International Journal of Statistics and Probability; Vol. 12, No. 1; January 2023
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Bayesian Predictive Inference Under Nine Methods for Incorporating Survey Weights

Lingli Yang[1], Balgobin Nandram[1], Jai Won Choi[2]

[1] Department of Mathematical Sciences, Worcester Polytechnic Institute, Worcester, MA, USA

[2] Statistical Consultant, Meho Inc., MD, USA

Correspondence: Balgobin Nandram, Department of Mathematical Sciences, Worcester Polytechnic Institute, Worcester, MA, USA. E-mail: balnan@wpi.edu

Received: November 10, 2022 Accepted: December 20, 2022 Online Published: December 31, 2022

doi:10.5539/ijsp.v12n1p33 URL: https://doi.org/10.5539/ijsp.v12n1p33

Abstract

Sample surveys play a significant role in obtaining reliable estimators of finite population quantities, and survey weights are used to deal with selection bias and non-response bias. The main contribution of this research is to compare the performance of nine methods with differently constructed survey weights, and we can use these methods for non-probability sampling after weights are estimated (e.g. quasi-randomization). The original survey weights are calibrated to the population size. In particular, the base model does not include survey weights or design weights. We use original survey weights, adjusted survey weights, trimmed survey weights, and adjusted trimmed survey weights into pseudo-likelihood function to build unnormalized or normalized posterior distributions. In this research, we focus on binary data, which occur in many different situations. A simulation study is performed and we analyze the simulated data using average posterior mean, average posterior standard deviation, average relative bias, average posterior root mean squared error, and the coverage rate of 95% credible intervals. We also performed an application on body mass index to further understand these nine methods. The results show that methods with trimmed weights are preferred than methods with untrimmed weights, and methods with adjusted weights have higher variability than methods with unadjusted weights.

Keywords: non-probability samples, normalized density, outliers, selection bias, trimmed survey weights

1. Introduction

Sample surveys play a significant role in obtaining reliable estimators of finite population quantities, such as means, totals, and ratios. In an ideal situation, the sampled population, where the sample was taken, can be identical to the target population. Therefore, probability sampling is the golden rule for finite population prediction and inference (Lohr, 2009). However, the ideal survey is hard to attain. The preference of probability sampling is challenged because of non-response, time, and cost. Survey response rates are declining steadily; rare events, such as crashes and diseases, need long-term observation; convenience samples are faster, easier, and cheaper to collect; massive data are increasingly available but unstructured and hard to analyze because there are no survey weights (Beaumont, 2020; Rao, 2020; Y. Chen, Li, & Wu, 2020).

The issue of survey weights also come up in non-probability sampling, which is defined as a sampling mechanism in which the researchers select samples based on the subjective judgment rather than random selection. In general, it is a sampling method in which not all units of the population have an equal chance of being selected, unlike probability sampling where every unit has a non-zero probability to be chosen. Accordingly, survey weights correct data disproportionality for the sample with respect to the target population of interest. Rao, Hidiroglou, Yung, and Kovacevic (2010) and Haziza and Beaumont (2017) summarized the typical weighting process in the multipurpose surveys. In the presence of a relevant probability sample and a set of common auxiliary variables, it is possible to use propensity scores and consider the non-probability samples as regular probability samples, which is also known as quasi-randomization (Lee & Valliant, 2009, Elliott & Valliant, 2017).

Survey weights depend on the survey designs as well as on the actual collected data. In that way, it can be constructed based on a combination of the unit non-response adjustment, the post-stratification adjustment and the inverse of its inclusion probability. Lohr (2007) mentioned good properties of using survey weights in estimation. For example, estimators have smaller mean squared errors and are robust to misspecifications of the superpopulation models. Q. Chen et al. (2017) showed that when sample units are chosen based on a complex design, the main reason to use survey weights in the analysis is to mitigate the biased inferences from a simple random sample, and they emphasized modifications of the basic design-based weight for a better analysis. In the matter of accessing quantities more complicated than means and quantiles, weights can be incorporated into inference as factors in the log-likelihood for each unit, which can lead to weighted linear or logistic regression (Pfeffermann, 1993). Gelman (2007) discussed incorporating covariates into the weighted regression model to make consistent estimates.

Under a specified distribution on the assumed model, it is easy to implement Bayesian inference, assuming the model holds for the sample. Royall and Pfeffermann (1982) discussed Bayesian inference on the finite population with normality assumption and flat priors on the parameters of a linear regression model. Musal, Soyer, McCabe, and Kharroubi (2012) present a Bayesian framework for population utility estimation. Pfeffermann, Da Silva Moura, and Do Nascimento Silva (2006) discussed an application of Bayesian modeling to make inferences from multilevel models under informative sampling. In addition, another way is fitting models on the non-probability sample and then making predictions on the response variable for units in the probability sample (Kim, Park, Chen, & Wu, 2018, Wang, Rothschild, Goel, & Gelman, 2015). Nandram (2007) discussed Bayesian prediction inference under informative sampling via surrogate samples. Beaumont (2020) and Rao (2020) reviewed available methods to use data from a nonprobability sample; specifically they discussed the literature on combining information from probability sample and nonprobability sample using survey weights or design weights.

In Table 1, we summarize the nine different methods by incorporating survey weights into our Bayesian models. Specifically, the baseline model with no survey weights is included; then the remaining eight methods are broadly categorized at original weights or trimmed weights, and each of these is further categorized by unnormalized or normalized distribution, and unadjusted or adjusted weights.

Table 1. A summary of the nine methods

(i).	**No weights**
	A. Baseline model
(ii).	**Original weights**
	B. Unnormalized distribution
	C. Normalized distribution
	D. Adjusted weights, unnormalized distribution
	E. Adjusted weights, normalized distribution
(iii).	**Trimmed weights**
	F. Unnormalized distribution
	G. Normalized distribution
	H. Adjusted weights, unnormalized distribution
	I. Adjusted weights, normalized distribution

Predictive inference can be obtained using a surrogate sample (Nandram 2007). That is, predict the entire population after the samples are obtained from the posterior density of the super-population parameters. For example,

$$y_1, \ldots, y_N | \underset{\sim}{\theta} \overset{ind}{\sim} f(y \mid \underset{\sim}{\theta})$$

is the population model with N the population size. We have sample data $(w_i, y_i), i = 1, \ldots, n$. With a prior on $\underset{\sim}{\theta}$, we have the posterior density $\pi(\theta | \underset{\sim}{y_s})$. Interest is on the finite population mean, $\bar{Y}$, say. Then, Bayesian predictive inference about the finite population mean is obtained as follows,

$$\pi(\bar{Y} | \underset{\sim}{y_s}) = \int f(\bar{Y} | \theta) \pi(\underset{\sim}{\theta} | \underset{\sim}{y_s}) d\underset{\sim}{\theta},$$

where $f(\bar{Y}|\theta)$, the population model, does not depend on $\underset{\sim}{y_s}$; see Nandram and Rao (2021). We simply draw $\underset{\sim}{\theta}$ from the posterior density $\pi(\underset{\sim}{\theta}|\underset{\sim}{y_s})$ and draw $\bar{Y}$ from the population model; this is surrogate sampling (e.g., Nandram, 2007).

The main idea of this research is to compare the performance of nine methods where we consider the base model with no survey weights, and incorporate four different types of survey weights into our Bayesian models. In Section 2, we review preparatory materials on the survey weights and then demonstrate these nine methods in detail: The base model, original survey weights, adjusted survey weights, trimmed survey weights, and adjusted trimmed survey weights incorporated into unnormalized or normalized posterior distributions. In Section 3, a simulation study is performed to gain a further understanding of these nine methods, and the simulated data are analyzed using average posterior mean, average posterior standard deviation, average relative bias, average posterior root mean squared error and the coverage rate of

95% credible intervals. In Section 4, we discuss an application of body mass index (BMI) data, which is taken from the Third National Health and Nutrition Examination Survey (NHANES III), and analyze the posterior means, and posterior standard deviations of these nine methods. Finally, in Section 5, we review the strengths and weaknesses of our study in more detail and provide some future research suggestions. The appendices have technical details.

2. Different Types of Survey Weights

In Section 2.1, we describe survey weights in greater detail. In Section 2.2, we describe adjusted weights and trimmed weights. In Section 2.3, we describe normalized weights.

2.1 Survey Weights

Design weights are reciprocal of selection probabilities, coming straight from the survey. Under a simple random sample, every sample of size, n, from a finite population of size, N, has the same probability of selection, and so each unit has the same selection probability, $\frac{n}{N}$ for sampling without replacement. In complex survey designs, the selection probabilities are different per unit, but units in certain groups may have equal probabilities (e.g., stratification and clustering). The sum of the design weights is the finite population size, N. So although we do not use covariates in this paper, we are actually using calibration weights (i.e., the sum of the original weights is the population size).

In a non-probability sample, the participation variable will be correlated with the study variable, and this correlation is a measure of the data defect of the non-probability sample; see Meng (2018) for a detailed discussion of this measure. One thing to note here as pointed out by Meng (2018), is that even a small correlation can cause a large selection bias for Big Data. Therefore, it is much needed to get a handle on the selection probabilities in the non-probability sample. This correlation is zero in a simple random sample, and there is no data defect. The correlation is non-zero in a complex survey (any survey with unequal selection probabilities or design weights). One prominent example is probability-proportional-to-size (PPS) sampling. The good thing though is that we know the selection probabilities in the probability sample, so there is no data defect if the selection probabilities are used properly in the analysis. In PPS sampling, the probability of selection is proportional to a measure of size, and this measure is correlated with the study variable (some PPS samples can have equal selection probabilities). So in this design, the study variable and the participation variable could be correlated. The good thing is that we have the selection probabilities, which should be used to avoid selection bias.

Probability samples have selection probabilities as part of the data. If the selection probabilities are not taken into account in an analysis, it is essentially a non-probability sample and the data defect will be nonzero, thereby creating a selection bias. The good thing is that when a probability sample is taken, the probabilities must be incorporated into the analysis, and an estimator (e.g., Horvitz-Thompson) of a finite population quantity (e.g., mean and total) is designed unbiased. In addition, there are usually problems with the sample selection. For example, if there is a non-response, and responding units are different from non-responding units, there will be a selection bias, and the selection probabilities need to be adjusted to form the survey weights; see below. A good probability sample can provide design-unbiased estimates, but it can turn into a non-probability sample if the survey weights are not used because there will be data defects.

