

Bayesian Inference of Non-probability Samples

Zhiqing Xu

Balgobin Nandram*

Abstract

In a world of big data, non-probability samples are fast and easy to collect than the traditional probability samples. The main issue of big data sources is the unknown probability of inclusion, and often the case that probability is related to the response variable. To deal with the nonignorable nonresponses, we propose a double logistic regression model which linked by the response variable, and fit by MCMC methods. We compare the nonignorable model we proposed in a simulation study with a ignorable model – single logistic regression, Horvitz-Thompson and Hajek estimators. We show that our model is favorable in terms of prediction precision.

Key Words: Bayesian model, logistic regression, MCMC, non-probability, selectivity.

1. Introduction

Compared to probability sampling, getting data with non-probability sampling is usually faster and cheaper. Non-probability sampling design includes methods such as quota sampling, whose sampling units are controlled to be distributed close to the population would be; convenience sampling, such as mall intercepts or volunteer samples, where only subjects who are readily available are recruited; and network sampling, where new subjects are recruited from existing study subjects' acquaintances. The biggest challenge come from those non-probability samples is to draw unbiased conclusions that the same kind of probability sampling provides.

Rubin (1976) introduced a formal framework for the field of incomplete data by classifying the missing data mechanisms into three categories: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). Under MCAR, the missingness variable I is independent of outcome Y ; under MAR, I is independent of outcome Y , given the observed y_{obs} ; and under MNAR, missingness I depends on the unobserved outcomes. An example of MNAR is the study that estimating the average income. The lower-earning families can be more difficult to locate and contact, which makes the response indicator variable correlated with the outcome variable - income. Thus, this coverage error will lead to overestimate the population income if not addressed properly.

There are two major categories of estimation procedure to deal with incomplete data: design-based and model-based approach. design-based, or often referred as quasi-randomization, is by estimating the pseudo-inclusion probabilities using the covariates available for both the samples and nonsample units, then incorporate those probabilities in unbiased estimators such as Horvitz-Thompson or Hajek estimators. The other approach, model-based inference, relies on the distribution of the quantity of interest. These methods treat the outcome variable, rather than the inclusion process as being the random variables (Baker et al. 2003). Models for MCAR and MAR are called ignorable, and for MNAR are nonignorable. Heckman selection model (1976) is a model-based approach that factorises the joint distribution of outcome variable and missing data indicator into two sub-models: one specifies the relationship of outcome variable with covariates $P(Y | \gamma, X)$, the other represents the missing data mechanism $P(I | \beta, X, Y)$. The later model depends on Y implies the

*Department of Mathematical Sciences, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609

missing data mechanism is MNAR. In this paper, we propose a full Bayesian modelling approach designed for MNAR which based on selection model formulation.

Using Bayesian methodology for missing data has gained popularity over recent years. The Bayesian paradigm is appealing because missing data can be treated as unknown parameters and are estimated within each MCMC iteration. Nandram and Choi (2002a, 2002b) use a hierarchical Bayesian selection model to model binary nonresponse data. They express uncertainty about ignorability through the odds ratio and apply small area estimation to address the issue of weak identifiability of the parameters. One big problem for nonignorable models is the weakly identifiable parameters. Nandram and Choi (2010) propose a nonignorable nonresponse model to analyse BMI data, which estimates the finite population means of small domains formed by crossing age, race and sex within counties. Woo, Nandram and Kim (2018) perform the Bayesian uncertainty analysis by placing priors on the nonidentifiable parameters, which allows a “borrowing of strength” from larger areas to improve estimation in smaller areas. Wang, Bartlett and Ryan (2017) proposed a Bayesian selection model for correcting for nonignorable nonresponse bias in a logistic modelling. They use sensitivity analysis by treating a parameter as fixed constant and examine results by varying this parameter. In our paper, we use a similar formulation, but approach the issue by full Bayesian modelling strategy.

In section 2, we provide the Bayesian model formulation. In section 3, we compare our model with an ignorable model which models responses and covariates by logistic regression, Horvitz-Thompson and Hajek estimators in a simulation study. Lastly, we conclude with a discussion and possible research extensions in section 4,.

