 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 integration involved in Bayesian inference is usually intractable, which is true for our logistic regression model. The approximation techniques are desired. The procedure we used to approximate the posterior density of the parameters of the logistic regression sub-area model, INNA, is similar to the integrated nested Laplace approximation (INLA) originally proposed by [27], but they are actually different. INLA is a quite popular algo-

rithm 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, the computation of modes becomes time-consuming and challenging for the logistic regression model or any generalized linear mixed models. Yet, INLA has found many useful applications, such as in Poisson regression by [28], and in spatial point pattern data by [29]. We note that INLA can be problematic, especially for logistic and Poisson hierarchical regression models, even if the modes can be computed. Ref. [30], attempted to improve INLA using a copula-based correction, which adds complexity to INLA. Our approximation method, INNA, which does not require finding posterior modes, uses a sampling-based procedure accommodated by the multiplication rule of probability. Instead of finding the posterior modes, INNA finds the approximate modes in closed form, facilitated by the empirical logistic transform ([31]) 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 [32–34].

In Section 2, a full description of a sub-area HB logistic regression model is given. In Section 3, we describe the integrated nested normal approximation (INNA) computation method and some theoretical results are provided. The exact MCMC method is presented in Appendix A. The exact method refers to MCMC methods without further approximation. In Section 4, we apply the 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. Some comparisons between INNA and the exact method are presented. Finally, in Section 5, we make concluding remarks and discuss the future work.

2. Sub-Area Logistic Regression Model

In this section, we discuss the sub-area HB logistic regression model at the unit level. In the NLSS II data, we have binary data (good health versus poor health) for each individual within a household, and these households are within wards. The observations are available at the unit level and so is the reliable auxiliary information. However, the model and method we proposed for small areas and sub-areas is not only for this application on NLSS II data. It can be also applied to other population-based surveys with binary responses which contain small areas or/and sub-areas.

Suppose that there are L small areas (wards) in the finite population and that, 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 the sampled households are sampled. Here, we assume the survey weights are the same within all households in each area. Actually, the design is almost self-weighting.

Let y_{ijk} , $k = 1, \dots, m_{ij}$, $j = 1, \dots, n_i$, $i = 1, \dots, \ell$ denote the binary responses. Let $y = (y_{ijk}, k = 1, \dots, m_{ij}, j = 1, \dots, n_i, i = 1, \dots, \ell)'$. Let $y_{ij} = \sum_{k=1}^{m_{ij}} y_{ijk}$ be the number with response 1 and m_{ij} be the total number of people who responded. Let $x_{ijk} = (1, x_{ijk1}, \dots, x_{ijkp})'$ be the $(p + 1)$ vector with p covariates for individuals and an intercept.

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

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

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

$$\begin{aligned} y_{ijk} | \beta, v_i, \mu_{ij} &\stackrel{\text{ind}}{\sim} \text{Bernoulli} \left\{ \frac{e^{x'_{ijk} \beta + v_i + \mu_{ij}}}{1 + e^{x'_{ijk} \beta + v_i + \mu_{ij}}} \right\}, k = 1, \dots, m_{ij}, \\ \mu_{ij} | \sigma^2 &\stackrel{\text{iid}}{\sim} \text{Normal}(0, \sigma^2), j = 1, \dots, n_i, \\ v_i | \delta^2 &\stackrel{\text{iid}}{\sim} \text{Normal}(0, \delta^2), i = 1, \dots, \ell, \\ \pi(\beta, \delta^2, \sigma^2) &\propto \frac{1}{(1 + \delta^2)^2} \frac{1}{(1 + \sigma^2)^2}, \delta^2 > 0, \sigma^2 > 0. \end{aligned} \quad (1)$$

Here, μ_{ij} , $i = 1, \dots, \ell; j = 1, \dots, n_i$ are the sub-area level random effects, which are not in the area-level model in [3]. v_i , $i = 1, \dots, \ell$ are the area random effects and $\beta = (\beta_0, \beta_1, \dots, \beta_p)'$ are the regression coefficients, with σ^2, δ^2 as the variance of the random effects, respectively.

In order to apply our approximation method and make an inference for posterior distribution, we use an equivalent model.

First, we separate β into β_0 and $\beta_{(0)}$, where $\beta_{(0)} = (\beta_1, \beta_2, \dots, \beta_p)'$. We set β_0 as the mean of v , and then we can omit the intercept term from the covariate x_{ijk} . Second, we introduce a new parameter, $w_{ij} = v_i + \mu_{ij}$, in order to set v_i and μ_{ij} independently and make it easy to make an inference for both of them. We have

$$\begin{aligned} y_{ijk} | w_{ij}, \beta_{(0)} &\stackrel{\text{ind}}{\sim} \text{Bernoulli} \left\{ \frac{e^{x'_{ijk} \beta_{(0)} + w_{ij}}}{1 + e^{x'_{ijk} \beta_{(0)} + w_{ij}}} \right\}, k = 1, \dots, m_{ij}, \\ w_{ij} | v_i, \sigma^2 &\stackrel{\text{ind}}{\sim} \text{Normal}(v_i, \sigma^2), j = 1, \dots, n_i, \\ v_i | \beta_0, \delta^2 &\stackrel{\text{iid}}{\sim} \text{Normal}(\beta_0, \delta^2), i = 1, \dots, \ell, \\ \pi(\beta, \delta^2, \sigma^2) &\propto \frac{1}{(1 + \delta^2)^2} \frac{1}{(1 + \sigma^2)^2}, \delta^2 > 0, \sigma^2 > 0. \end{aligned} \quad (2)$$

The joint posterior density for the parameters is

$$\begin{aligned} \pi(v, w, \beta, \sigma^2, \delta^2 | y) &\propto \prod_{i=1}^{\ell} \prod_{j=1}^{n_i} \prod_{k=1}^{m_{ij}} \left[\frac{e^{(x'_{ijk} \beta_{(0)} + w_{ij}) y_{ijk}}}{1 + e^{x'_{ijk} \beta_{(0)} + w_{ij}}} \right] \times \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp \left\{ - \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \frac{(w_{ij} - v_i)^2}{2\sigma^2} \right\} \\ &\times \left(\frac{1}{\sqrt{2\pi\delta^2}} \right)^l \exp \left\{ - \sum_{i=1}^{\ell} \frac{(v_i - \beta_0)^2}{2\delta^2} \right\} \frac{1}{(1 + \sigma^2)^2} \frac{1}{(1 + \delta^2)^2}. \end{aligned} \quad (3)$$

The posterior density is a non-standard multivariate 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 the INNA method for the sub-area HB logistic regression model. It is an extension of the INNA method in [3]. INNA method is not required to find the posterior modes. Due to the large amount of sub-areas, it would be time-consuming to find all posterior modes, which is why we did not choose the popular INLA method. In detail, we discuss the approximation of the joint posterior density (3).

Notice that the joint posterior density (3) is very complicated and it is the expit part, $\prod_{i=1}^{\ell} \prod_{j=1}^{n_i} \prod_{k=1}^{m_{ij}} \left[\frac{e^{(x'_{ijk} \beta_{(0)} + w_{ij}) y_{ijk}}}{1 + e^{x'_{ijk} \beta_{(0)} + w_{ij}}} \right]$, that causes the difficulties. In the following, we discuss how 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). This is the key contribution in the paper. Then we use the multiplication rule to approximate the joint posterior density,

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

where the first three densities on the right-hand side are all multivariate normal densities. Therefore, we can draw samples and make inference through the approximate joint posterior density.