Here our primary concern is about survey weights, which have several adjustments (e.g., non-response, demographics, and other features of the population), but the survey weights must be calibrated to the population size (i.e., they must be adjusted so that they sum up to the population size). However, the reciprocals of the survey weights we still call selection probabilities (or more appropriately propensity scores). One can think that survey weights are surrogates for the design weights, but they are not designed weights. We can also think that the survey weight of a unit is the number of units that it represents in the population, including itself. However, survey weights are not in general equal to inverse probabilities of selection but they are typically constructed using a combination of probability calculations and non-response adjustments (possibly multi-stage processes). We agree with Gelman (2007) that survey weights are not just features of individual units, but the survey weight of an individual unit depends on other units in the entire population. This suggests that survey weights are correlated, and they should have a joint distribution. This is not a simple task because it is not possible to find the correlation in a single sample without additional structure. One way to get an idea of what the correlation might be is to use the method of random grouping; see, for example, Choi and Nandram (2021, 2022). We do not pursue this issue further in this paper.

We use an example from Q. Chen et al. (2017) to illustrate some of these points. In data sets from population surveys, the weights attached to the units can include adjustments for unit non-response, and post-stratification to match the distributions of auxiliary variables with known distributions in the population. Thus, a more general form of weight for the i^{th} sampled unit with design weight, d_i, can be expressed as

$$w_i = d_i \times w_i^{(n)} \times w_i^{(p)},$$

where $w_i^{(n)}$ is a unit non-response adjustment, and $w_i^{(p)}$ is a post-stratification adjustment. Weighting units are a convenient way to help correct the effects of differential inclusion into the sample, but the resulting estimates can be very inefficient. Q. Chen et al. (2017) reviews methods that attempt to help mitigate this inefficiency, either by modifying the weights or by model-based approaches that treat weights as covariates. As stated by Q. Chen et al. (2017), many surveys are multi-purpose surveys because the information is collected on a possibly large number of characteristics of interest. Then, the weights are generally constructed so that they may be applied to any characteristic of interest, and these weights are often referred to as multi-purpose weights. It does not matter what weights we have, in principle, we can use a weighted likelihood to make inferences about a finite population. Smoother weights are preferred (e.g., trimmed weights).

We believe that the dialogue between Gelman (2007) and Lohr (2007) is historical and of scientific value, so it behooves us to cite it here. Gelman (2007) started his paper with the statement, "Survey weighting is a mess. It is not always clear how to use weights in estimating anything more complicated than a simple mean or ratio, and standard errors are tricky even with simple weighted means." Lohr (2007) responded in her comments, "I do not think that survey weighting is a mess, but I do think that many people ask too much of the weights." We concur with Lohr (2007) that one cannot expect too much from survey weights, but we should try to get as much as possible from them. Design weights are generally adjusted several times. However, we agree with Gelman (2007) that survey weights are complicated constructs that a survey sampler must confront.

In design-based analyzes, inverse probability weighting is used to estimate finite population means or totals; see Y. Chen et al. (2020) and Robbins, Ghosh-Dastidar, and Ramchand (2021). They both used inverse probability weighting via quasi-randomization, but their procedures are different. While Y. Chen et al. (2020) used weighted likelihoods to obtain propensity scores, Robbins et al. (2021) focused on weights using design-based methods such as calibration and propensity score weighting (without parametric models). They both used the Horvitz-Thompson estimator to make inferences about finite population means or totals. Note that these weights are surrogates for survey weights, so quasi-randomization is the probabilistic basis for inference. Our original weights are calibrated to the population size. However, our concern is not how to estimate the weights, but rather to show how to incorporate the weights into a parametric model. The idea is the same for any parametric model, simple (Appendix A) or hierarchical (Appendix C).

There are Bayesian parametric models, and we present one piece of evidence here. A model-based method for incorporating survey weights was described by Nandram and Rao (2021) for integration of a non-probability sample and a probability sample; see also Nandram, Choi, and Liu (2021). They used adjusted weights with a normalized likelihood, which is the likelihood divided by its integral (sum) for the continuous (discrete) response. Under normality, the normalization constant does not matter for inference, but it could matter under other densities (e.g., Bernoulli distribution); the Bernoulli is a running example, but see Appendix A. We compare nine methods for incorporating the survey weights from a single probability sample (not data integration) similar to the one presented by these authors. Again, note that we are constructing survey weights; they are available to us. The procedure for a non-probability sample, once the unknown selection probabilities are estimated, is simply the same. Of course, a problem with parametric models is that they are generally not robust to their assumptions, but if their assumptions are true, they will be more efficient than design-based methods; design-based methods rely on consistency and unbiasedness possibly via asymptotic theory to obtain linear approximations.

2.2 Adjusted Weights and Trimmed Weights

For unit i, let y_i denote the response (study variable), x_i vector of covariates, and W_i the survey weight. Also, let z_i be an indicator (participation) variable, where $z_i = 1$ if the unit i is in the sample and $z_i = 0$ if the unit i is not in the sample, and π_i the selection probability. Under the assumption that the selection probability only depends on observed covariates (i.e., ignorable selection),

$$\pi_i = P(z_i = 1 \mid \underset{\sim}{x_i}, y_i) = P(z_i = 1 \mid \underset{\sim}{x_i}), \quad i = 1, \ldots, N.$$

When a sampling plan is implemented in a finite population of size N to draw a sample of size n, with given selection probabilities, $\pi_1, \ldots, \pi_N$, with π_i corresponding to the i^{th} selected unit, since $d_i = 1/\pi_i$, $i = 1, \ldots, N$, Horvitz-Thompson estimators of population total and population size are

$$\hat{T} = \sum_{i=1}^{n} d_i y_i,$$

$$\hat{N} = \sum_{i=1}^{n} d_i,$$

and these are design-unbiased estimators. With survey weights, these estimators are approximately design-unbiased and are generally used in a similar manner.

The effective sample size indicates the degree to which the variance increases due to the unequal weight. Then, the adjusted weight (Potthoff, Woodbury, & Manton, 1992) is required to get the appropriate variance,

$$w_i = \frac{n_e W_i}{\sum_{j=1}^n W_j}, \quad i = 1, \ldots, n,, \tag{1}$$

$$n_e = \frac{\left(\sum_{j=1}^n W_j\right)^2}{\sum_{j=1}^n W_j^2}, \tag{2}$$

where n_e is the effective sample size; see also Kish (1965). Here, we use capital W for original survey weights, and small w for adjusted survey weights. The effective sample size n_e has some interesting properties. For example, by calculation, we get $n_e = \sum_{i=1}^n w_i = \sum_{i=1}^n w_i^2$; when W_i are almost equal, $n_e \approx n$. Weight trimming can make this happen.

To improve statistical efficiency and increase the robustness of statistical inferences, Winsorization is an effective way to deal with outliers (Rao, 1966, Basu, 1971, Haziza & Beaumont, 2017). Outliers here are defined as observations above $Q_3 + 1.5(Q_3 - Q_1)$, where Q_1=1st quartile, Q_3=3rd quartile. Let W^* be weights after trimming and $W_0 = Q_3 + 1.5(Q_3 - Q_1)$ denote the threshold value, then

$$W_i^* = \begin{cases} W_0, & W_i \geq W_0 \\ aW_i, & W_i < W_0, \end{cases} \tag{3}$$

where a is a rescaling parameter such that $\sum_{i=1}^n W_i^* = \sum_{i=1}^n W_i = \hat{N}$.

Then, it is reasonable to consider the adjusted trimmed weights w_i^*,

$$w_i^* = \frac{n_e^* W_i^*}{\sum_{j=1}^n W_j^*}, \quad i = 1, \ldots, n, \tag{4}$$

where

$$n_e^* = \frac{\left(\sum_{j=1}^n W_j^*\right)^2}{\sum_{j=1}^n W_j^{*2}}, \tag{5}$$

the effective sample size.

2.3 Normalized Weighted Density

We assume that

$$y_1, \ldots, y_N \mid \underset{\sim}{\theta} \overset{iid}{\sim} f(y \mid \underset{\sim}{\theta}),$$

and a probability sample of size n is taken from the finite population. That is, putting the sample first,

$$y_1, \ldots, y_n \mid \underset{\sim}{\theta} \overset{iid}{\sim} f(y \mid \underset{\sim}{\theta}).$$

We want to incorporate the survey weights in the likelihood.

Then, the loglikelihood for the entire population is

$$L(\underset{\sim}{\theta}) = \sum_{i=1}^N \log\{f(y_i|\underset{\sim}{\theta})\}.$$

The Horvitz-Thompson estimator of $L(\underset{\sim}{\theta})$ is

$$\widehat{L(\underset{\sim}{\theta})} = \sum_{i=1}^n W_i \log\{f(y_i|\underset{\sim}{\theta})\}.$$

Note that in this pseudo-loglikelihood, if we consider $\underset{\sim}{\theta}$ as fixed, then the y_i are independent; e.g., see Y. Chen et al. (2020). This motivates our normalized density.

We form a density function for y_i by taking

$$g(y_i|\theta) \propto f(y_i \mid \underset{\sim}{\theta})^{W_i}, i = 1, \ldots, n.$$

This is an unnormalized 'density'. We obtain the normalized density by inserting the normalization constant, $\int f(y_i \mid \underset{\sim}{\theta})^{W_i} dy_i$. Therefore,

$$g(y_i|\theta) = \frac{(f(y_i|\underset{\sim}{\theta}))^{W_i}}{\int (f(y_i|\underset{\sim}{\theta}))^{W_i} dy_i}. \quad (6)$$

is the normalized density (i.e., $g(y_i \mid \underset{\sim}{\theta})$ integrates to 1, but $f(y_i \mid \underset{\sim}{\theta})^{W_i}$ does not). A Bayesian should use the normalized density rather than the unnormalized density whether original weights or trimmed weights are being used. Note that $\int f(y_i \mid \underset{\sim}{\theta})^{W_i} dy_i$ is a function of $\underset{\sim}{\theta}$, and this is an adjustment to the density in (6). We can use $f(y_i \mid \underset{\sim}{\theta})^{W_i}$ as survey samplers do, but this is incorrect from a Bayesian point of view because, as we just stated, the denominator in (6) is a function of $\underset{\sim}{\theta}$.

In Appendix A, we show several illustrative examples of normalized densities, $g(y_i|\underset{\sim}{\theta})$ in order to demonstrate generality (pointed out by both reviewers). We specifically show how we can construct such normalized densities. Although we describe both continuous and discrete study variables in this paper we focus on binary variables. Our main problem is how to include survey weights into

$$y_1, \ldots, y_N \mid \theta \overset{ind}{\sim} \text{Bernoulli}(\theta)$$

and to make inference about the finite population proportion, $P = \frac{1}{N}\sum_{i=1}^{N} y_i$, when data, $(W_i, y_i), i = 1, \ldots, n$, are available. One issue that we do not address here is the incorporation of covariates, an important problem, but it does not help to understand the main issue of survey weights, but see Appendix D, where we describe the the problem under study. This is a problem of future study.