2. Hierarchical Bayesian Model

2.1 Model Formulation

Adapting the formulation from Smith (Sugden and Smith, 1984), we consider the joint density of the population vector of an analysis variable, $\mathbf{Y} = (Y_1, Y_2, \dots, Y_N)$ and the binary inclusion indicator variable $\mathbf{I} = (I_1, I_2, \dots, I_N)$. Suppose we have $p - 1$ covariates, that is known for the population, which will be used for both constructing estimator and modeling selection mechanism. We use \mathbf{X} to denote the $N \times p$ matrix that contains a column of 1s and the $p - 1$ columns of X variables. Without loss of generality, let $y_i, i = 1, \dots, n$, be the sample responses, and $y_i, i = n + 1, \dots, N$ be the nonsample units, accordingly, $I_i = 1$ for $i = 1, \dots, n$ and $I_i = 0, i = n + 1, \dots, N$. The prediction of the population proportion is the quantity of interest.

The model assumes both the sample and nonsample units follow the same population model, which we called the full data model. In our example, the full data model for \mathbf{Y} satisfy the logistic regression model, with $\boldsymbol{\gamma}$ be regression coefficients vector with the length of p .

$$y_i | \boldsymbol{\gamma} \overset{ind}{\sim} \text{Bernoulli} \left\{ \frac{\exp(\mathbf{x}'_i \boldsymbol{\gamma})}{1 + \exp(\mathbf{x}'_i \boldsymbol{\gamma})} \right\}, \quad i = 1, \dots, N \quad (1)$$

The selection mechanism is modeled by another logistic regression model, but included \mathbf{Y} as an additional predictors to reflect the NMAR selection scheme. Denote the vector of parameters associated with \mathbf{X} by $\boldsymbol{\beta}_{(p)}$ and the parameter for \mathbf{Y} by β_{p+1} . When it is not necessary to split into $\boldsymbol{\beta}_{(p)}$ and β_{p+1} , we use $\boldsymbol{\beta}$ to represent the combined vector.

$$I_i | \mathbf{y}, \boldsymbol{\beta} \overset{ind}{\sim} \text{Bernoulli} \left\{ \frac{\exp(\mathbf{x}'_i \boldsymbol{\beta}_{(p)} + \beta_{p+1} y_i)}{1 + \exp(\mathbf{x}'_i \boldsymbol{\beta}_{(p)} + \beta_{p+1} y_i)} \right\}, \quad i = 1, \dots, N \quad (2)$$

By putting these two models together, the joint likelihood of the i^{th} subject is

$$f(y_i, I_i | \boldsymbol{\beta}, \boldsymbol{\gamma}) \propto \left(\frac{\exp(\mathbf{x}'_i \boldsymbol{\gamma})}{1 + \exp(\mathbf{x}'_i \boldsymbol{\gamma})} \right)^{y_i} \left(1 - \frac{\exp(\mathbf{x}'_i \boldsymbol{\gamma})}{1 + \exp(\mathbf{x}'_i \boldsymbol{\gamma})} \right)^{1-y_i} \\ \times \left(\frac{\exp(\mathbf{x}'_i \boldsymbol{\beta}_{(p)} + \beta_{p+1} y_i)}{1 + \exp(\mathbf{x}'_i \boldsymbol{\beta}_{(p)} + \beta_{p+1} y_i)} \right)^{I_i} \left(1 - \frac{\exp(\mathbf{x}'_i \boldsymbol{\beta}_{(p)} + \beta_{p+1} y_i)}{1 + \exp(\mathbf{x}'_i \boldsymbol{\beta}_{(p)} + \beta_{p+1} y_i)} \right)^{1-I_i}. \quad (3)$$

To improve the mixing performance in MCMC, we choose multivariate normal $N(\hat{\boldsymbol{\beta}}, 10^2 \hat{\boldsymbol{\Sigma}})$ to be the priori for $\boldsymbol{\beta}$, where $\hat{\boldsymbol{\beta}}$ is the mode of the likelihood for $\boldsymbol{\beta}$, and $\hat{\boldsymbol{\Sigma}}$ is the inverse of the Hessian of its negative log-likelihood. Notice we multiply the estimated covariance matrix by 10^2 to avoid the issue of double use of data, and make the priori less informative. Using the flat priori for $\boldsymbol{\gamma}$ with $\Pi(\boldsymbol{\gamma}) = \mathbf{1}$, we derive the posterior joint density of the model parameters and the nonsampled units, as presented in formula (4).