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

Lemma 1. Let $h(\underline{\tau})$ be a logconcave density function with the parameter $\underline{\tau}$. Then, $\underline{\tau}$ approximately has a multivariate normal distribution,

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

Proof. Simply applying the second-order multivariate Taylor series of $h(\underline{\tau})$ at $\underline{\tau}^*$ gives

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

Due to the logconcavity of $h(\underline{\tau})$, its Hessian Matrix $-H$ is posit-definite, which can be the covariance matrix. Notice that we are not required to use the mode of $h(\underline{\tau})$. We do not need to find the solution of the gradient vector $\underline{g} = 0$. Therefore, $\underline{\tau}^*$ does not have to be the solution but some other point. It is worth noticing that the term, $-H^{-1}\underline{g}$, is a correction to $\underline{\tau}^*$. \square

To illustrate the approximation steps, we start with a simpler model with flat priors for $\underline{\beta}_{(0)}$ and the \underline{w} , according to model (2). That is,

$$y_{ijk} \mid w_{ij}, \underline{\beta}_{(0)} \stackrel{\text{ind}}{\sim} \text{Bernoulli} \left\{ \frac{e^{x'_{ijk} \underline{\beta}_{(0)} + w_{ij}}}{1 + e^{x'_{ijk} \underline{\beta}_{(0)} + w_{ij}}} \right\}, k = 1, \dots, m_{ij}, j = 1, \dots, n_i, i = 1, \dots, \ell,$$

$$p(\underline{w}, \underline{\beta}_{(0)}) = 1. \quad (4)$$

The joint posterior density is

$$\pi(\underline{w}, \underline{\beta}_{(0)} \mid \underline{y}) \propto \prod_{i=1}^{\ell} \prod_{j=1}^{n_i} \prod_{k=1}^{m_{ij}} \left\{ \frac{e^{(x'_{ijk} \underline{\beta}_{(0)} + w_{ij}) y_{ijk}}}{1 + e^{x'_{ijk} \underline{\beta}_{(0)} + w_{ij}}} \right\}. \quad (5)$$

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

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

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

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

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

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

The first derivative of $q(\beta_{(0)})$ is

$$\begin{aligned} q'(\beta_{(0)}) &= \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \sum_{k=1}^{m_{ij}} \left\{ x_{ijk} y_{ijk} - \frac{x_{ijk} e^{(x'_{ijk} \beta_{(0)} + \hat{w}_{ij}^*)}}{1 + e^{x'_{ijk} \beta_{(0)} + \hat{w}_{ij}^*}} \right\} \\ &= \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \sum_{k=1}^{m_{ij}} \left\{ x_{ijk} y_{ijk} - x_{ijk} [1 + e^{-(x'_{ijk} \beta_{(0)} + \hat{w}_{ij}^*)}]^{-1} \right\}. \end{aligned}$$

Usually we should set $q'(\beta_{(0)})$ equal to zero and find the modes as the maximum likelihood estimator (MLE) of $\beta_{(0)}$. But here, it is not easy to solve the equation due to the complexity of $q'(\beta_{(0)})$. We use the first-order Taylor series to approximate it and then simplify $q'(\beta_{(0)})$ so that we can get quasi-modes of $\beta_{(0)}$.

The first-order Taylor expansion of $(1 + e^{x'_{ijk} \beta_{(0)} + \hat{w}_{ij}^*})^{-1}$ equals $(1 - e^{-(x'_{ijk} \beta_{(0)} + \hat{w}_{ij}^*)})$. Notice that by Taylor series, $e^{-(x'_{ijk} \beta_{(0)} + \hat{w}_{ij}^*)} \approx 1 - (x'_{ijk} \beta_{(0)} + \hat{w}_{ij}^*)$. Then we can get

$$\begin{aligned} q'(\beta_{(0)}) &\approx \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \sum_{k=1}^{m_{ij}} \left\{ x_{ijk} y_{ijk} - x_{ijk} [(1 - e^{x'_{ijk} \beta_{(0)} + \hat{w}_{ij}^*})] \right\} \\ &\approx \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \sum_{k=1}^{m_{ij}} \left\{ x_{ijk} y_{ijk} - x_{ijk} [(1 - 1 + (x'_{ijk} \beta_{(0)} + \hat{w}_{ij}^*))] \right\} \\ &= \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \sum_{k=1}^{m_{ij}} \left\{ x_{ijk} (y_{ijk} - \hat{w}_{ij}^*) - x_{ijk} x'_{ijk} \beta_{(0)} \right\}. \end{aligned}$$

We can get the quasi-modes of $\beta_{(0)}$ by solving the equation $q'(\beta_{(0)}) = 0$. That is,

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

Second, we obtain quasi-modes for the w_{ij} , a refinement of the z_i . Plug $\beta_{(0)}^*$ into the likelihood function Δ and consider it as function w_{ij} only:

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

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

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

We can obtain the approximate posterior mode of w_{ij} by solving the equation $g'(w_{ij}) = 0$.

$$w_{ij}^* = \log \left\{ \frac{\sum_{k=1}^{m_{ij}} e^{-x'_{ijk} \beta_{(0)}^*}}{m_{ij}(1 - \bar{y}_{ij})} \right\}.$$

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

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

Let $\tau^{*'} = (\mu^{*'}, \beta_{(0)}^{*'})$. Next, we evaluate g and H at the quasi-modes $\tau = \tau^*$ can also be obtained as

$$g = \left(\frac{\partial \Delta}{\partial w_{11}} \quad \cdots \quad \frac{\partial \Delta}{\partial w_{\ell n_\ell}} \quad \frac{\partial \Delta}{\partial \beta_{(0)}} \right)^T_{w=w^*, \beta_{(0)}=\beta_{(0)}^*},$$

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

The partial derivatives can be expressed in terms of response y_{ijk} and covariates x_{ijk} as

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

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

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

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

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

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

For the convenience of computation, denote $g = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}$ and $H = - \begin{pmatrix} D & C' \\ C & B \end{pmatrix}$, where

$$g_1 = \left(\frac{\partial \Delta}{\partial w_{11}} \quad \cdots \quad \frac{\partial \Delta}{\partial w_{\ell n_\ell}} \right)^T, g_2 = \frac{\partial \Delta}{\partial \beta_{(0)}},$$

$$B = - \frac{\partial^2 \Delta}{\partial \beta_{(0)}^2}, C = - \left(\frac{\partial^2 \Delta}{\partial w_{11} \partial \beta_{(0)}} \quad \cdots \quad \frac{\partial^2 \Delta}{\partial w_{\ell n_\ell} \partial \beta_{(0)}} \right), D = - \begin{pmatrix} \frac{\partial^2 \Delta}{\partial w_{11}^2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\partial^2 \Delta}{\partial w_{\ell n_\ell}^2} \end{pmatrix}.$$

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

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

Lemma 2. 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, $\tau|y$ in (5), is logconcave.

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

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

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

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

$$C = (c_{ij}), \quad c_{ij} = \frac{\partial^2 \Delta}{\partial w_{ij} \partial \beta_{(0)}} = \sum_{k=1}^{m_{ij}} p_{ijk}(1 - p_{ijk}) x_{ijk},$$

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 [35]). 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

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

It is now easy to show that

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

Therefore, $-H$ is positive-definite, and $\tau|y$ is logconcave. \square

Finally, according to the Lemmas 1 and 2, we can establish the approximation Theorem.

Theorem 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, $\tau|y$ in (5) is approximately a multivariate normal density, and the conditional posterior density of $w|\beta_{(0)}, y$ and $\beta_{(0)}|y$ can also be approximated by multivariate normal distributions.