As a summary, we use the original survey weights W which are the 'inverse' of the inclusion probabilities for each unit to construct adjusted survey weights $\underset{\sim}{w}$ which make the sum of all adjusted weights equal the effective sample size, trimmed survey weights $\underset{\sim}{W^*}$ which trimmed outliers, and adjusted trimmed survey weights $\underset{\sim}{w^*}$, which make the sum of all adjusted trimmed weights equal the effective sample size. We also use normalized and unnormalized densities, but our preference is for normalized adjusted weights (original or trimmed).

3. Bayesian Methodology

Let $y_1, \ldots, y_N$, be the variable (study variable) of interest, where N is the number of units in the population.

By adjusting the sample weights, different surrogate samples from the original finite population were drawn and we compared their performance in making inferences about the original finite population (Nandram, 2007). Specifically,

$$y_1 \ldots, y_N \mid \theta \overset{\text{ind}}{\sim} \text{Bernoulli } (\theta). \quad (7)$$

Since there is no information about θ, we consider the proper but non-informative prior under the Bayesian framework,

$$\theta \sim \text{uniform } (0, 1). \quad (8)$$

Here (7) and (8) form the population model, which we assume to be true. Of course, we can have a more general prior, $\theta \sim$ Beta (α, β), where (α, β) must be specified based on the amount of prior information that is available.

Let $\underset{\sim}{y_s} = (y_1, \ldots, y_n)'$ and $\underset{\sim}{y_{ns}} = (y_{n+1}, \ldots, y_N)'$ denote respectively the sampled and non-sampled units.

Predictive inference about the finite population proportion, $P = \frac{1}{N}\sum_{i=1}^{N} y_i$, is obtained as follows,

$$\pi(P|\underset{\sim}{y_s}) = \int f(P|\theta)\pi(\theta|\underset{\sim}{y_s})d\theta.$$

where $f(P|\theta)$, the population model, does not depend on $\underset{\sim}{y_s}$. We simply draw θ from the posterior density $\pi(\theta|\underset{\sim}{y_s})$ and draw P from the population model; this is surrogate sampling (e.g., Nandram, 2007). Actually, letting $T = \sum_{i=1}^{N} y_i$, we have

$$\pi(T|\underset{\sim}{y_s}) = \int f(T|\theta)\pi(\theta|\underset{\sim}{y_s})d\theta,$$

where $T|\underset{\sim}{\theta} \sim \text{Binomial}(N, \theta)$; thereby providing a simpler computation.

Next, we describe the nine methods.

Method A: Without survey weights, the posterior distribution is

$$\theta \mid \underset{\sim}{y_s} \sim \text{Beta}\left(\sum_{i=1}^{n} y_i + 1, n - \sum_{i=1}^{n} y_i + 1\right). \quad (9)$$

As we discussed above, method A can lead to biased estimates, because the survey weights are not used. Considering this one as the baseline model, we can analyze how biased samples affect our predictive inference of a population quantity.

We now adjust the Bernoulli population model using the survey weights to reflect the sampling bias. Therefore, the weights are now to be added to

$$y_1, \ldots, y_n|\theta \overset{\text{ind}}{\sim} \text{Bernoulli}(\theta),$$

where the original survey weights, W, of $y_1, \ldots, y_n$ are known, and we can also construct adjusted survey weights, $\underset{\sim}{w}$, trimmed survey weights, $\underset{\sim}{W}^*$, and adjusted trimmed survey weights, $\underset{\sim}{w}^*$.

The remaining eight methods are broadly categorized as original weights or trimmed weights, and each of these is further categorized by unnormalized or normalized distribution, and unadjusted or adjusted weights, as we discussed in Table 1. This gives the sample model, an adjustment to correct for the selection bias.

Method B: Replace sample total with estimator of population total by using original survey weights,

$$\theta \mid \underset{\sim}{y_s} \sim \text{Beta}\left(\sum_{i=1}^{n} y_i W_i + 1, \sum_{i=1}^{n} (1 - y_i) W_i + 1\right). \tag{10}$$

Method C: Consider use the normalized likelihood function with original survey weights to update the posterior distribution,

$$f_C(\theta \mid \underset{\sim}{y_s}) \propto \frac{\theta^{\sum_{i=1}^{n} y_i W_i}(1-\theta)^{\sum_{i=1}^{n}(1-y_i)W_i}}{\prod_{i=1}^{n}\left[\theta^{W_i} + (1-\theta)^{W_i}\right]}. \tag{11}$$

In Appendix B, we show how to do the computation for big W_i.

The following methods are generated by replacing original survey weights W with adjusted survey weights w (method D and method E), trimmed survey weights $\underset{\sim}{W}^*$ (method F and method G), and adjusted trimmed survey weights $\underset{\sim}{w}^*$ (method H and method I).

Method D: Beta distribution with adjusted survey weights $\underset{\sim}{w}$ is

$$\theta \mid \underset{\sim}{y_s} \sim \text{Beta}\left(\sum_{i=1}^{n} y_i w_i + 1, \sum_{i=1}^{n} (1 - y_i) w_i + 1\right). \tag{12}$$

Method E: The normalized likelihood with adjusted survey weights $\underset{\sim}{w}$ is

$$f_E(\theta \mid \underset{\sim}{y_s}) \propto \frac{\theta^{\sum_{i=1}^{n} y_i w_i}(1-\theta)^{\sum_{i=1}^{n}(1-y_i)w_i}}{\prod_{i=1}^{n}\left[\theta^{w_i} + (1-\theta)^{w_i}\right]}. \tag{13}$$

Method F: Beta distribution with trimmed survey weights $\underset{\sim}{W}^*$ is

$$\theta \mid \underset{\sim}{y_s} \sim \text{Beta}\left(\sum_{i=1}^{n} y_i W_i^* + 1, \sum_{i=1}^{n} (1 - y_i) W_i^* + 1\right). \tag{14}$$

Method G: The normalized likelihood with trimmed survey weights $\underset{\sim}{W}^*$ is

$$f_G(\theta \mid \underset{\sim}{y_s}) \propto \frac{\theta^{\sum_{i=1}^{n} y_i W_i^*}(1-\theta)^{\sum_{i=1}^{n}(1-y_i)W_i^*}}{\prod_{i=1}^{n}\left[\theta^{W_i^*} + (1-\theta)^{W_i^*}\right]}. \tag{15}$$

In Appendix B, we also show how to do the computation for big W_i^*.

Method H: Beta distribution with adjusted trimmed survey weights $\underset{\sim}{w}^*$ is

$$\theta \mid \underset{\sim}{y_s} \sim \text{Beta}\left(\sum_{i=1}^{n} y_i w_i^* + 1, \sum_{i=1}^{n} (1 - y_i) w_i^* + 1\right). \tag{16}$$

Method I: The normalized likelihood with adjusted trimmed survey weights $\underset{\sim}{w}^*$ is

$$f_I(\theta \mid \underset{\sim}{y_s}) \propto \frac{\theta^{\sum_{i=1}^n y_i w_i^*}(1-\theta)^{\sum_{i=1}^n (1-y_i) w_i^*}}{\prod_{i=1}^n \left[\theta^{w_i^*} + (1-\theta)^{w_i^*}\right]}. \tag{17}$$

In short, there are nine methods for incorporating different survey weights into the Bayesian models.

In Appendix C, we consider a more general example of small area estimation because one reviewer asked about a more general example of a hierarchical Bayesian model. However, as we stated, there is very little loss in generality in our simple Bernoulli; the difference is essentially one of computation.

4. Simulation Study

The design of our simulation is inspired by the one implemented in Nandram (2007). The samples are given a random selection mechanism with unequal probabilities. But in the simulation section, to generate probability samples, it is assumed that at the stage of analysis, the selection probabilities are known, and our goal is to adjust for the selection bias by using a probability sample whose weights are known. We conduct the simulation under nine Bayesian methods.

Regardless of which methods we use, suppose we have random samples of H iterations from the posterior distribution, denoted by $\theta_1, \ldots, \theta_H$. By using surrogate sampling (e.g. Nandram, 2007), we have

$$\sum_{i=1}^N y_i \mid \theta_i \sim \text{Binomial}(N, \theta_i), i = 1, \ldots, H$$

Then,

$$f\left(\sum_{i=1}^N y_i | \underset{\sim}{y_s}\right) = \int_\Theta f\left(\sum_{i=1}^N y_i | \underset{\sim}{y_s}, \theta_i\right) \pi(\theta | y_s)\, d\theta_i.$$

From each iteration, we can draw θ_i, and then get the prediction of population quantities, such as finite population proportion, $\sum_{i=1}^N y_i/N$.

In addition, probability proportional to size (PPS) sampling is a special case within the Bayesian framework that deserves special attention. We can construct selection probabilities,

$$\pi_i = \frac{n x_i}{\sum_{i=1}^N x_i}, \quad i = 1, \ldots, N. \tag{18}$$

Besides, to compare the different drawing methods, Poisson sampling is also considered. Poisson sampling is a sampling process where every unit of the population is subjected to an independent Bernoulli trial can determine whether the unit is a part of the sample or not. Samples are selected as follows,

$$I_i \sim \text{Bernoulli}(\pi_i), \quad i = 1, \ldots, N;$$

again when $I_i = 1$, unit i is selected and when $I_i = 0$, unit i is not selected from population. Letting $n_0 = \sum_{i=1}^N I_i$ denote the size of the sample in Poisson sampling, we have $E(n_0) = \sum_{i=1}^N E(I_i) = n$, and $Var(n_0) = \sum_{i=1}^N \pi_i(1-\pi_i)$, so $n_0 \approx n$.

Consider a finite population of size N, and the sample units $y_1, \ldots, y_N$ are drawn with probability proportional to measures of size $x_1, \ldots, x_N$, which should be non-negative. Here, x is auxiliary variable and z is a latent variable, generated as follows:

$$x_i \overset{\text{iid}}{\sim} \text{Gamma}(\alpha, \beta), \tag{19}$$

$$z_i \mid x_i \overset{\text{iid}}{\sim} N\left(\frac{\rho}{\sigma_x}(x_i - 1.2\mu_x), 1-\rho^2\right), \quad i = 1 \ldots N, \tag{20}$$

where ρ is the correlation coefficient, $\mu_x = \frac{\alpha}{\beta}$ and $\sigma_x^2 = \frac{\alpha}{\beta^2}$.