Denote the non-sampled units by \mathbf{y}_{ns} , and the sampled ones by \mathbf{y}_s . Taking $I_i = 1$ for $i = 1, \dots, n$, and $I_i = 0$ for $i = n + 1, \dots, N$, the posterior joint density is

$$f(\mathbf{y}_{ns}, \boldsymbol{\beta}, \boldsymbol{\gamma} | \mathbf{I}, \mathbf{y}_s) \propto \prod_{i=1}^n \frac{\exp(\mathbf{x}'_i \boldsymbol{\gamma} y_i)}{1 + \exp(\mathbf{x}'_i \boldsymbol{\gamma})} \frac{\exp(\mathbf{x}'_i \boldsymbol{\beta}_{(p)} + \beta_{p+1} y_i)}{1 + \exp(\mathbf{x}'_i \boldsymbol{\beta}_{(p)} + \beta_{p+1} y_i)} \\ \times \prod_{i=n+1}^N \frac{\exp(\mathbf{x}'_i \boldsymbol{\gamma} y_i)}{1 + \exp(\mathbf{x}'_i \boldsymbol{\gamma})} \frac{1}{1 + \exp(\mathbf{x}'_i \boldsymbol{\beta}_{(p)} + \beta_{p+1} y_i)} \quad (4) \\ \times \exp \left(-\frac{1}{2} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T (10^2 \hat{\boldsymbol{\Sigma}})^{-1} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \right).$$

Before drawing samples from this posterior distribution, we have to check its propriety first. Since the product on second row of (4) is always less than one and we know the normal distribution is proper, we only need to show the first row

$$\prod_{i=1}^n \frac{\exp(\mathbf{x}'_i \boldsymbol{\gamma} y_i)}{1 + \exp(\mathbf{x}'_i \boldsymbol{\gamma})} \times \prod_{i=1}^n \frac{\exp(\mathbf{x}'_i \boldsymbol{\beta}_{(p)} + \beta_{p+1} y_i)}{1 + \exp(\mathbf{x}'_i \boldsymbol{\beta}_{(p)} + \beta_{p+1} y_i)} \quad (5)$$

is proper. We see that $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are independent in (5), and their individual densities are log-concave (see detailed proof on Nandram et al., 2018), thus the joint density must be log-concave. Due to the property of log-concavity, we know (5) is proper, which proves (4) is proper.

Now, our goal is to apply Markov chain Monte Carlo (MCMC) methods to draw samples of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $y_i, i = n + 1, \dots, N$ from posterior distribution (4) and combine each set of the prediction units with the sample units to obtain the inferences of population proportion. In the next part, we will describe the samplers and algorithm applied.

2.2 MCMC algorithm

It is easy to derive the conditional densities of vector $\boldsymbol{\beta}, \boldsymbol{\gamma}$. The derivation of (iii) is less straightforward, we leave the proof at Appendix. The conditional posterior densities are

$$(i) f(\boldsymbol{\beta}, | \mathbf{y}, \mathbf{I}) \propto \prod_{i=1}^N \frac{\exp((\mathbf{x}'_i \boldsymbol{\beta}_{(p)} + \beta_{p+1} y_i) I_i)}{1 + \exp(\mathbf{x}'_i \boldsymbol{\beta}_{(p)} + \beta_{p+1} y_i)} \times \exp \left(-\frac{1}{2} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T (10^2 \hat{\boldsymbol{\Sigma}})^{-1} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \right);$$

$$(ii) f(\gamma | \mathbf{y}, \mathbf{I}) \propto \prod_{i=1}^N \frac{\exp(\mathbf{x}'_i \gamma y_i)}{1 + \exp(\mathbf{x}'_i \gamma)}$$

$$(iii) P(y_i = 1 | \gamma, \beta) = \frac{\exp(\mathbf{x}'_i \gamma)(1 + \exp(\mathbf{x}'_i \beta_{(p)}))}{\exp(\mathbf{x}'_i \gamma)(1 + \exp(\mathbf{x}'_i \beta_{(p)})) + 1 + \exp(\mathbf{x}'_i \beta_{(p)} + \beta_{p+1})},$$

$$P(y_i = 0 | \gamma, \beta) = 1 - P(y_i = 1 | \gamma, \beta), \quad i = n + 1, \dots, N.$$

Because the density of β or γ is not in a simple form, we use Metropolis-Hastings algorithm for these two sets of parameters. Whereas the conditional distribution of $y_i, i = n + 1, \dots, N$ has close form, which is Bernoulli distribution with probability of success calculated by (iii), it is nature to use Gibbs sampler with univariate Metropolis steps. It can be shown that each of the components preserves the stationary distribution of the chain, and the convergence will obtain in the combination steps (Carlin and Louis, 2009).