Proof. The proof is given in Appendix B. \square

Therefore, we can approximate that logit expit term $\prod_{i=1}^{\ell} \prod_{j=1}^{n_i} \prod_{k=1}^{m_{ij}} \left[\frac{e^{(x'_{ijk}\beta_{(0)} + w_{ij})y_{ijk}}}{1 + e^{x'_{ijk}\beta_{(0)} + w_{ij}}} \right]$ into two multivariate densities by Theorem 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

$$\pi(w, \nu, \beta, \sigma^2, \delta^2 | y) \propto \pi(y|w, \beta_{(0)}) \pi(w | \nu, \sigma^2) \pi(\nu | \beta_0, \delta^2) \pi(\beta_{(0)}, \beta_0, \sigma^2, \delta^2)$$

The likelihood function $\pi(y|\underline{w}, \underline{\beta}_{(0)})$ can be approximated by the multivariate normal distribution by Theorem 1. Combining the prior values of \underline{w} and \underline{v} given by our Bayesian Logistic model and the results in Theorem 1, we can obtain our INNA model

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

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

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

$$\underline{v}|\beta_0, \delta^2 \stackrel{\text{ind}}{\sim} \text{Normal}(\beta_0 \underline{j}, \delta^2 I),$$

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

where $\underline{\mu}'_v = (\underbrace{v_1, \dots, v_1}_{n_1} \dots \underbrace{v_\ell, \dots, v_\ell}_{n_\ell})'$ and \underline{j} is a vector of ones.

Using Bayes' Theorem and the multiplication rule, the posterior density $\pi(\underline{w}, \underline{v}, \underline{\beta}, \sigma^2, \delta^2 | y)$ can be approximated as

$$\begin{aligned} \pi_a(\underline{w}, \underline{v}, \underline{\beta}, \sigma^2, \delta^2 | y) &\propto \pi_a(\underline{w} | \underline{v}, \underline{\beta}_{(0)}, \sigma^2, y) \pi_a(\underline{v} | \beta_0, \delta^2, y) \pi_a(\underline{\beta}_{(0)} | y) \pi_a(\underline{\beta}, \sigma^2, \delta^2 | y) \\ &= e^{-\frac{1}{2} \left\{ \left[\underline{w} - (\underline{\mu}_w - D^{-1}C'(\underline{\beta}_{(0)} - \underline{\mu}_\beta)) \right]' D \left[\underline{w} - (\underline{\mu}_w - D^{-1}C'(\underline{\beta}_{(0)} - \underline{\mu}_\beta)) \right] \right\}} \\ &\times e^{-\frac{1}{2} \left\{ \left[\underline{w} - \underline{\mu}_v \right]' (\sigma^2 I)^{-1} \left[\underline{w} - \underline{\mu}_v \right] + \left[\underline{v} - \beta_0 \underline{j} \right]' (\delta^2 I)^{-1} \left[\underline{v} - \beta_0 \underline{j} \right] + \left[\underline{\beta}_{(0)} - \underline{\mu}_\beta \right]' G^{-1} \left[\underline{\beta}_{(0)} - \underline{\mu}_\beta \right] \right\}} \\ &\times \frac{|D|^{1/2}}{|\delta^2 I|^{1/2} |\sigma^2 I|^{1/2} |G|^{1/2}} \frac{1}{(1 + \sigma^2)^2} \frac{1}{(1 + \delta^2)^2} \end{aligned} \quad (6)$$

Therefore, we can get the following key result.

Theorem 2. Using the multiplication rule, the joint posterior density, $\pi(\underline{w}, \underline{v}, \underline{\beta}, \sigma^2, \delta^2 | y)$ in (6), can be approximated by

$$\pi_a(\underline{w}, \underline{v}, \underline{\beta}, \sigma^2, \delta^2 | y) \propto \pi_a(\underline{w} | \underline{v}, \underline{\beta}_{(0)}, \sigma^2, y) \pi_a(\underline{v} | \beta_0, \delta^2, y) \pi_a(\underline{\beta}_{(0)} | y) \pi_a(\underline{\beta}, \sigma^2, \delta^2 | y),$$

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

Proof. The proof is given in Appendix C. \square

The INNA is actually a random sampler. First, we draw samples for σ^2, δ^2 from $\pi(\sigma^2, \delta^2 | y)$. The posterior distribution of $\sigma^2, \delta^2 | y$ does not have standardized form. Here, we use the 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 $\phi_1, \phi_2 | y$ is

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

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

$$\pi(\phi_1, \phi_2 | y) = \pi(\phi_2 | \phi_1) \pi(\phi_1 | y) = \pi(\phi_2 | \phi_1) \int_0^1 \pi(\phi_1, \phi_2 | y) d\phi_2.$$

We plug each grid of $\phi_1 \in (0, 1)$ into $\int_0^1 \pi(\phi_1, \phi_2 | y) d\phi_2$ and then use numerical integration to get the density of $(\phi_1 | y)$. After we plug all the 100 grids, we can get 100 value of $\pi(\phi_1 | y)$ 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 $\phi_2^{(h)}$. We repeat those steps 10,000 times to get the sample of $(\phi_1^{(h)}, \phi_2^{(h)})$, $h = 1, \dots, 10,000$. 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 , we can simply draw samples of β from the approximate multivariate normal distribution $\pi_a(\beta | \sigma^2, \delta^2, y)$. Third, we can draw samples of v_i independently given β, δ^2 and data from the approximate normal distribution $\pi_a(v | \beta_0, \delta^2, y)$. Finally, samples of w_{ij} independently given v, β, σ^2 can be obtained from the approximate normal distribution $\pi_a(w | v, \beta_{(0)}, \sigma^2, y)$. Notice that the last three steps are very simple, just drawing samples from normal densities. In addition, w_{ij} and v_i are all independent so that we can draw them simultaneously. Therefore, those latter steps permit fast computing.

In order to check if INNA method can provide resonal results, we apply the MCMC logistic regression exact method to the sub-area model. The idea of exact method is to get full conditional posterior distributions for all of the parameters in the model, and then get a large number of independent samples of each parameter with its full conditional posterior density. Details are given in Appendix A.

There are two differences between these two methods. First, both methods are sampling-based. The approximate method implements random samples and the exact method uses numerical integration method and Markov chains. Second, $\pi_a(\beta, \sigma^2, \delta^2 | y)$ is used for the INNA method. In the exact method, a Metropolis step is used for the $\pi(\beta, \sigma^2, \delta^2 | y)$. This is very time-consuming in the exact method. On the other hand, the exact method actually uses the INNA method. We use M-H sampler draw samples for v and w , respectively. Proposal functions are $\pi_a(v | \beta_{(0)}, \beta_0, \sigma^2, y)$ and $\pi_a(w | v, \beta_{(0)}, \sigma^2, y)$, respectively, from the INNA method.

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 the NLSS II is to track changes in and the progress of national living standards and social indicators of the 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 Primary Sampling Units (PSUs) and, within each PSU, there are a number of households (sub-area) selected. All household members in the sample were interviewed.

In detail, the NLSS II has records for 20,263 individuals from 3912 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 the strata using probability proportional to size (PPS) sampling and 12 households were systematically selected from each PSU. The survey is self-weighted and some adjustments were made after conducting the survey for non-responses 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	1152	408	1224	3912
Individuals	1949	1954	1467	5755	2104	7034	20,263

We chose four relevant covariates which can influence health status from the same NLSS II survey for our two-fold logistic regression model. They are age, nativity, sex and religion. We created binary variables for 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 four covariates. In the model fitting, we standardized the age covariate. Older age and a child's age are more vulnerable times than younger age. Indigenous people can have different health statuses from migrated people.