Then, we use the latent variable to generate binary responses,

$$y_i = \begin{cases} 1, & z_i \geq 0 \\ 0, & z_i < 0 \end{cases}, \quad i = 1 \ldots N. \tag{21}$$

In this simulation study, our interest is in the finite population proportion, $P = \frac{1}{N}\sum_{i=1}^{N} z_i$; in the simulation study P is known, and we look to see how well our model predicts it. In a biased sample, the sampled values are taken with unequal selection probabilities, which depend on the characteristic y. The sample model is assumed to correct for the selection bias, which is accommodated by the survey weights.

Now, we can perform the simulation study to access the estimators of the finite population under probability proportional to size (PPS) and Poisson sampling with respect to the measure of size $x_i, i = 1, \ldots, N$, and the effective size of samples should be equal to or around $n = 100$ according to different sampling method. Keeping the population size fixed at $N = 1000$ and $\beta = 1$ (i.e., $\mu_x = \alpha$), we can generate $H = 1000$ datasets at $\rho = \{0.2, 0.5, 0.8\}$ and $\alpha = \{2, 5, 15\}$, which means there are nine design points for each posterior distribution.

To evaluate the repeated surrogate sampling properties of our nine methods, average posterior mean (APM), average posterior standard deviation (APSD), average relative bias (ARB), average posterior root mean squared error (APRMSE) and the proportion of the 95% highest posterior density intervals (HPDI) of P containing the true P (PCI) are calculated as below:

Let $P = \sum_{i=1}^{N} y_i/N$ denote the finite population proportion. Then,

$$PM = E(P|\underset{\sim}{y_s}), \tag{22}$$

$$PSD = \sqrt{Var(P|\underset{\sim}{y_s})}, \tag{23}$$

$$95\%\ \text{HPDI} = (C_{025}, C_{975}), \tag{24}$$

where C_{025} and C_{975} denote the end points, not necessarily percentiles.

Performing simulations H times, we get

$$PM^{(h)}, PSD^{(h)}, (C_{025}^{(h)}, C_{975}^{(h)}), \quad h = 1, \ldots, H.$$

Now, we compute

$$APM = \frac{1}{H}\sum_{h=1}^{H} PM^{(h)}, \tag{25}$$

$$APSD = \frac{1}{H}\sum_{h=1}^{H} PSD^{(h)}, \tag{26}$$

$$ARB = \frac{1}{H}\sum_{h=1}^{H} \left(PM^{(h)} - P\right)/P, \tag{27}$$

$$APRMSE = \frac{1}{H}\sum_{h=1}^{H} \sqrt{(PM^{(h)} - P)^2 + (PSD^{(h)})^2}, \tag{28}$$

$$PCI = \frac{1}{H}\sum_{h=1}^{H} I\left(C_{025}^{(h)} \leq P \leq C_{975}^{(h)}\right), \tag{29}$$

where P is the true finite population proportion. These are standard quantities used in most simulation studies of this kind.

In Table 2 and Table 3, keeping ρ fixed and increasing α, the population mean decreases, and we can find the estimators from eight methods with survey weights performed better than method A without survey weights. With increased α and ρ, population distribution is more right-skewed. In this case, methods with survey weights still performed better than method A. This shows methods with survey weights work better when the population is skewed. Besides, since the effective sample size is almost equal to the sample size among methods with original survey weights and methods with trimmed survey weights, the number of outliers could be small and there is no significant difference in average posterior means. Estimators from methods including survey weights are more close to the population mean, which means incorporating survey weights into Bayesian models can reduce survey bias in both PPS sampling and Poisson sampling.

Table 2. PPS: Comparisons of the average posterior mean (APM) using nine posterior distributions of the finite population by ρ and α

ρ	α	A	B	C	D	E	F	G	H	I	P
0.2	2	0.5270	0.4753	0.4890	0.4761	0.4740	0.4847	0.4943	0.4850	0.4849	0.4766
	5	0.4960	0.4619	0.4840	0.4627	0.4618	0.4648	0.4860	0.4656	0.4654	0.4643
	15	0.4564	0.4355	0.4708	0.4368	0.4363	0.4361	0.4717	0.4374	0.4372	0.4382
0.5	2	0.5580	0.4308	0.4633	0.4333	0.4253	0.4544	0.4802	0.4556	0.4547	0.4328
	5	0.4894	0.4028	0.4462	0.4051	0.4021	0.4115	0.4531	0.4133	0.4122	0.4043
	15	0.3938	0.3433	0.3921	0.3464	0.3447	0.3451	0.3943	0.3479	0.3467	0.3454
0.8	2	0.5659	0.3590	0.4057	0.3640	0.3443	0.3981	0.4409	0.4005	0.3983	0.3654
	5	0.4624	0.3287	0.3771	0.3325	0.3245	0.3389	0.3869	0.3422	0.3389	0.3333
	15	0.3269	0.2548	0.3006	0.2594	0.2552	0.2563	0.3021	0.2610	0.2576	0.2576

Table 3. Poisson sampling: Comparisons of the average posterior mean (APM) using nine posterior distributions of the finite population by ρ and α

ρ	α	A	B	C	D	E	F	G	H	I	P
0.2	2	0.5307	0.4760	0.4894	0.4769	0.4741	0.4879	0.4957	0.4882	0.4879	0.4763
	5	0.4986	0.4638	0.4856	0.4646	0.4637	0.4674	0.4876	0.4681	0.4677	0.4636
	15	0.4609	0.4395	0.4733	0.4407	0.4402	0.4405	0.4741	0.4417	0.4414	0.4383
0.5	2	0.5662	0.4349	0.4659	0.4373	0.4299	0.4621	0.4834	0.4630	0.4623	0.4332
	5	0.4934	0.4067	0.4499	0.4088	0.4061	0.4149	0.4564	0.4166	0.4158	0.4047
	15	0.3983	0.3479	0.3969	0.3508	0.3492	0.3497	0.3989	0.3525	0.3515	0.3451
0.8	2	0.5799	0.3700	0.4165	0.3747	0.3571	0.4103	0.4506	0.4124	0.4106	0.3646
	5	0.4752	0.3398	0.3888	0.3435	0.3366	0.3501	0.3992	0.3533	0.3505	0.3346
	15	0.3298	0.2566	0.3024	0.2613	0.2570	0.2583	0.3042	0.2629	0.2596	0.2577

As for comparisons of the APSD (see Table 4 and Table 5), in general, methods B, C, F, and G have smaller APSDs than others. The reason is that no matter whether in our simulation or real datasets, the survey weights could be very large. For example, in the method B,

$$\theta \mid \underset{\sim}{y_s} \sim \text{Beta}\left(\sum_{i=1}^{n} y_i W_i + 1, \sum_{i=1}^{n} (1 - y_i) W_i + 1\right),$$

when survey weights W_i are large, $\text{Var}(\theta \mid \underset{\sim}{y_s}) \approx 0$. After we normalized the distribution, the APSDs of the variable decreased. However, if we adjusted the survey weights first, APSDs of methods D, E, H, and I are similar to that of method A. Because the effective sample size is almost equal to the sample size among methods with original survey weights and methods with trimmed survey weights, the number of outliers could be small and there is no significant difference in average posterior standard deviations. From these tables, it seems the APSDs are mostly related to parameter ρ. The larger ρ is, the lower APSDs will be. Clearly, and as is well-known in the survey literature, posterior standards under methods B, C, F and G are not appropriate, and therefore we can weed out the use of W_i and W_i^* if posterior standard deviation is needed as is usually the case.

For ARBs in Table 6 and Table 7, as the population means decrease, the ARBs of method A are increasing. When α is fixed and ρ is increasing, ARBs of all these methods are increasing. For methods B and F (survey weights and trimmed survey weights with beta distribution), the ARBs are more robust and stable than others. But after considering normalized the distribution (method C and G), ARBs are larger when ρ is larger. But for adjusted original survey weights (method D and method E) and adjusted trimmed survey weights (method H and method I), there are no significant differences in unnormalized distribution and normalized distribution.

Table 8 and Table 9 indicate that method A (no survey weights) has larger APRMSEs than the other methods. Keeping ρ constant and changing α from 2 to 15, we can see that APRMSEs of almost all methods are negatively related to α, except for method C and method G (a normalized distribution with original survey weights and trimmed survey weights), APRMSEs get larger when α increases. If the α is constant, APRMSEs of method A (no survey weights) tend to be

Table 4. PPS: Comparisons of the average posterior standard deviation (APSD) using nine posterior distributions of the finite population by ρ and α

ρ	α	A	B	C	D	E	F	G	H	I
0.2	2	0.0514	0.0220	0.0161	0.0664	0.0692	0.0221	0.0162	0.0576	0.0580
	5	0.0515	0.0221	0.0161	0.0564	0.0571	0.0221	0.0162	0.0548	0.0550
	15	0.0512	0.0220	0.0161	0.0526	0.0527	0.0220	0.0161	0.0523	0.0524
0.5	2	0.0512	0.0219	0.0159	0.0657	0.0690	0.0220	0.0161	0.0575	0.0580
	5	0.0515	0.0217	0.0158	0.0555	0.0565	0.0218	0.0159	0.0540	0.0544
	15	0.0503	0.0211	0.0154	0.0505	0.0508	0.0211	0.0154	0.0503	0.0506
0.8	2	0.0510	0.0212	0.0155	0.0640	0.0687	0.0216	0.0158	0.0565	0.0573
	5	0.0513	0.0208	0.0153	0.0532	0.0549	0.0210	0.0153	0.0521	0.0529
	15	0.0483	0.0193	0.0144	0.0464	0.0468	0.0194	0.0144	0.0464	0.0468

Table 5. Poisson sampling: Comparisons of the average posterior standard deviation (APSD) using nine posterior distributions of the finite population by ρ and α

ρ	α	A	B	C	D	E	F	G	H	I
0.2	2	0.0513	0.0222	0.0161	0.0660	0.0686	0.0222	0.0162	0.0575	0.0578
	5	0.0517	0.0222	0.0162	0.0567	0.0574	0.0222	0.0162	0.0550	0.0552
	15	0.0514	0.0221	0.0161	0.0527	0.0528	0.0221	0.0161	0.0526	0.0527
0.5	2	0.0511	0.0221	0.0159	0.0661	0.0695	0.0222	0.0161	0.0576	0.0580
	5	0.0515	0.0219	0.0158	0.0556	0.0566	0.0219	0.0159	0.0542	0.0546
	15	0.0505	0.0212	0.0154	0.0507	0.0511	0.0212	0.0154	0.0506	0.0509
0.8	2	0.0508	0.0215	0.0156	0.0644	0.0688	0.0219	0.0159	0.0568	0.0575
	5	0.0515	0.0211	0.0153	0.0538	0.0553	0.0213	0.0154	0.0526	0.0534
	15	0.0484	0.0194	0.0145	0.0465	0.0470	0.0194	0.0145	0.0464	0.0469