Algorithm: For $(t = 1, \dots, T)$, repeat:

1. Draw β^* from multivariate t-distribution, where the proposal mean vector is the mode of (i), and covariance matrix is given by its inverse of the Hessian of the negative log-likelihood function (-LL). Keep the degrees of freedom ν fixed through out the iterations. It is standard to tune the Metropolis step by varying ν . Denote this proposal density by $g(\beta, | \mathbf{y}, \mathbf{I})$

2. Compute the acceptance ratio

$$r = \frac{f(\beta^*, | \mathbf{y}, \mathbf{I})g(\beta^{(t-1)}, | \mathbf{y}, \mathbf{I})}{g(\beta^*, | \mathbf{y}, \mathbf{I})f(\beta^{(t-1)}, | \mathbf{y}, \mathbf{I})}$$

3. Draw u from uniform $U(0, 1)$, if $u \leq \min(1, r)$, set $\beta^{(t)} = \beta^*$
else, set $\beta^{(t)} = \beta^{(t-1)}$.

Keep a record of the jumping number.

4. Similarly, adapt steps 1-3 to draw $\gamma^{(t)}$ based on formula (ii)
5. For each $i, i = n + 1, \dots, N$, calculate the probability of success by formula (iii), and draw the corresponding y from its Bernoulli distribution.

After iterations are done, we need to check if the jumping rates for β or γ are within the range of (25%, 75%). If the jumping rate is not acceptable, i.e. too low or too high, we have to change the value for ν and repeat the above steps.

3. Simulation and Results

3.1 Dataset Description

We use simulated dataset to test our model. First, we generate values for the covariates *Age*, *Race*, *Gender*, *Education* from distributions listed in Table 1: *Age* is from normal distribution with mean 50 and standard deviation 2, *Race*, *Gender*, and *Education* are from Bernoulli with probabilities of 0.45, 0.3, and 0.4 respectively. We draw $N = 10000$ from each.

Second, we formulate a relation between the covariates variable \mathbf{X} and the analysis variable \mathbf{Y} :

$$v_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{3i} + b_4 x_{4i} + \epsilon_i$$

Table 1: Distributions for covariates

Covariate	Distribution
<i>Age</i>	$N(50, 2^2)$
<i>Race</i>	Bernoulli(0.45)
<i>Gender</i>	Bernoulli(0.3)
<i>Education</i>	Bernoulli(0.4)

where x_2, x_3, x_4 are samples from the distributions in Table 1, and x_1 are the values of *Age* scaled to mean = 0, sd = 1. In our simulation dataset, we use $b_0 = 50, b_1 = 2, b_2 = 4, b_3 = 3, b_4 = 8$, and $\epsilon_i \sim N(0, 2^2)$. Then, to get the binary outcome variable Y , we use a threshold with the value of 56, which is about the median of the population. Y is defined by

$$y_i = \begin{cases} 1 & v_i > 56 \\ 0 & v_i \leq 56. \end{cases}$$

Lastly, to mimic a pattern of nonignorable nonresponse, we resample by the rule

$$\begin{cases} P(I_i = 1 | y_i = 1) = 0.13 \\ P(I_i = 1 | y_i = 0) = 0.07. \end{cases}$$

Use the strategy above, we obtain $n = 1009$ samples, out of $N = 10000$ in the population. The sample proportion is 0.615, while the true population proportion is 0.470. We see that the sample proportion overestimates the population proportion by 31%.

3.2 Results and Conclusion

By running 50,000 iterations, burn-in the first 5000, and thinning by 45, we obtain 1,000 sets of MCMC samples. The jumping rate for γ is 0.68, for β is 0.60, which are acceptable as these are between 0.25 and 0.75. The parameter estimates and HPD intervals are shown in Table 2. Our program is written in R language, and it takes about 5 hours to finish 50,000 iterations. How to improve the program efficiency could be in our extension study.