Table 2. The descriptives of 4 covariates.

Covariates		Frequency	Percentage
Age	0–14	7765	38.32
	15–59	10,951	54.04
	60+	1547	7.64
Gender	Male	9763	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	3385	19.17

According to the 2001 census data, only about 0.091% of households and only 0.904% of PSU were sampled. The NLSS II was designed to provide reliable estimates only at the stratum level, or even larger areas than the stratum. It cannot give estimates in small areas (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 chose the binary variable, health status, from the health section of the questionnaire.

4.2. Numerical Comparison

We used data from NLSS II to illustrate our sub-area logistic regression model. We predicted the household proportions of members in good health for 18,924 households (sampled and non-sampled). Bayesian bootstrap by [36] was applied to get non-sampled auxiliary information. This analysis was based on 1224 sample households from 102 wards (PSUs) in strata 6. Our primary purposes were 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.

We used Rcpp [37] and RcppArmadillo [38] packages in R [39] to fit the model based on both the approximation, INNA, method and the exact method to this NLSS II dataset. For the INNA method, we began with 10,000 iterations and a burn-in of 1000 and we kept only every ninth sample. Finally, 1000 samples were obtained for constructing the posterior distributions of all the parameters. The exact method was very time-consuming, taking about 30 h to finish. However, the INNA approximation method can get samples in 8 min. When we have a large number of areas or sub-areas, the approximation method will make enormous savings.

Convergence diagnostics were conducted. The convergences of the hyperparameters ($\beta, \sigma^2, \delta^2$) were monitored by the Geweke test of stationarity [40] and the effective sample sizes. The p-values and effective sample sizes are shown in Table 3, resulting in good

convergence for both methods. Table 3 also shows the posterior means (PMs) and associated posterior standard deviations (PSDs) of the hyperparameters. The PMs are very close between these two methods. The PSDs are slightly larger for the exact method than for the INNA method, but they are reasonably close.

Table 3. Posterior means (PM), associated posterior standard deviations (PSD), Geweke test p -values and effective sample sizes (ESS) of hyperparameters based on the INNA and exact method.

Method		INNA			Exact			
Estimator	PM	PSD	p -Value	ESS	PM	PSD	p -Value	ESS
β_0	0.802	0.323	0.688	780	0.840	0.405	0.264	1000
β_1	0.675	0.015	0.759	857	0.654	0.016	0.156	858
β_2	0.302	0.017	0.453	894	0.313	0.019	0.655	1000
β_3	−0.862	0.017	0.605	915	−0.824	0.017	0.418	863
β_4	−0.338	0.013	0.718	937	−0.360	0.011	0.839	1000
σ^2	115.248	35.601	0.408	1000	108.959	38.719	0.615	911
δ^2	30.002	1.146	0.448	731	29.034	1.697	0.490	780

In Figures 1–3, we compare the PMs, PSDs and posterior coefficient of variations (PCVs) in the household level as our primary purpose. We can see that the PMs are very close, nearly lying on the 45-degree line through the origin. The PSDs are slightly spread out and thicker, but all points still lie on the 45-degree line and so do the PCVs. Overall, these approximations are acceptable in the data analysis. Figures 4–6 we compare respectively to the PMs, PSDs and PCVs at the ward level. The plots of the PMs are still very good. Notice that other two plots of PSDs and PCVs are more spread out than those in the household level. Again, though, the approximate method and the exact method are reasonably close.

We also compare the approximate method with the exact method using the five number summaries (the minimum values, the first quartiles, the median, the third quartiles, and the maximum values) with respect to the PMs, PSDs and PCVs of the finite population proportions at the household level and ward level in Tables 4 and 5. The PMs from both methods at the household level have larger variations than those at the ward level. The PCVs at the ward level are generally much smaller than at the household level. The summaries of the PMs, PSDs and PCVs within households and wards between the approximate and exact methods are very close.

Table 4. Comparison of posterior inference about the finite population proportions using the five-number summaries at the household level.

Households		Min	Q1	Mean	Q3	Max
PM	INNA	0.049	0.541	0.565	0.584	0.981
	Exact	0.050	0.542	0.566	0.584	0.984
PSD	INNA	0.040	0.246	0.441	0.465	0.500
	Exact	0.038	0.248	0.442	0.466	0.501
PCV	INNA	0.041	0.379	0.762	0.852	1.549
	Exact	0.039	0.384	0.768	0.852	1.577

Table 5. Comparison of posterior inference about the finite population proportions using five-number summaries at the ward level.

Wards		Min	Q1	Mean	Q3	Max
PM	INNA	0.449	0.537	0.563	0.589	0.684
	Exact	0.450	0.537	0.564	0.590	0.683
PSD	INNA	0.056	0.059	0.063	0.066	0.077
	Exact	0.058	0.061	0.064	0.066	0.077
PCV	INNA	0.095	0.103	0.113	0.121	0.163
	Exact	0.097	0.103	0.113	0.122	0.165

We conclude that the approximation method at the household level is reasonable. The approximation is desirable because one can perform the computations in real time.

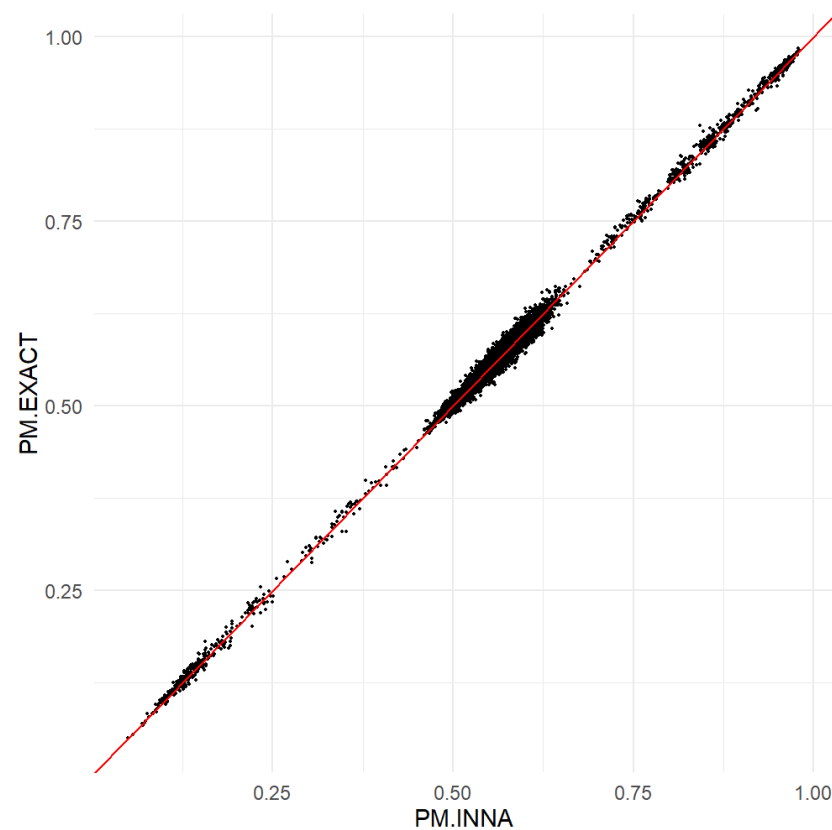


Figure 1. Comparison of the INNA method and the exact method using the PSDs of the household proportions.