Table 6. PPS: Comparisons of the average relative bias (ARB) using nine posterior distributions of the finite population by ρ and α

ρ	α	A	B	C	D	E	F	G	H	I
0.2	2	0.1202	0.1104	0.0639	0.1064	0.1178	0.0913	0.0553	0.0893	0.0904
	5	0.0976	0.0928	0.0674	0.0906	0.0929	0.0887	0.0672	0.0869	0.0877
	15	0.0918	0.0891	0.0951	0.0873	0.0883	0.0886	0.0961	0.0869	0.0876
0.5	2	0.2898	0.1170	0.1087	0.1126	0.1297	0.1093	0.1206	0.1078	0.1089
	5	0.2138	0.1017	0.1273	0.0990	0.1043	0.1004	0.1378	0.0988	0.1005
	15	0.1570	0.0994	0.1542	0.0974	0.0996	0.0996	0.1589	0.0978	0.0993
0.8	2	0.5497	0.1183	0.1477	0.1116	0.1514	0.1353	0.2141	0.1362	0.1374
	5	0.3885	0.1120	0.1592	0.1090	0.1222	0.1148	0.1806	0.1136	0.1181
	15	0.2780	0.1234	0.1943	0.1208	0.1256	0.1237	0.1985	0.1216	0.1248

Table 7. Poisson sampling: Comparisons of the average relative bias (ARB) using nine posterior distributions of the finite population by ρ and α

ρ	α	A	B	C	D	E	F	G	H	I
0.2	2	0.1253	0.1093	0.0625	0.1055	0.1164	0.0947	0.0561	0.0925	0.0936
	5	0.1010	0.0893	0.0678	0.0873	0.0895	0.0860	0.0684	0.0842	0.0851
	15	0.0966	0.0908	0.0984	0.0893	0.0902	0.0904	0.0993	0.0889	0.0892
0.5	2	0.3075	0.1181	0.1115	0.1137	0.1306	0.1165	0.1260	0.1152	0.1163
	5	0.2213	0.0990	0.1312	0.0973	0.1016	0.0998	0.1413	0.0985	0.0999
	15	0.1723	0.1071	0.1705	0.1057	0.1073	0.1075	0.1745	0.1063	0.1082
0.8	2	0.5915	0.1259	0.1728	0.1216	0.1542	0.1599	0.2420	0.1613	0.1622
	5	0.4210	0.1122	0.1812	0.1113	0.1195	0.1203	0.2057	0.1210	0.1231
	15	0.2888	0.1213	0.1972	0.1190	0.1237	0.1221	0.2029	0.1202	0.1239

Table 8. PPS: Comparisons of the average posterior root mean square error (APRMSE) using nine posterior distributions of the finite population by ρ and α

ρ	α	A	B	C	D	E	F	G	H	I
0.2	2	0.0812	0.0596	0.0361	0.0881	0.0942	0.0514	0.0323	0.0758	0.0765
	5	0.0727	0.0510	0.0363	0.0745	0.0757	0.0494	0.0361	0.0720	0.0724
	15	0.0689	0.0474	0.0456	0.0688	0.0692	0.0472	0.0460	0.0684	0.0687
0.5	2	0.1368	0.0577	0.0507	0.0862	0.0940	0.0547	0.0553	0.0785	0.0792
	5	0.1037	0.0490	0.0551	0.0723	0.0745	0.0487	0.0589	0.0712	0.0719
	15	0.0777	0.0427	0.0568	0.0639	0.0646	0.0427	0.0584	0.0638	0.0644
0.8	2	0.2074	0.0507	0.0577	0.0799	0.0935	0.0564	0.0806	0.0800	0.0809
	5	0.1405	0.0452	0.0567	0.0680	0.0724	0.0461	0.0634	0.0680	0.0698
	15	0.0893	0.0393	0.0533	0.0588	0.0600	0.0394	0.0543	0.0589	0.0598

Table 9. Poisson sampling: Comparisons of the average posterior root mean square error (APRMSE) using nine posterior distributions of the finite population by ρ and α

ρ	α	A	B	C	D	E	F	G	H	I
0.2	2	0.0831	0.0592	0.0354	0.0875	0.0932	0.0530	0.0325	0.0768	0.0773
	5	0.0738	0.0494	0.0362	0.0734	0.0747	0.0481	0.0366	0.0712	0.0716
	15	0.0703	0.0480	0.0468	0.0693	0.0697	0.0478	0.0472	0.0691	0.0693
0.5	2	0.1439	0.0582	0.0518	0.0869	0.0947	0.0578	0.0574	0.0810	0.0817
	5	0.1062	0.0481	0.0566	0.0717	0.0737	0.0484	0.0603	0.0710	0.0718
	15	0.0817	0.0450	0.0620	0.0660	0.0668	0.0451	0.0633	0.0661	0.0667
0.8	2	0.2216	0.0532	0.0662	0.0824	0.0940	0.0646	0.0901	0.0867	0.0875
	5	0.1508	0.0454	0.0638	0.0687	0.0721	0.0477	0.0716	0.0698	0.0710
	15	0.0916	0.0392	0.0541	0.0590	0.0603	0.0394	0.0555	0.0592	0.0602

greater with increasing ρ, but there is no significant pattern for other methods.

Table 10 and Table 11 are highly related to Table 4 and Table 5. As we mentioned before, unnormalized or normalized distributions with original survey weights (method B and method C) or with trimmed survey weights (method F and method G) have smaller APSDs than others, since they used the large survey weights without adjustment. In this situation, it causes the width of the 95% confidence interval to be very short. Then the probability of this 95% confidence interval to cover the population mean decreases. For unnormalized or normalized distribution with adjusted original survey weights (method D and method E) or with adjusted trimmed survey weights (method H and method I), they have similar APSDs, so their width of 95% confidence intervals is larger, but the average relative bias of method A is the largest, which leads to the lower 95% confidence interval coverage when APMs of method A are biased. When we focus on the models with high PCI, we can figure out that model D is almost the same as model E. Similarly, model H performs as model I. So no matter whether the model used adjusted original weights or adjusted trimmed weights, there is no winner between the unnormalized model and the normalized model.

The main reason why methods (D, E) differ very little from methods (H, I) is that the simulation proceeds with very few outliers. This makes trimming the weights less useful. However, if there are outliers in the weights, these two will differ. Also, there are very little differences between unnormalized and normalized whether the adjusted weights come from the original weights or the trimmed weights. As we will see in an example on body mass index data, there are differences among these methods, and the main cause is whether there are outliers or not in the weights.

5. Application on Body Mass Index

In this section, we apply our methods to the body mass index (BMI) from NHANES III (Nandram & Choi, 2005, 2010). We use eight counties, which are included in California, from NHANES III. In these datasets, original sample weights for each county are given.

The datasets contain *age*, *race* and *sex* as observed covariates, where *age* is collected as integers from 20 to 90; *race* uses {0, 1} to denote Hispanic and non-Hispanic; *sex* is represented by 0 for male and 1 for female. Body mass index (BMI) is a simple index of weight for height that is commonly used to classify overweight and obesity in adults. It is defined

Table 10. PPS: Comparisons of the proportion of these 95% credible intervals containing θ (PCI) using nine posterior distributions of the finite population by ρ and α

ρ	α	A	B	C	D	E	F	G	H	I
0.2	2	0.858	0.490	0.633	0.971	0.959	0.581	0.709	0.964	0.964
	5	0.930	0.569	0.544	0.963	0.955	0.590	0.551	0.962	0.960
	15	0.948	0.628	0.340	0.959	0.958	0.634	0.346	0.963	0.961
0.5	2	0.320	0.495	0.296	0.971	0.957	0.539	0.223	0.941	0.950
	5	0.651	0.587	0.321	0.970	0.964	0.599	0.260	0.965	0.962
	15	0.877	0.677	0.315	0.982	0.972	0.676	0.298	0.980	0.976
0.8	2	0.016	0.565	0.339	0.981	0.963	0.515	0.157	0.916	0.912
	5	0.261	0.603	0.328	0.978	0.970	0.602	0.281	0.969	0.964
	15	0.708	0.656	0.312	0.985	0.979	0.657	0.296	0.980	0.976

Table 11. Poisson sampling: Comparisons of the proportion of these 95% credible intervals containing θ (PCI) using nine posterior distributions of the finite population by ρ and α

ρ	α	A	B	C	D	E	F	G	H	I
0.2	2	0.827	0.487	0.649	0.969	0.960	0.553	0.681	0.966	0.964
	5	0.89	0.600	0.545	0.976	0.970	0.636	0.528	0.969	0.963
	15	0.944	0.606	0.324	0.970	0.971	0.606	0.319	0.969	0.967
0.5	2	0.251	0.497	0.295	0.976	0.960	0.496	0.201	0.935	0.89
	5	0.614	0.605	0.302	0.977	0.970	0.603	0.251	0.974	0.972
	15	0.833	0.646	0.275	0.966	0.967	0.651	0.268	0.967	0.964
0.8	2	0.005	0.539	0.269	0.977	0.959	0.440	0.114	0.876	0.878
	5	0.205	0.619	0.270	0.985	0.979	0.589	0.222	0.968	0.963
	15	0.709	0.680	0.294	0.971	0.967	0.682	0.276	0.970	0.969

as a person's weight in kilograms divided by the square of his height in meters (kg/m^2). The World Health Organization defined obesity as a BMI greater than or equal to 30. In this dataset, we focus on obesity, which means $y_i = I(\text{BMI}_i \geq 30)$, $i = 1, \ldots, n$. Here, our sample size is $n = 1867$ for eight counties combined in California. We also give separate analyses for each county.

We present posterior summaries in Tables 12 & 13. Note that n is the sample size, n_e is the effective sample size generated by adjusted (according to original survey weights or trimmed survey weights), n_e^* is the effective sample size generated by adjusted weights (according to original survey weights or trimmed survey weights), n_O is the number of outliers detected in weight trimming.