Table 2: The 95% HPD interval for parameters γ, β and \bar{y}

γ	Mean	HPD interval	β	Mean	HPD interval	\bar{y}	HPD interval
γ_0	-6.03	(-6.88, -5.13)	β_0	-2.49	(-2.59, -2.40)	0.479	(0.456, 0.507)
γ_1	2.02	(1.71, 2.36)	β_1	0.03	(-0.04, 0.10)		
γ_2	4.28	(3.50, 4.99)	β_2	-0.00	(-0.14, 0.14)		
γ_3	2.91	(2.26, 3.54)	β_3	-0.16	(-0.30, -0.02)		
γ_4	8.00	(7.08, 9.05)	β_4	0.19	(-0.11, 0.48)		
			β_5	0.49	(0.10, 0.86)		

It is worth noting that the HPD interval for β_5 does not contains 0, meaning y is significant in the selection model and thus correctly detected the MNAR mechanism. The posterior mean for the population proportion of Y is 0.479, with HPD interval of (0.456, 0.507). We see that this is close to the true population proportion, showing that the model successfully adjusts the sample bias and predicts the population proportion. The histogram of the population proportion predictions is in Figure 1.

Furthermore, we conduct Geweke’s Diagnostics to check the convergence of each parameter to its target distribution. From Table 3, We see that all the p-values from the Z tests are greater than 0.05, and the effective sample size for each parameter is acceptable.

Table 3: Geweke’s Diagnostics for γ and β

γ	p-value	effective size	α	p-value	effective size
γ_0	0.74	1098	β_0	0.14	866
γ_1	0.73	1130	β_1	0.99	642
γ_2	0.24	805	β_2	0.75	660
γ_3	0.35	1430	β_3	0.90	772
γ_4	0.90	1000	β_4	0.74	533
			β_5	0.66	484

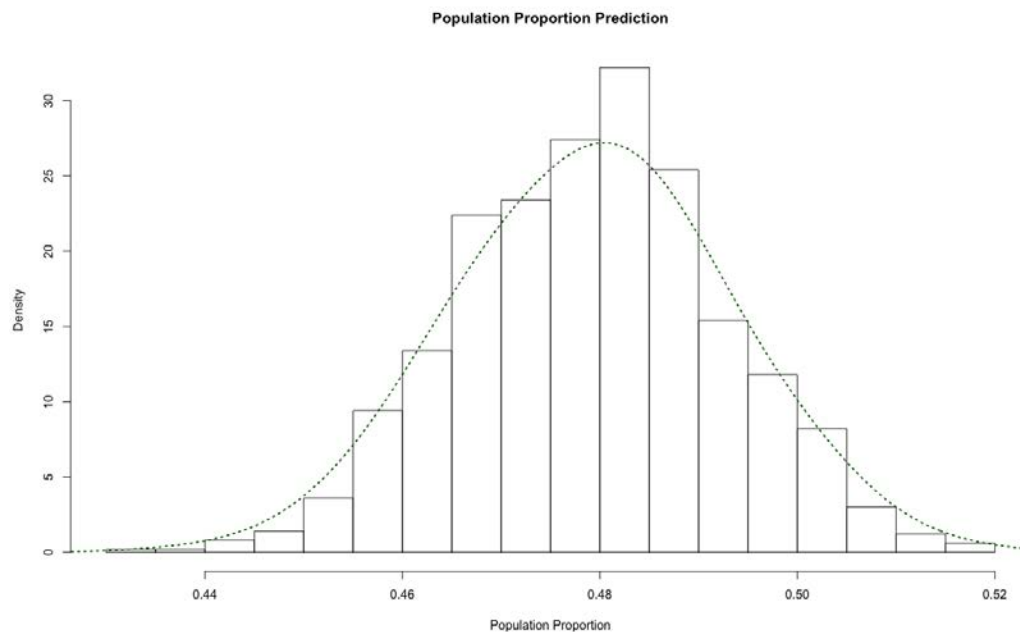


Figure 1: Histogram of Population Proportion Prediction

3.3 Model Comparison

3.3.1 Logistic Regression Model

If the missingness scheme is not MNAR, we would use a single logistic regression as ignorable selection model for binary outcome data. Here, we use it as our baseline model to compare results. By fitting the model with the samples, and applying the model to the nonsample covariates, we get the estimation of population proportion as 0.501.

3.3.2 Horvitz-Thompson Estimator

Horvitz-Thompson (H-T) estimator (1952) has long been used as an unbiased estimator when the weight of samples are correctly specified. The expression to calculate population

Table 4: Results comparison

	True Value	Our Model (95% HPD)	Ignorable model	H-T	Hajek
Sampled	0.615				
Population	0.470	0.479 (0.456, 0.507)	0.501	0.5216	0.5220

mean is

$$\hat{Y}_{HT} = \sum_{i=1}^n \frac{y_i}{Pr_i},$$

where Pr_i is the inclusion probability, and it can be obtained in a model with covariates. Here, we use propensity scores to construct the probability of selection.