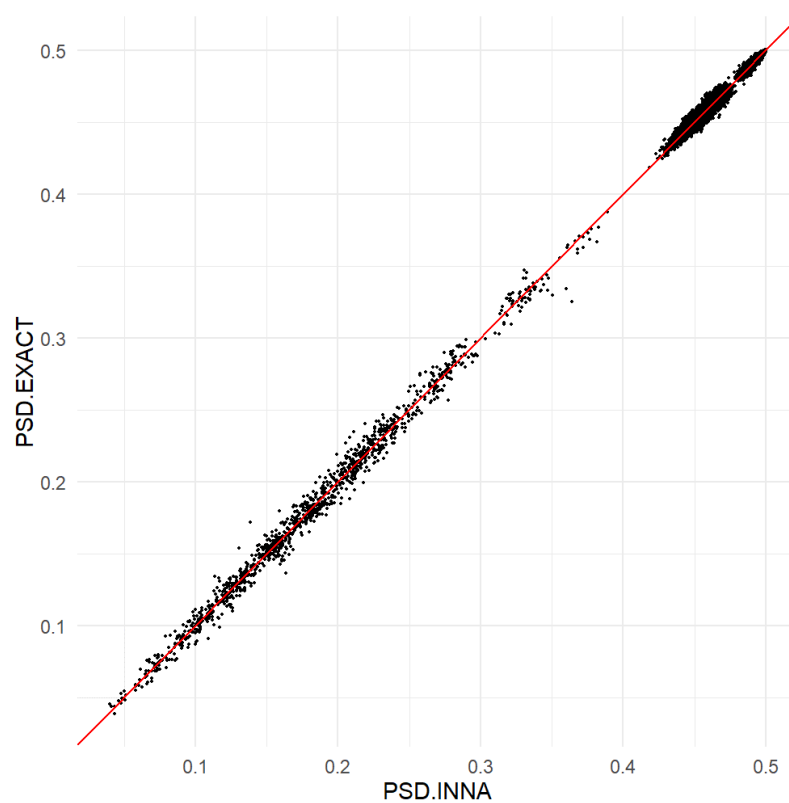


Figure 2. Comparison of the INNA method and the exact method using the PSDs of the household proportions.

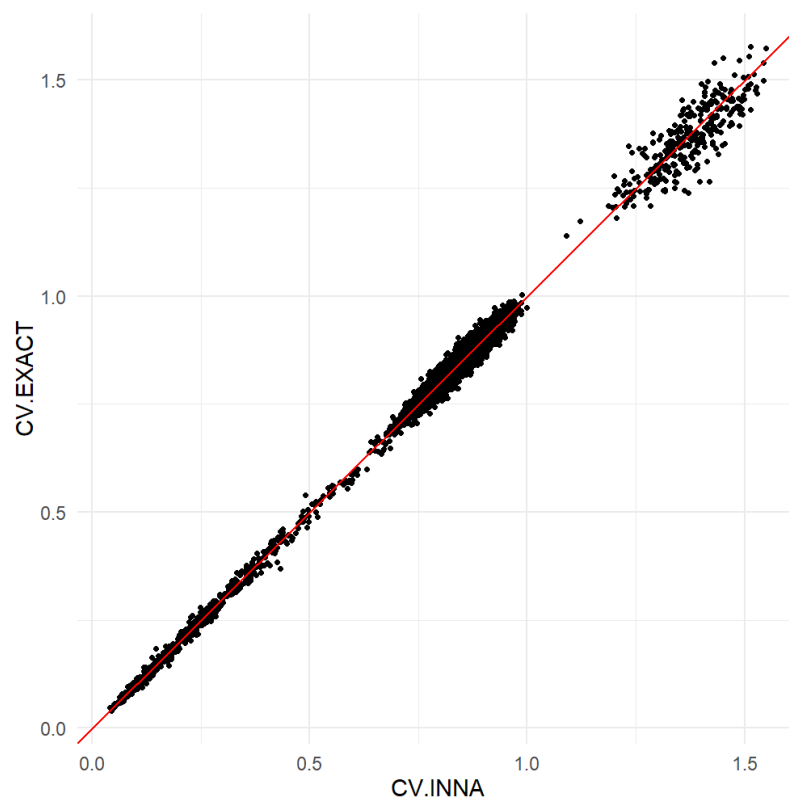


Figure 3. Comparison of the INNA method and the exact method using the CVs of the household proportions.

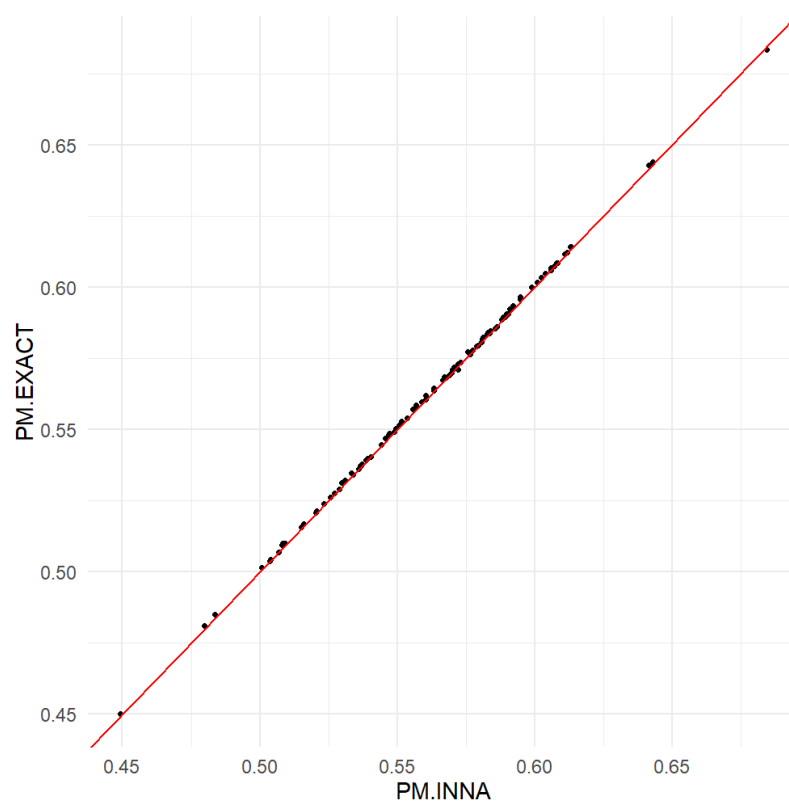


Figure 4. Comparison of the INNA method and the exact method using the PMs of the ward proportions.