Table 12. Posterior mean (PM) of nine methods for BMI data by different areas

n	A	B	C	D	E	F	G	H	I	n_e	n_e^*	n_O
164	0.254	0.220	0.221	0.228	0.188	0.230	0.231	0.237	0.221	76	86	3
176	0.246	0.168	0.168	0.182	0.040	0.257	0.257	0.258	0.255	50	159	23
795	0.228	0.235	0.235	0.239	0.192	0.234	0.235	0.234	0.220	180	593	113
162	0.182	0.127	0.128	0.148	0.047	0.154	0.155	0.165	0.127	33	64	9
125	0.234	0.209	0.209	0.224	0.144	0.233	0.234	0.241	0.224	33	94	26
141	0.251	0.159	0.159	0.170	0.133	0.234	0.235	0.240	0.230	41	124	31
128	0.215	0.150	0.151	0.161	0.089	0.172	0.172	0.179	0.147	56	83	7
176	0.231	0.172	0.174	0.181	0.085	0.218	0.218	0.221	0.212	69	142	37
1867	0.228	0.192	0.192	0.195	0.147	0.221	0.221	0.221	0.199	498	1301	323

Table 13. Posterior standard deviation (PSD) of nine methods for BMI data by different areas

n	A	B	C	D	E	F	G	H	I	n_e	n_e^*	n_O
164	0.036	0.013	0.013	0.049	0.049	0.013	0.014	0.047	0.048	76	86	3
176	0.035	0.011	0.012	0.054	0.034	0.014	0.014	0.036	0.037	50	159	23
795	0.020	0.014	0.014	0.034	0.035	0.013	0.013	0.022	0.022	180	593	113
162	0.031	0.011	0.011	0.060	0.029	0.012	0.012	0.047	0.046	33	64	9
125	0.040	0.013	0.013	0.070	0.072	0.014	0.013	0.045	0.046	33	94	26
141	0.039	0.012	0.012	0.058	0.060	0.013	0.014	0.040	0.040	41	124	31
128	0.038	0.011	0.011	0.048	0.029	0.012	0.012	0.043	0.042	56	83	7
176	0.035	0.011	0.012	0.047	0.028	0.013	0.013	0.038	0.039	69	142	37
1867	0.017	0.012	0.012	0.022	0.012	0.013	0.013	0.017	0.017	498	1301	323

From Table 12, PMs of method E (a normalized distribution with adjusted original survey weights) are smaller than PMs of other methods. It is possible that outliers of survey weights affected the estimators badly. After trimming these outliers, PMs from methods F, G, H, and I (distributions with trimmed survey weights) are more stable and reasonable. There are outliers in the original survey weights, but these are removed in the trimmed weights. The outliers can cause the adjusted weights to be dominated by many small adjusted weights (weights much smaller than unity, with just a few weights larger than unity). This is the reason why D and E are so different in some cases.

From Table 13, PSDs of models with adjusted survey weights (methods D, E, H, and I) are greater than that of models with unadjusted survey weights (methods B, C, F, and G). This is one finding from the simulation study, but it should be obvious to most survey statisticians. More importantly, there are some examples where the PSDs in E are much smaller than those in D, but this is not so for H and I. Clearly, the outliers in the original weights can make a significant difference.

Also, in the simulation, the sampling fraction is $100/(100 + 900) \approx 0.10$, but in the BMI dataset, for county 59, there are 162 people in the sample and 1,353,001 people in the non-sample (i.e., the sampling fraction, $162/(162 + 1,353,001) \approx 0.00012$, is very small). To assure that the population size is not an issue in the simulation study, we also modified the original survey weights such that the sum of weights is around 1000 as the population size of the simulation. But because multiplying one constant does not have any influence on the effective sample size or boundary of outliers, there are no big differences between PM tables as well as PSD tables.

In this BMI application, trimming survey weights is essential for inference. It appears that if there are outliers in the original weights, normalization will make a difference; otherwise, it does not (compare D and E and H and I). In the simulation, the effective sample size is close to the sample size, which did not lead to the same conclusion as the BMI dataset, and there were no outliers in the original weights.

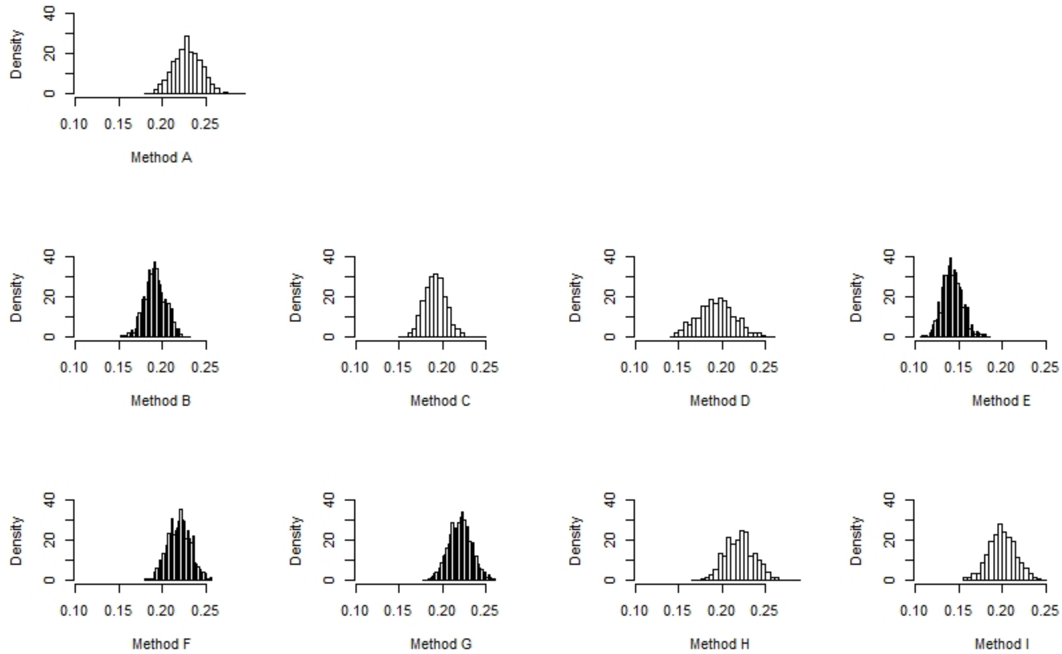

Figure 1. Histogram of surrogate posterior means on eight counties under the nine methods

To see more details, we draw the histograms of the finite population proportions under the nine methods for the whole dataset, where $n = 1867$. Without survey weights, the distribution of PMs of method A is a little bit on the right of others. Also, after we trimmed the survey weights outliers, the distributions under methods F, G H and I are on the right of those under methods B, C, D and E. The normalization does not affect a lot between methods B and C (distributions with original survey weights), methods F and G (trimmed survey weights), and methods H and I (adjusted trimmed survey weights), but distribution under method E is significantly sharper and smaller than that under method D since outliers are critical in normalization. Too small or too large weights can cause problems in large national surveys like NHANES, done using probability-proportional-to-size survey designs.

6. Concluding Remarks

The discussion is motivated by the desire to make predictions and inferences about a finite population quantity from biased samples by generating surrogate samples. The advantage of using a Bayesian method is evident when we use the correct density of y, which is generally an awkward density; see Appendix A. The correct density, based on $f(y_i \mid \theta)$, should be normalized and the normalization constant is a function of θ. In the Bayesian approach, essentially we need the posterior density of θ, and we can get samples from it regardless of its complexity. Besides we can input any available prior information.

According to our simulation study, the performance of these nine methods is different. The results of PCI tables (Table 10 and Table 11) show that models with adjusted survey weights (method D, E, H, and I) higher coverage than models with unadjusted survey weights (method A, B, C, F, and G), no matter whether it is the unnormalized or normalized model with original survey weights or trimmed survey weights. It is clear that the W_i should not be used in a parametric model if we need a measure of uncertainty at the same time.

In the analysis of the BMI data, we have seen that these nine methods are able to deal with data that contain extreme survey weights and make proper inferences about the population, and this is an advantage of the Bayesian framework. Method E (a normalized distribution with adjusted original survey weights) is worse for inference because of the outliers. If we do not trim the weights, there will be differences between the unnormalized and normalized densities. The normalized density will provide smaller posterior standard deviations, but this is not necessarily correct.

In future work, it is reasonable to consider how to incorporate probability samples with non-probability samples to make full use of known information. Once the survey weights are obtained, they can be incorporated into a Bayesian model, regardless of the form as we have done here.

When there are covariates, $\underset{\sim}{x_i}$ ($p-1$ covariates and an intercept), it is a standard practice to use a logistic regression model. So we assume logistic regression for the population model,

$$y_i|\underset{\sim}{\beta} \overset{ind}{\sim} \text{Bernoulli}\left\{\frac{e^{\underset{\sim}{x_i}'\underset{\sim}{\beta}}}{1+e^{\underset{\sim}{x_i}'\underset{\sim}{\beta}}}\right\}, i=1,\dots,N,$$

where N, the population size; see Appendix D for the sample model, where the survey weights are included. The non-sampled covariates, $\underset{\sim}{x_i}$, are typically unknown; see Appendix D for further details.

Acknowledgement

The authors thank the two reviewers for very informative comments that led to an improved manuscript. Balgobin Nandram was supported by a grant from the Simons Foundation (#353953, Balgobin Nandram).

Appendix A. Examples of Normalized Densities with Survey Weights

We give several examples of normalized densities when survey weights are incorporated into a likelihood. Our main purpose is to show that we can deal with other data besides the simple Bernoulli model that is used to illustrate the general principle. In all examples, we use adjusted original weights, w_i, the examples are similar for W_i or trimmed weights, w_i^*, and W_i^*. Obviously, the list is not exhaustive.

Example 1 Normal distribution, $f(y_i|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(y_i-\mu)^2}$, $-\infty<\mu<\infty, \sigma^2>0$.

$$y_i \mid \mu,\sigma^2 \overset{\text{ind}}{\sim} \text{Normal}\left(\mu, \frac{\sigma^2}{w_i}\right).$$

Suppose the prior distribution is $f(\mu,\sigma^2) = \frac{1}{\sigma^2}$. Then, the joint distribution is

$$f(\underset{\sim}{y},\mu,\sigma^2) = \prod_{i=1}^{n}\left\{\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(y_i-\mu)^2}\right\}\frac{1}{\sigma^2}.$$

In this way, it is easy to use the multiplication rule to draw parameters μ and σ^2 from their closed-form posterior distributions,

$$\mu|\sigma^2,\underset{\sim}{y} \sim \text{Normal}\left(\frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i}, \frac{\sigma^2}{\sum_{i=1}^n w_i}\right),$$

$$\sigma^2|\underset{\sim}{y} \sim \text{InvG}\left(\frac{n-1}{2}, \frac{\sum_{i=1}^n w_i(y_i-\bar{y})^2}{2}\right).$$

Notice that $\sigma^2|\underset{\sim}{y}$ has much smaller variance than if W_i is used, but $\frac{\sum_{i=1}^n W_i y_i}{\sum_{i=1}^n W_i}$ and $\frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i}$ are the same.