Using logit link function, propensity score can be calculated by:

$$\pi_i = P(R_i = 1|x_i; \alpha) = \frac{e^{x_i' \alpha}}{1 + e^{x_i' \alpha}}, \quad (6)$$

for $i = 1, 2, \dots, N$. Thus, with the selection probability is defined by $Pr_i = \frac{n\pi_i}{\sum_{i=1}^N \pi_i}$.

For the simulated data, we obtain the value for H-T estimator is 0.5216.

3.3.3 Hajek Estimator

Hajek estimator of the population mean, defined by Sarndal, Swensson and Wretman (2003) is

$$\hat{Y}_H = \frac{\sum_{i=1}^n y_i / Pr_i}{\sum_{i=1}^n 1 / Pr_i},$$

where Pr_i is the selection probability. We use the same formula for selection probability as for Hovitz-Thompson estimator.

The result is 0.5220, with the comparison among those models listed in Table 4.

4. Conclusion and Possible Research Extension

It is usually difficult to learn about the missing data mechanism. However, incorporating the outcome variable into the selection model and testing its statistical significance give us a sense about the selection mechanism. Moreover, our model can automatically adjust to different missing data mechanisms with varied significance of parameters, providing the model and covariates are correctly specified.

There are a few interesting research extensions. First, instead of using nonprobability samples alone, DiSogra et al. (2012) introduced an idea referred to as “blended calibration” in which a small probability sample is combined with a parallel nonprobability sample. Many researches have been conducted on how to combine these two to increase the accuracy of estimates derived from nonprobability samples alone (Sakshaug et al., 2019). By using our modeling strategy, probability samples should be straightforward to be entered into likelihood function.

Second, we could further extend our research to the case where the response is categorical data, so multinomial distribution would be used to model as such.

Third, there are likely to be heterogeneity in big data, which can be addressed by post-stratifying the population into small areas. Thus, poststratification and small area estimation could be used to borrow strength from the neighboring cells.

Furthermore, nearest neighbor approach with variability incorporated can be studied.

A. Appendix

Conditional Distribution for y_i

We know that Y is binary variable, which means it is from Bernoulli distribution and characterized by the success probability $P(y_i = 1 | \dots), i = n + 1, \dots, N$. From (3), we see that

$$P(y_i = 1 | I_i = 0, \mathbf{x}_i, \boldsymbol{\gamma}, \boldsymbol{\beta}, \beta_{p+1}) = k \frac{\exp(\mathbf{x}'_i \boldsymbol{\gamma})}{1 + \exp(\mathbf{x}'_i \boldsymbol{\gamma})} \frac{1}{1 + \exp(\mathbf{x}'_i \boldsymbol{\beta} + \beta_{p+1})} \quad (7)$$

where k is a constant of which the value is the product of the rest of fractions in the joint distribution. Similarly, by plugging in $y_i = 0$, we have

$$P(y_i = 0 | I_i = 0, \mathbf{x}_i, \boldsymbol{\gamma}, \boldsymbol{\beta}, \beta_{p+1}) = k \frac{1}{1 + \exp(\mathbf{x}'_i \boldsymbol{\gamma})} \frac{1}{1 + \exp(\mathbf{x}'_i \boldsymbol{\beta})} \quad (8)$$

where k is the product of the same fractions, and thus has the same value as in (4).

By using the property that $P(y_i = 1 | \dots) + P(y_i = 0 | \dots) = 1$ we can solve for k ,

$$k = \frac{(1 + \exp(\mathbf{x}'_i \boldsymbol{\gamma}))(1 + \exp(\mathbf{x}'_i \boldsymbol{\beta} + \beta_{p+1}))(1 + \exp(\mathbf{x}'_i \boldsymbol{\beta}))}{\exp(\mathbf{x}'_i \boldsymbol{\gamma})(1 + \exp(\mathbf{x}'_i \boldsymbol{\beta})) + 1 + \exp(\mathbf{x}'_i \boldsymbol{\beta} + \beta_{p+1})} \quad (9)$$