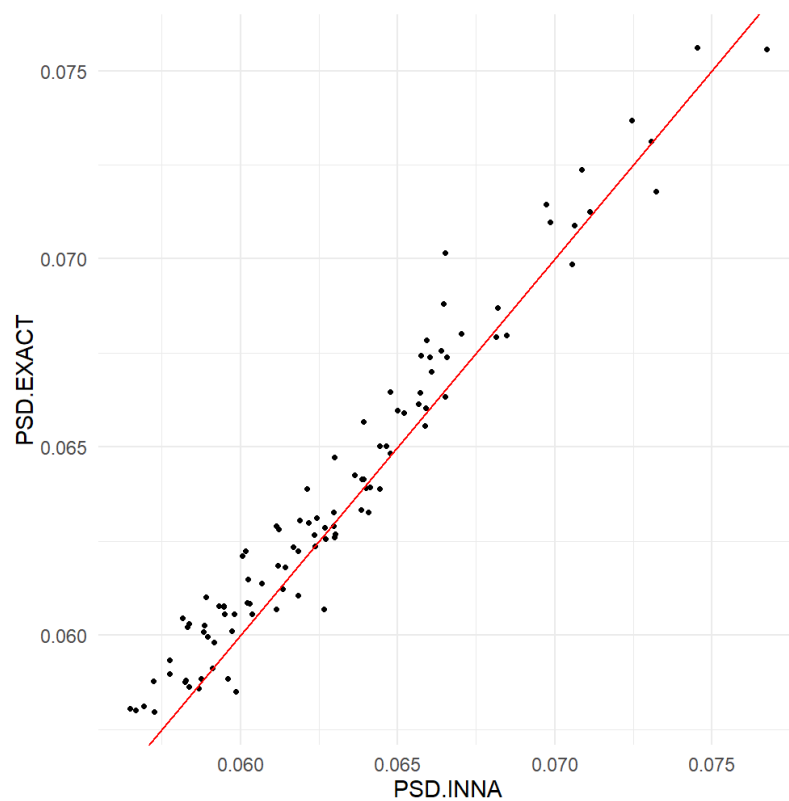


Figure 5. Comparison of the INNA method and the exact method using the PSDs of the ward proportions.

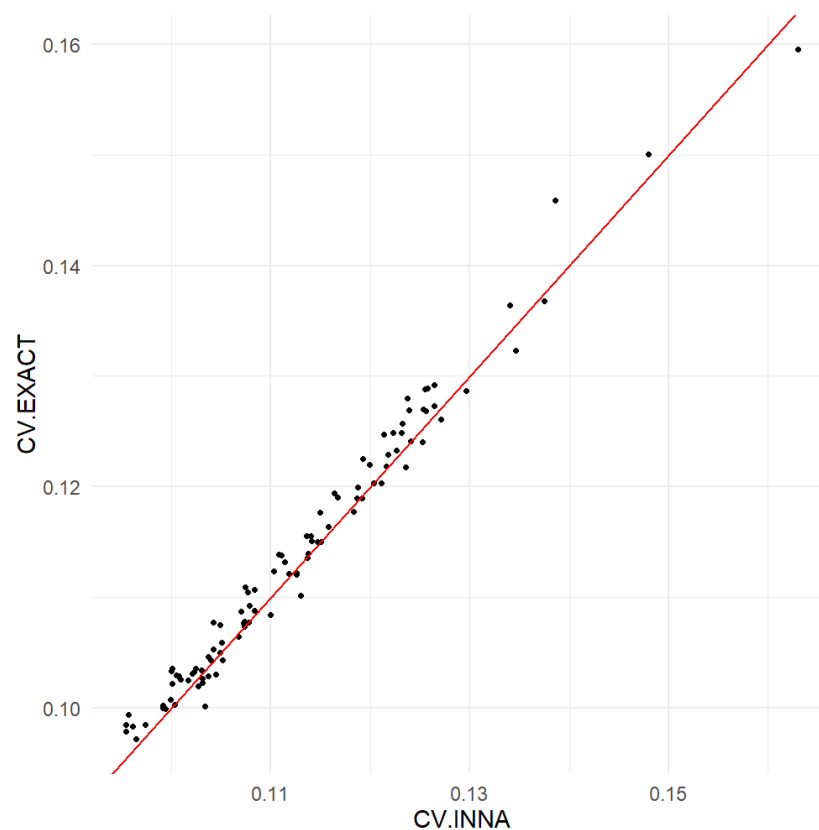


Figure 6. Comparison of the INNA method and the exact method using the CVs of the ward proportions.

5. Conclusions and Future Work

The sub-area HB logistic regression model can be applied to analyze the binary response variable. This model is an extension of the HB logistic regression area-level model, which ignores the actual hierarchical structure of the data. We propose an approximation method, INNA, to fit the model. For large datasets, it is very unrealistic to use the MCMC method to fit the model. We propose the approximation method, INNA, which saves time significantly because 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 the NLSS II is presented in order to compare the approximation method and the exact method. It shows that, when there are a large number of areas and sub-areas, the approximation method can be efficient and it can also provide reasonable estimates.

INNA is a method for approximate Bayesian inference based on Laplace's method, the second-order multivariate Taylor-series approximation and the empirical logistic transform (ELT). It can be applied to all HB logistic regression models, for which it can be a fast and accurate alternative to the Markov chain Monte Carlo methods. The comparison and model results illustrate the performance of the INNA methods based on the sub-area model.

There will be many future works on the two-fold small areas model. First, in this paper, we assume equal survey weights since the NLSS II is a self-weighted sampling. However, after the data are 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 in the model. The NLSS II is a national population-based survey. We should rescale the sample weights to sum to an equivalent sample size. That is, we consider the adjusted weight as

$$w_{ijk}^* = \hat{n} \left(\frac{w_{ijk}}{\sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \sum_{k=1}^{m_{ij}} w_{ijk}} \right), \text{ where } \hat{n} = \frac{\left(\sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \sum_{k=1}^{m_{ij}} w_{ijk} \right)^2}{\sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \sum_{k=1}^{m_{ij}} w_{ijk}^2} \text{ as an equivalent sample. Introducing the}$$

sampling weights, we can obtain an updated normalized likelihood function. Based on the updated likelihood function and the same prior in the two-fold model, we can have a full Bayesian analysis on the updated model and then project the finite population proportion of the family members with good health in each household.

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

Author Contributions: Conceptualization, B.N. and L.C.; methodology, B.N. and L.C.; software, L.C.; validation, B.N. and L.C.; formal analysis, L.C.; investigation, L.C.; resources, B.N.; writing—original draft preparation, L.C.; writing—review and editing, B.N. and L.C.; visualization, L.C.; supervision, B.N. All authors have read and agreed to the published version of the manuscript.

Funding: Balgobin Nandram was supported by a grant from the Simons Foundation (#353953, Balgobin Nandram).

Conflicts of Interest: The authors declare no conflict of interest.

Appendix 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

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

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 $\underline{\beta}, \sigma^2, \delta^2$ together and then draw \underline{v} given $\underline{\beta}, \sigma^2, \delta^2$ samples. At last, we use MH method to draw \underline{w} given $\underline{v}, \underline{\beta}, \sigma^2, \delta^2$ samples.

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

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

Notice that the integrant is not any simple distribution function, so we use Monte Carlo numerical integration to approximate the integrals. Let $z_{ij}^w = \frac{w_{ij} - v_i}{\sigma}$. 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 $[p_{a-1}, p_a], a = 1, \dots, M$. Then we can get an approximate but very accurate joint density

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

Let $\bar{z}_a^w = \frac{p_a - p_{a-1}}{2}$, which is the midpoint of each interval $[p_{a-1}, p_a], a = 1, \dots, M$. 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

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

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

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

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

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

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

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

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

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

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

Appendix B. Proof of Theorem 1

Proof. By Lemma 2, the posterior density is logconcave. Then according to Lemma 1, the posterior distribution $\tau | y$ is approximately a multivariate normal distribution.