Example 2 Lognormal distribution, $f(y_i|\mu,\sigma^2) = \frac{1}{y_i\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(ln(y_i)-\mu)^2}$, $-\infty<\mu<\infty, \sigma^2>0$.

we can define $z_i = \ln(y_i)$, $y_i > 0$. Then,

$$z_i \mid \mu,\sigma^2 \overset{\text{ind}}{\sim} \text{Normal}\left(\mu, \frac{\sigma^2}{w_i}\right),$$

where again it is easy to draw parameters μ and σ^2 as above.

Example 3 Gamma distribution, $f(y_i \mid \alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}y_i^{\alpha-1}e^{-\beta y_i}$, $\alpha>1, \beta>0$.

The pseudo-likelihood function with weights is

$$y_i \mid \alpha,\beta \overset{\text{ind}}{\sim} \text{Gamma}\left(w_i\alpha - w_i + 1, \beta w_i\right).$$

Notice that the general restriction of gamma distribution is $\alpha > 0$ rather than $\alpha > 1$. But to assure the pseudo-likelihood function is well-defined, we need $w_i\alpha - w_i + 1 > 0$ for $i = 1,\dots,n$, which means $\alpha > \max\frac{w_i-1}{w_i}$. And to simplify it, we just set it as $\alpha > 1$.

Example 4 Student's t distribution, $f(y_i|\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\left(1 + \frac{y_i^2}{\nu}\right)^{-\frac{\nu+1}{2}}$.

The pseudo-likelihood function is hard to calculate because of the denominator. Suppose,

$$\frac{(y_i - \mu)}{\sigma} \overset{\text{ind}}{\sim} t_\nu,$$

where $f(y_i|\mu, \sigma^2) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi\sigma^2}\Gamma(\frac{\nu}{2})}\left(1 + \frac{(y_i-\mu)^2}{\nu\sigma^2}\right)^{-\frac{\nu+1}{2}}$. The numerator of pseudo-likelihood function with weights is,

$$(f(y_i|\nu))^{w_i} = \left(\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi\sigma^2}\Gamma\left(\frac{\nu}{2}\right)}\right)^{w_i}\left(1 + \frac{(y_i - \mu)^2}{\nu\sigma^2}\right)^{-\frac{w_i(\nu+1)}{2}}.$$

Let $a_i + 1 = w_i(\nu + 1)$, with $\nu > 0$ and $a_i > w_i - 1$. By calculating, the pseudo-likelihood function is,

$$\frac{(y_i - \mu)}{\sigma\frac{a_i+1-w_i}{a_iw_i}} \overset{\text{ind}}{\sim} t_{a_i},$$

where

$$g(y_i|\mu, \sigma^2) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{a_i\pi\sigma^2\left(\frac{a_i+1-w_i}{a_iw_i}\right)}\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \frac{(y_i - \mu)^2}{a_i\sigma^2\frac{a_i+1-w_i}{a_iw_i}}\right)^{-\frac{a+1}{2}}.$$

When $\underset{\sim}{w}$ are the original survey weights, the condition $a_i > w_i - 1$ leads to a big degree of freedom and makes it approximate to a normal distribution, not a student's t distribution. When $\underset{\sim}{w}$ are the standardized or adjusted survey weights as method D, E, H, and I, it is possible to get some $w_i < 1$ and cause a_i negative. Therefore, $\frac{a_i+1-w_i}{a_iw_i} > 0$ is not practical for all samples.

To deal with this situation and get closed-form likelihood function of y_i,

$$y_i \mid a^2, \mu, \sigma^2 \overset{\text{ind}}{\sim} \text{Normal}\left(\mu, \frac{a^2\sigma^2}{w_i}\right),$$

$$\frac{\nu}{a^2} \sim \text{Gamma}\left(\frac{\nu}{2}, \frac{1}{2}\right).$$

In this model, the pseudo-likelihood function is easy to get,

$$\frac{\sqrt{w_i}\,(y_i - \mu)}{\sigma} \overset{\text{ind}}{\sim} t_\nu,$$

where $g(y_i|\mu, \sigma^2) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi\frac{\sigma^2}{w_i}}\Gamma(\frac{\nu}{2})}\left(1 + \frac{w_i(y_i-\mu)^2}{\nu\sigma^2}\right)^{-\frac{\nu+1}{2}}$.

Example 5 Multinomial distribution, $f(\underset{\sim}{y_i}|p_1, \ldots, p_k) = \frac{p_1^{w_iy_{i1}} \cdots p_k^{w_iy_{ik}}}{p_1^{w_i} + \cdots + p_k^{w_i}}$, $\sum_{i=1}^k p_i = 1$, $\sum_{j=1}^k y_{ij} = 1$, and $k > 0$ is the number of groups.

Then, by calculation, the pseudo-likelihood function is

$$g(\underset{\sim}{y_i}|p_1, \ldots, p_k) \overset{\text{ind}}{\sim} \text{Multi}\left(1, \frac{p_1^{w_1}}{\sum_{j=1}^k p_j^{w_j}}, \cdots, \frac{p_k^{w_k}}{\sum_{j=1}^k p_j^{w_j}}\right)$$

.

B. Algorithm for Original Survey Weights $\underset{\sim}{W}$

For equation,

$$\pi(p|\underset{\sim}{y}) \propto \prod_{i=1}^n \left\{\frac{p^{y_iW_i}(1 - p)^{(1-y_i)W_i}}{p^{W_i} + (1 - p)^{W_i}}\right\},$$

when survey weights are large (W), since $0 \leq p \leq 1$, the denominator would be close to 0. Then, it is hard to calculate the probability density at each p. To solve this problem, we can rewrite $\pi(p|\underset{\sim}{y})$ and make a transformation. In this case, p is approximately a point mass at a possibly unknown point. Then, we have

$$\pi(p|\underset{\sim}{y}) = \prod_{\{i:y_i=1\}} \frac{p^{W_i}}{p^{W_i} + (1-p)^{W_i}} \prod_{\{i:y_i=0\}} \frac{(1-p)^{W_i}}{p^{W_i} + (1-p)^{W_i}}$$
$$= \prod_{\{i:y_i=1\}} \frac{(\frac{p}{1-p})^{W_i}}{1 + (\frac{p}{1-p})^{W_i}} \prod_{\{i:y_i=0\}} \frac{1}{1 + (\frac{p}{1-p})^{W_i}}.$$

If we assume that $0 \leq p \leq \frac{1}{2}$ (right-skewed density), then we can make the transformation, $Q = \frac{p}{1-p}$. This is true for obesity. The Jacobian is $\frac{1}{(1+Q)^2}$ and $0 \leq Q \leq 1$. A similar procedure can be carried out if $0 \geq p \leq \frac{1}{2}$ (left-skewed density). With our assumption that the density is right skewed,

$$\pi(Q|\underset{\sim}{y}) \propto \frac{1}{(1+Q)^2} \prod_{\{i:y_i=1\}} \frac{Q^{W_i}}{1+Q^{W_i}} \prod_{\{i:y_i=0\}} \frac{1}{1+Q^{W_i}} \quad (30)$$
$$= \frac{Q^{\sum_{i=1}^{n} W_i}}{(1+Q)^2} \prod_{i=1}^{n} \frac{1}{1+Q^{W_i}}. \quad (31)$$

Because survey weights $\underset{\sim}{w}$ are large, $\prod_{i=1}^{n} \frac{1}{1+Q^{W_i}} \approx 1$ and $\pi(Q|\underset{\sim}{y}) \approx \frac{Q^{\sum_{i=1}^{n} W_i}}{(1+Q)^2}$.

We will use the grid method to draw Q from (31). Then P can be obtained by re-transformation, $P = \frac{Q}{1+Q}$.

For normalized posterior distributions with original survey weights (method C) or trimmed survey weights (method G), we need this approach.

APPENDIX C: Hierarchical Bayesian Model for Binary Data

We consider a small area model, where there are ℓ areas each comes from a population. (A reviewer requested us to provide a hierarchical model to show generality.) The population size of the i^{th} area is $N_i, i = 1, \ldots, \ell$. Let $y_{ij}, j = 1, \ldots, N_i, i = 1, \ldots, \ell$. We are omitting covariates from this example. Inference is required for the finite population proportions $P_i = \frac{1}{n_i} \sum_{j=1}^{N_i} y_{ij}$. The population model is

$$y_{ij} \mid p_i \overset{ind}{\sim} \text{Bernoulli}(p_i), j = 1, \ldots, N_i,$$
$$p_i \mid \mu, \rho \overset{ind}{\sim} \text{Beta}\left\{\mu\frac{1-\rho}{\rho}, (1-\mu)\frac{1-\rho}{\rho}\right\}, i = 1, \ldots, \ell.$$

See Nandram (2016) for this reparameterization of the beta distribution. Finally, we assume

$$\pi(\mu, \rho) = 1, 0 < \mu, \rho < 1.$$

We assume that this population model is correct.

For the sample model, we adjust the population model using only the observed data. We have a sample of size $n_i, i = 1, \ldots, \ell$, from the i^{th} area. We also have survey weights W_{ij} and adjusted trimmed survey weights w_{ij}^* Then, using the normalized adjusted trimmed survey weights (our preference), we have

$$y_{ij} \mid p_i \overset{ind}{\sim} \text{Bernoulli}\left\{\frac{p_i^{w_{ij}^*}}{p_i^{w_{ij}^*} + (1-p_i)^{w_{ij}^*}}\right\}, j = 1, \ldots, n_i$$
$$p_i \mid \mu, \rho \overset{ind}{\sim} \text{Beta}\left\{\mu(\frac{1-\rho}{\rho}), (1-\mu)(\frac{1-\rho}{\rho})\right\}, i = 1, \ldots, \ell,$$
$$\pi(\mu, \rho) = 1, 0 < \mu, \rho < 1.$$

Let $\underset{\sim}{y}$ denote the sample data. Then, using Bayes' theorem, the joint posterior density is

$$\pi(\underset{\sim}{p}, \mu, \rho \mid \underset{\sim}{y}) \propto$$

$$\prod_{i=1}^{\ell}\prod_{j=1}^{n_i}\frac{p_i^{y_{ij}w_{ij}^*}(1-p_i)^{(1-y_{ij})w_{ij}^*}}{p_i^{w_{ij}^*}+(1-p_i)^{w_{ij}^*}}\prod_{i=1}^{\ell}\frac{p_i^{\mu(\frac{1-\rho}{\rho})-1}(1-p_i)^{(1-\mu)\frac{1-\rho}{\rho}-1}}{B\{\mu(\frac{1-\rho}{\rho}),(1-\mu)(\frac{1-\rho}{\rho})\}},$$

where $0 < p_i < 1, i = 1, \ldots, \ell, 0 < \mu, \rho < 1$. To make inference about the finite population proportions, we only need to sample $p_i, i = 1, \ldots, \ell$, from their marginal posterior density. It is easy to show that if $0 < \sum_{j=1}^{n_i} y_{ij} < n_i, i = 1, \ldots, \ell$ (distinctly), the joint posterior density, $\pi(\underset{\sim}{p}, \mu, \rho \mid \underset{\sim}{y})$ is proper. Also, we can sample the posterior density using a technique similar to Nandram (2016).