Putting (7) and (9) together, we arrived at

$$\begin{aligned} P(y_i = 1 | \mathbf{x}_i, \boldsymbol{\gamma}, \boldsymbol{\beta}, \beta_{p+1}) &= \frac{(1 + \exp(\mathbf{x}'_i \boldsymbol{\gamma}))(1 + \exp(\mathbf{x}'_i \boldsymbol{\beta} + \beta_{p+1}))(1 + \exp(\mathbf{x}'_i \boldsymbol{\beta}))}{\exp(\mathbf{x}'_i \boldsymbol{\gamma})(1 + \exp(\mathbf{x}'_i \boldsymbol{\beta})) + 1 + \exp(\mathbf{x}'_i \boldsymbol{\beta} + \beta_{p+1})} \\ &\times \frac{\exp(\mathbf{x}'_i \boldsymbol{\gamma})}{1 + \exp(\mathbf{x}'_i \boldsymbol{\gamma})} \frac{1}{1 + \exp(\mathbf{x}'_i \boldsymbol{\beta} + \beta_{p+1})} \\ &= \frac{\exp(\mathbf{x}'_i \boldsymbol{\gamma})(1 + \exp(\mathbf{x}'_i \boldsymbol{\beta}))}{\exp(\mathbf{x}'_i \boldsymbol{\gamma})(1 + \exp(\mathbf{x}'_i \boldsymbol{\beta})) + 1 + \exp(\mathbf{x}'_i \boldsymbol{\beta} + \beta_{p+1})} \end{aligned}$$

REFERENCES

- Baker, R., Brick, J. M., Bates, N. A., Battaglia, M., Couper, M. P., Dever, J. A., Gile, K. J., Tourangeau, R. (2013), "Summary Report of the AAPOR Task force on Non-probability Sampling," *Journal of Survey Statistics and Methodology*, 1, 90–143.
- Carlin, B. P., Louis, T. A. (2009), *Bayesian Methods for Data Analysis* (3rd ed.), Boca Raton, FL: CRC Press
- Choi, S. M., and Kim, D. H. (2014), "Sensitivity analysis in Bayesian nonignorable selection model for binary responses," *Journal of the Korean Data & Information Science Society 2014*, 25, 187–194.
- DiSogra, C., Cobb, C., Chan, E., and Dennis J. (2012) "Using Probability-Based Online Samples to Calibrate Non-Probability Opt-in Samples." *Presentation at: 67th Annual Conference of the American Association for Public Opinion Research (AAPOR)*.
- Heckman, J. J. (1976), "The common structure of statistical models of truncation, sample selection, and limited dependent variables and a simple estimator for such models," *Annals of Economic and Social Measurement*, 5, 475–492.
- Horvitz, D., and Thompson, D. (1952), "A Generalization of Sampling Without Replacement from a Finite Universe," *Journal of the American Statistical Association*, 47, 663–685
- Nandram, B., Chen, L., Fu, S., and Manandhar, B. (2018), "Bayesian Logistic Regression for Small Areas with Numerous Households," *Statistics and Application*, 16, 171–205
- Nandram, B., and Choi, J. W. (2002a), "A Bayesian analysis of a proportion under non-ignorable non-response" *Statistics in Medicine*, 21, 1189–1212
- Nandram, B., and Choi, J. W. (2002b), "Hierarchical Bayesian Nonresponse Models for Binary Data From Small Areas with Uncertainty about Ignorability" *Journal of the American Statistical Association*, 97(458), 381–388
- 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
- Rubin, D. B. (1976), "Inference and missing data," *Biometrika*, 63, 581–592
- Sakshaug, J. W., Wisniowski, A., Ruiz, D. A. P., and Blom, A. G. (2019), "Supplementing Small Probability Samples with Nonprobability Samples: A Bayesian Approach," *Journal of Official Statistics*, 35(3), 653–681
- Sarndal, C. E., Swensson, B. and Wretman, J. (2003), *Model assisted survey sampling*, NY: Springer
- Sugden, R. A., and Smith, T. M. F. (1984), "Ignorable and informative designs in survey sampling inference," *Biometrika*, 71(3), 495–506
- Wang, J.J.J., Bartlett, M. and Ryan, L. (2017), "Non-ignorable missingness in logistic regression," *Statistics in Medicine*, 36, 3005–3021
- Woo, N., Nandram, B., and Kim, D. (2018), "Bayesian Small Area Models for Three-way Contingency Tables with Nonignorability," *Statistica Sinica*, 28, 1839–1866.