By Lemma 1, evaluating all quantities at τ^* , the mean is

$$\begin{pmatrix} \mu_w \\ \mu_\beta \end{pmatrix} = \tau^* - H^{-1} \underline{g} = \begin{pmatrix} \underline{w}^* \\ \beta_{(0)}^* \end{pmatrix} + \begin{pmatrix} E & F' \\ F & G \end{pmatrix} \begin{pmatrix} \underline{g}_1 \\ \underline{g}_2 \end{pmatrix} = \begin{pmatrix} \underline{w}^* + E \underline{g}_1 + F' \underline{g}_2 \\ \beta_{(0)}^* + F \underline{g}_1 + G \underline{g}_2 \end{pmatrix}.$$

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 1, the approximate joint posterior density of $\underline{w}, \beta_{(0)} | y$ is

$$\begin{pmatrix} \underline{w} \\ \beta_{(0)} \end{pmatrix} | y \sim \text{Normal} \left\{ \begin{pmatrix} \mu_w \\ \mu_\beta \end{pmatrix}, \begin{pmatrix} E & F' \\ F & G \end{pmatrix} \right\}.$$

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

$$\underline{w} | \beta_{(0)}, y \sim \text{Normal} \{ \mu_w - D^{-1} C' (\beta_{(0)} - \mu_\beta), D^{-1} \} \text{ and } \beta_{(0)} | y \sim \text{Normal} \{ \mu_\beta, G \},$$

where

$$\mu_w = \underline{w}^* + E \underline{g}_1 + F' \underline{g}_2 \text{ and } \mu_\beta = \beta_{(0)}^* + F \underline{g}_1 + G \underline{g}_2.$$

□

Appendix C. Proof of Theorem 2

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

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

$$\text{where } \underline{\Sigma}_w = \left[\underline{w} - \left(D + \frac{1}{\sigma^2} I \right)^{-1} \left(D \underline{\mu}_w - C'(\underline{\beta}_{(0)} - \underline{\mu}_\beta) + \frac{1}{\sigma^2} \underline{\mu}_v \right) \right].$$

Then it can show that the $\pi_a(\underline{w} | \underline{v}, \underline{\beta}_{(0)}, \sigma^2, \underline{y})$ is

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

Notice that $(D + \frac{1}{\sigma^2} I)$ is diagonal matrix. Then given $\underline{v}, \underline{\beta}_{(0)}, \sigma^2, \underline{y}$, 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 \underline{w} has a multivariate normal distribution, we can integrate out \underline{w} from the joint approximate posterior density $\pi_a(\underline{w}, \underline{v}, \underline{\beta}, \sigma^2, \delta^2 | \underline{y})$, and obtain the joint posterior density of $\underline{v}, \underline{\beta}, \sigma^2$ and δ^2

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

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

Look at the exponent only containing $v_i, i = 1, \dots, \ell$, in the $\pi_a(\underline{v}, \underline{\beta}, \sigma^2, \delta^2 | \underline{y})$

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

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

Then it is easy to see that

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

Similarly, we can use parallel computing to draw v_i , $i = 1, \dots, \ell$ as well since all of them are independent given $\beta_{(0)}, \beta_0, \sigma^2, \delta^2$. Then we can integrate out \underline{v} from the joint approximate posterior density $\pi_a(\underline{v}, \beta, \sigma^2, \delta^2 | \underline{y})$ and obtain the joint posterior density of β, σ^2 and δ^2

$$\begin{aligned} \pi_a(\beta, \sigma^2, \delta^2 | \underline{y}) &\propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^{\ell} \left(\frac{1}{1/\sum_{j=1}^{n_i} \sigma_{ij}^2} + \delta^2 \right)^{-1} \left[\bar{\mu}_{w_i} - \bar{\ell}'_i(\beta_{(0)} - \underline{\mu}_{\beta}) - \beta_0 \right]^2 \right\} \\ &\quad \times \exp \left\{ -\frac{1}{2} \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \frac{1}{\sigma_{ij}^2} \left[\{(\bar{\ell}_i - \underline{\ell}_{ij})'(\beta_{(0)} - \underline{\mu}_{\beta}) - (\bar{\mu}_{w_i} - \underline{\mu}_{w_{ij}})\}^2 \right] \right\} \\ &\quad \times \prod_{i=1}^{\ell} \left(\frac{1}{\sum_{j=1}^{n_i} \sigma_{ij}^2} + \frac{1}{\delta^2} \right)^{\frac{1}{2}} \frac{1}{|\delta^2 I|^{1/2} |D + \frac{1}{\sigma^2} I|^{1/2}} \frac{1}{(1 + \sigma^2)^2} \frac{1}{(1 + \delta^2)^2} \\ &= e^{-\frac{1}{2} (\underline{\mu}_w - D^{-1} C'(\beta_{(0)} - \underline{\mu}_{\beta}) - \beta_0 \underline{j})' (D^{-1} + \sigma^2 I + \delta^2 I)^{-1} (\underline{\mu}_w - D^{-1} C'(\beta_{(0)} - \underline{\mu}_{\beta}) - \beta_0 \underline{j})} \\ &\quad \times e^{-\frac{1}{2} (\beta_{(0)} - \underline{\mu}_{\beta})' G^{-1} (\beta_{(0)} - \underline{\mu}_{\beta})} \\ &\quad \times \prod_{i=1}^{\ell} \left(\frac{1}{\sum_{j=1}^{n_i} \sigma_{ij}^2} + \frac{1}{\delta^2} \right)^{\frac{1}{2}} \frac{1}{|\delta^2 D + \frac{\delta^2}{\sigma^2} I|^{1/2}} \frac{1}{(1 + \sigma^2)^2} \frac{1}{(1 + \delta^2)^2}. \end{aligned}$$

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

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

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

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

So the exponent terms are

$$\begin{pmatrix} \beta_0 - \omega_0 \\ \beta_{(0)} - \omega_{(0)} \end{pmatrix}' \begin{pmatrix} \delta_0^2 & \gamma' \\ \gamma & \Delta_{(0)} \end{pmatrix} \begin{pmatrix} \beta_0 - \omega_0 \\ \beta_{(0)} - \omega_{(0)} \end{pmatrix}.$$

Consider the exponent terms containing $\beta_{(0)}$ and β_0

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

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

$$\begin{aligned}\Delta_{(0)} &= CD^{-1}(D^{-1} + \sigma^2 I + \delta^2 I)^{-1} D^{-1} C' + G^{-1}, \\ \delta_0^2 &= \tilde{j}'(D^{-1} + \sigma^2 I + \delta^2 I)^{-1} \tilde{j}, \\ \gamma &= CD^{-1}(D^{-1} + \sigma^2 I + \delta^2 I)^{-1} \tilde{j}, \\ \begin{pmatrix} \omega_0 \\ \varpi_{(0)} \end{pmatrix} &= \begin{pmatrix} \delta_0^2 & \gamma' \\ \gamma & \Delta_{(0)} \end{pmatrix}^{-1} \begin{pmatrix} (\mu_w + D^{-1} C' \mu_\beta)'(D^{-1} + \sigma^2 I + \delta^2 I)^{-1} \tilde{j} \\ (\mu_w + D^{-1} C' \mu_\beta)'(D^{-1} + \sigma^2 I + \delta^2 I) D^{-1} C' + \mu_\beta' G^{-1} \end{pmatrix}.\end{aligned}$$