Predictive inference is now exactly the same for the simple Bernoulli example as we discussed in this paper using surrogate samples (e.g. Nandram, 2007).

APPENDIX D: A Model for Binary Study Variable with Covariates

Our basic objective is to show how to incorporate covariates for binary data. Therefore, we incorporate the survey weights into the logistic regression model. The Bayesian analysis is currently under study. A reviewer has requested that we consider calibration weights, and this is why we have considered covariates. In our development we are actually using calibration weights, but we are not using covariates. So by simply raking up the original weights so that they sum to the population size, we are essentially using the calibration weights. But calibration makes sense when the total covariate vector is known for the population. It is just a simple optimization step to go from the original survey weights to calibration weights (Haziza & Beaumont, 2017).

We assume logistic regression with p covariates, including an intercept, for the population model,

$$y_i \mid \underset{\sim}{\beta} \overset{ind}{\sim} \text{Bernoulli}\left\{\frac{e^{\underset{\sim}{x}_i'\underset{\sim}{\beta}}}{1+e^{\underset{\sim}{x}_i'\underset{\sim}{\beta}}}\right\}, i = 1, \ldots, N,$$

where N, the population size, and the nonsampled $\underset{\sim}{x}_i$ may not be known, and these can come from an external source.

Then, the normalized density with the adjusted trimmed survey weights, $w_i^*, i = 1, \ldots, n$, is

$$y_i \mid \underset{\sim}{\beta} \overset{ind}{\sim} \text{Bernoulli}\left\{\frac{e^{w_i^*\underset{\sim}{x}_i'\underset{\sim}{\beta}}}{1+e^{w_i^*\underset{\sim}{x}_i'\underset{\sim}{\beta}}}\right\}, i = 1, \ldots, n.$$

Again, the normalized form is more appropriate under the Bayesian paradigm, and in this form the covariates are adjusted to $\underset{\sim}{\tilde{x}}_i = w_i^*\underset{\sim}{x}_i, i = 1, \ldots, n$. Then, using a flat prior on $\underset{\sim}{\beta}$, $\pi(\underset{\sim}{\beta}) = 1$, the joint posterior density is

$$\pi(\underset{\sim}{\beta} \mid \underset{\sim}{y}) \propto \left\{\frac{e^{\sum_{i=1}^n y_i w_i^*\underset{\sim}{x}_i'\underset{\sim}{\beta}}}{\prod_{i=1}^n(1+e^{w_i^*\underset{\sim}{x}_i'\underset{\sim}{\beta}})}\right\}, \ \underset{\sim}{\beta} \in R^p.$$

With a flat prior on $\underset{\sim}{\beta}$, under some mild conditions relating the x_i and y_i, it is well known that if $X = (\underset{\sim}{x}_i')$ is full rank, the joint posterior density, $\pi(\underset{\sim}{\beta} \mid \underset{\sim}{y})$, is proper; see M.-H. Chen, Ibrahim, and Kim (2008) for more details with Jeffreys' prior. Also, with a proper prior on $\underset{\sim}{\beta}$, because the likelihood is bounded, the posterior density will be proper. We can actually use the Gibbs sample (or the Metropolis sampler) to get samples of $\underset{\sim}{\beta}$. Inference about the finite population proportion P is now straight forward. Because we have covariates, it is possible to start with calibration weights, which replace the original weights. These can be adjusted and trimmed survey weights.

The only practical issue that remains is when the nonsampled covariates are unknown, a typical scenario, how to estimate them because they are needed to predict the finite population proportion. If all the covariates are discrete with each having just a few levels, then it is possible. If some variables are continuous, they can be discretized to a few levels. In the BMI data, there are three covariates, which are generally used, and these are all discrete, but age has about 70 levels (age runs from 20 to 89) and race and sex, each has two levels. So there are 280 distinct vectors $\underset{\sim}{x}$ of covariates. If the sample is fairly large (BMI data we use have nearly 2000 observations), it is reasonable to assume that the sample covariates are the only ones in the population, and if some are not observed, they can be structurally eliminated. Using the original survey weights, we have a Horvitz-Thompson estimate of the frequency of each distinct vector covariate.

References

Basu, D. (1971). *An essay on the logical foundations of survey sampling, part i. foundations of statistical inferences, vp godambe and da sprott.* Holt, Rinehart and Winston, Toronto, Canada.

Beaumont, J.-F. (2020). Are probability surveys bound to disappear for the production of official statistics? *Survey Methodology*, *46*(1), 1–28.

Chen, M.-H., Ibrahim, J. G., & Kim, S. (2008). Properties and implementation of Jeffreyss prior in binomial regression models. *Journal of the American Statistical Association*, *103*(484), 1659–1664. https://doi.org/10.1198/016214508000000779

Chen, Q., Elliott, M. R., Haziza, D., Yang, Y., Ghosh, M., Little, R. J., Sedransk, J., & Thompson, M. (2017). Approaches to improving survey-weighted estimates. *Statistical Science*, *32*(2), 227–248. https://doi.org/10.1214/17-STS609

Chen, Y., Li, P., & Wu, C. (2020). Doubly robust inference with nonprobability survey samples. *Journal of the American Statistical Association*, *115*(532), 2011–2021. https://doi.org/10.1080/01621459.2019.1677241

Choi, J. W., & Nandram, B. (2021). Large sample problems. *International Journal of Statistics and Probability*, *10*(2), 81-89.

Choi, J. W., & Nandram, B. (2022). Combining correlated p-values from primary data analyses. *International Journal of Statistics and Probability*, *11*(6), 1-25.

Elliott, M. R., & Valliant, R. (2017). Inference for nonprobability samples. *Statistical Science*, 249–264. https://doi.org/10.1214/16-STS598

Gelman, A. (2007). Struggles with survey weighting and regression modeling. *Statistical Science*, *22*(2), 153–164. https://doi.org/10.1214/088342306000000691

Haziza, D., & Beaumont, J.-F. (2017). Construction of weights in surveys: A review. *Statistical Science*, *32*(2), 206–226. https://doi.org/10.1214/16-STS608

Kim, J. K., Park, S., Chen, Y., & Wu, C. (2018). Combining non-probability and probability survey samples through mass imputation. *arXiv preprint arXiv:1812.10694*. https://doi.org/10.1111/rssa.12696

Kish, L. (1965). *Survey sampling*. New York: Wiley.

Lee, S., & Valliant, R. (2009). Estimation for volunteer panel web surveys using propensity score adjustment and calibration adjustment. *Sociological Methods & Research*, *37*(3), 319–343. https://doi.org/10.1177/0049124108329643

Lohr, S. (2007). Comment: Struggles with survey weighting and regression modeling. *Statistical Science*, *22*(2), 175–178.

Lohr, S. (2009). *Sampling: Design and Analysis*. Nelson Education.

Meng, X.-L. (2018). Statistical paradises and paradoxes in big data (i) law of large populations, big data paradox, and the 2016 us presidential election. *The Annals of Applied Statistics*, *12*(2), 685–726. https://doi.org/10.1214/18-AOAS1161SF

Musal, R. M., Soyer, R., McCabe, C., & Kharroubi, S. A. (2012). Estimating the population utility function: A parametric bayesian approach. *European journal of operational research*, *218*(2), 538–547. https://doi.org/10.1016/j.ejor.2011.11.017

Nandram, B. (2007). Bayesian predictive inference under informative sampling via surrogate samples. *Bayesian statistics and its applications*, 356–374.

Nandram, B. (2016). Bayesian predictive inference of a proportion under a twofold small-area model. *Journal of Official Statistics (JOS)*, *32*(1).

Nandram, B., & Choi, J. W. (2005). Hierarchical Bayesian nonignorable nonresponse regression models for small areas: An application to the NHANES data. *Survey Methodology*, *31*(1), 73–84.

Nandram, B., & 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*(489), 120–135. https://doi.org/10.1198/jasa.2009.ap08443

Nandram, B., Choi, J. W., & Liu, Y. (2021). Integration of nonprobability and probability samples via survey weights. *International Journal of Statistics and Probability*, *10*(6), 1–5.

Nandram, B., & Rao, J. (2021). A bayesian approach for integrating a small probability sample with a non-probability sample. *In JSM Proceedings, Survey Research Methods Section, Alexandria, VA: American Statistical Association*, 1568-1603.

Pfeffermann, D. (1993). The role of sampling weights when modeling survey data. *International Statistical Review/Revue Internationale de Statistique*, 317–337. https://doi.org/10.2307/1403631

Pfeffermann, D., Da Silva Moura, F. A., & Do Nascimento Silva, P. L. (2006). Multi-level modelling under informative sampling. *Biometrika*, *93*(4), 943–959.

Potthoff, R. F., Woodbury, M. A., & Manton, K. G. (1992). Equivalent sample size and equivalent degrees of freedom refinements for inference using survey weights under superpopulation models. *Journal of the American Statistical Association*, *87*(418), 383–396. https://doi.org/10.2307/2290269

Rao, J. (1966). Alternative estimators in pps sampling for multiple characteristics. *Sankhyā: The Indian Journal of Statistics, Series A*, 47–60.

Rao, J. (2020). On making valid inferences by integrating data from surveys and other sources. *Sankhya B*, 1–31.

Rao, J., Hidiroglou, M., Yung, W., & Kovacevic, M. (2010). Role of weights in descriptive and analytical inferences from survey data: An overview. *Journal of the Indian Society of Agricultural Statistics*, *64*, 129–135.

Robbins, M. W., Ghosh-Dastidar, B., & Ramchand, R. (2021). Blending probability and nonprobability samples with applications to a survey of military caregivers. *Journal of Survey Statistics and Methodology*, *9*(5), 1114–1145. https://doi.org/10.1093/jssam/smaa037

Royall, R. M., & Pfeffermann, D. (1982). Balanced samples and robust bayesian inference in finite population sampling. *Biometrika*, *69*(2), 401–409. https://doi.org/10.2307/2335414

Wang, W., Rothschild, D., Goel, S., & Gelman, A. (2015). Forecasting elections with non-representative polls. *International Journal of Forecasting*, *31*(3), 980–991. https://doi.org/10.1016/j.ijforecast.2014.06.001

Copyrights