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

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

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

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

□

References

- Central Bureau of Statistics Thapathali. *Nepal living Standards Survey 2003/04*; Statistical Report; Central Bureau of Statistics Thapathali: Kathmandu, Nepal, 2004; Volume 1.
- Rao, J.N.; Molina, I. *Small Area Estimation*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
- Nandram, B.; Chen, L.; Fu, S.-T.; Manandhar, B. Bayesian Logistic Regression for Small Areas with Numerous Households. *Stat. Appl.* **2018**, *1*, 171–205.
- Yan, G.; Sedransk, J. Bayesian Diagnostic Techniques for Detecting Hierarchical Structure. *Bayesian Anal.* **2007**, *2*, 735–760.
- Yan, G.; Sedransk, J. A Note on Bayesian Residuals as A Hierarchical Model Diagnostic Technique. In *Stat. Pap.* **2010**, *51*, 1. [CrossRef]
- Fuller, W.A.; Goyeneche, J.J. Estimation of the State Variance Component. Unpublished Manuscript. 1988,
- Torabi, M.; Rao, J.N.K. On Small Area Estimation under a Sub-Area Level Model. In *J. Multivar. Anal.* **2014**, *127*, 36–55. [CrossRef]
- Chen, L.; Nandram, B.; Cruze, N.B. Hierarchical Bayesian Model with Inequality Constraints for US County Estimates. *J. Off. Stat.* **2022**, *38*, 709–732. [CrossRef]
- Erciulescu, A. L.; Cruze, N. B.; Nandram, B. Model-Based County Level Crop Estimates Incorporating Auxiliary Sources of Information. *J. R. Stat. Soc. Ser. (Stat. Soc.)* **2019**, *182*, 283–303. [CrossRef]
- Fay, R.E.; Herriot, R.A. Estimates of Income for Small Places: an Application of James-Stein Procedures to Census Data. *J. Am. Stat. Assoc.* **1979**, *74*, 269–277. [CrossRef]
- Stukel, D.M.; Rao, J.N.K. Estimation of Regression Models with Nested Error Structure and Unequal Error Variances Under Two and Three Stage Cluster Sampling. *Stat. Probab. Lett.* **1997**, *35*, 401–407. [CrossRef]
- Stukel, D.M.; Rao, J.N.K. On Small-Area Estimation under Two-Fold Nested Error Regression Models. *J. Stat. Plan. Inference* **1999**, *78*, 131–147. [CrossRef]
- Nandram, B.; Sedransk, J. Bayesian Predictive Inference for a Finite Population Proportion: Two-Stage Cluster Sampling. *J. R. Stat. Soc. Ser. B (Methodol.)* **1993**, *55*, 399–408. [CrossRef]
- You, Y.; Reiss, P. Hierarchical Bayes Small Area Estimation of Response Rates for an Expenditure Survey. In *Proceedings of the Survey Methods Section*; Statistical Society of Canada: Ottawa, ON, Canada, 2000; pp. 123–128.
- Nandram, B. Bayesian Predictive Inference of a Proportion Under a Twofold Small-Area Model. *J. Off. Stat.* **2016**, *32*, 187–208. [CrossRef]

16. Lee, D.; Nandram, B.; Kim, D. Bayesian Predictive Inference of a Proportion under a Two-Fold Small Area Model with Heterogeneous Correlations. *Surv. Methodol.* **2017**, *17*, 69–92.
17. Chen, L.; Nandram, B. A Hierarchical Bayesian Beta-Binomial Model for Sub-areas. In *Applied Statistical Methods. ISGES 2020, Springer Proceedings in Mathematics & Statistics*; Hanagal, D.D., Latpate, R.V., Chandra, G., Eds.; Springer: Singapore, 2022; pp. 23–40.
18. Nandram, B. Discrimination between Complementary Log-log and Logistic Model for Ordinal Data. *Commun. Stat. Theory Methods* **1989**, *18*, 2155–2164. [[CrossRef](#)]
19. Roberts, G.; Rao, J.N.K.; Kumar, S. Logistic Regression Analysis of Sample Survey Data. *Biometrika* **1987**, *74*, 1–12. [[CrossRef](#)]
20. Nandram, B.; Chen, M.-H. Reparameterizing the Generalized Linear Model to Accelerate Gibbs Sampler Convergence. *J. Stat. Comput. Simul.* **1996**, *54*, 129–144. [[CrossRef](#)]
21. Albert, J.H.; Chib, S. Bayesian Analysis of Binary and Polychotomous Response Data. *J. Am. Stat. Assoc.* **1993**, *88*, 669–679. [[CrossRef](#)]
22. Farrell, P.J.; MacGibbon, B.; Tomberlin, T.J. Empirical Bayes Small Area Estimation Using Logistic Regression Models and Summary Statistics. *J. Bus. Econ. Stat.* **1997**, *15*, 101–108.
23. Nandram, B.; Erhardt, E. Fitting Bayesian Two-Stage Generalized Linear Models Using Random Samples via the SIR Algorithm. *Sankhya* **2010**, *66*, 733–755.
24. Nandram, B.; Choi, J.W. A Bayesian Analysis of Body Mass Index Data from Small Domains Under Nonignorable Nonresponse and Selection. *J. Am. Stat. Assoc.* **2010**, *105*, 120–135. [[CrossRef](#)]
25. Scott, S.L.; Blocker, A.W.; Bonassi, F.V.; Chipman, H.A.; George, E.I.; McCulloch, R.E. *Bayes and Big Data: The Consensus Monte Carlo Algorithm*; Technical Report; Google, Inc.: Mountain View, CA, USA, 2013; pp. 1–22.
26. Miroshnikov, A.; Wei, Z.; Conlon, E.M. Parallel Markov chain Monte Carlo for non-Gaussian posterior distributions. *Stat* **2015**, *4*, 304–319. [[CrossRef](#)]
27. Rue, H.; Martino, S.; Chopin, IN. Approximate Bayesian Inference for Latent Gaussian Models Using Integrated Nested Laplace Approximations. *J. R. Stat. Soc. (Ser. B)* **2009**, *71*, 319–392. [[CrossRef](#)]
28. Fong, Y.; Rue, H.; Wakefield, J. Bayesian inference for generalized linear mixed models. *Biostatistics* **2010**, *11*, 397–412. [[CrossRef](#)] [[PubMed](#)]
29. Illian, J. B.; Sørbye, S. H.; Rue, H. A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). *Ann. Appl. Stat.* **2012**, *6*, 1499–1530. [[CrossRef](#)]
30. Ferkingstad, E.; Rue, H. Improving the INLA approach for approximate Bayesian inference for latent Gaussian models. *Electron. J. Stat.* **2015**, *9*, 2706–2731. [[CrossRef](#)]
31. Cox, D.R.; Snell, E.J. *Analysis of Binary Data*, 2nd ed.; Chapman and Hall/CRC: London, UK, 1989.
32. Wang, J.; Fuller, W.A. The Mean Squared Error of Small Area Predictors Constructed With Estimated Area Variances. *J. Am. Stat. Assoc.* **2003**, *98*, 716–723. [[CrossRef](#)]
33. Yan, G.; Sedransk, J. Small Area Estimation Using Area Level Models and Estimated Sampling Variances. *Surv. Methodol.* **2006**, *32*, 97–103.
34. Erciulescu, A.L.; Berg, E. Small area estimates for the conservation effects assessment project. In *In Frontiers of Hierarchical Modeling in Observational Studies, Complex Surveys and Big Data: A Conference Honoring Professor Malay Ghosh*, College Park, MD, USA; Women in Statistics Conference: Cary, NC, USA, 2014.
35. Boyd, S.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK, 2004.
36. Rubin, D.B. The Bayesian Bootstrap. *Ann. Stat.* **1981**, *9*, 130–134. [[CrossRef](#)]
37. Eddebuettel, D.; Francois, R. Rcpp: Seamless R and C++ Integration. *J. Stat. Softw.* **2011**, *40*, 1–18. [[CrossRef](#)]
38. Eddebuettel, D.; Sanderson, C. Rcpparmadillo: Accelerating R with High-performance C++ Linear Algebra. *Comput. Stat. Data Anal.* **2014**, *71*, 1054–1063. [[CrossRef](#)]
39. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019.
40. Geweke, J. Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. *Bayesian Stat.* **1992**, 169–193.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.