Discussion of Five Papers at JSM2023 in an Invited Session in Honor of Joe Sedransk

Balgobin Nandram*

Abstract

At JSM2023, I organized an invited session of five speakers on **Contributions to Inference from Survey Samples: In Honor of Joe Sedransk**. I also served as the discussant of these five papers, which were presented by Qixuan Chen, Lu Chen, Glen Meeden, Mary Meyer and Mary Thompson in this order. This paper is a summary of my discussions at the meeting. The first 2.5 minutes was used to say congratulation to Professor Joe Sedransk, and because my time was limited to 15 minutes, I spoke about 2.5 minutes on each paper, and I made a small point about each paper. I highlighted some of Joe's contributions and my collaborations with him.

Key Words: Bayesian surveys, Fay-Herriot model, Nonparametric methods, Order-restricted inference, Scott-Smith model, Selection bias, Small area estimation

1. Introduction

This session is to honor Professor Joe Sedransk on his 85th birthday. Joe has made many contributions to the American Statistical Association (ASA), Survey Sampling and other areas of Statistics. He has served on Advisory Committees to several US government agencies, i.e., the US Census Bureau, Energy Information Administration, Environmental Protection Agency and Social Security. He was the chair of the first two and was a founding member of the Federal Economic Statistics Advisory Committee. He has served as a faculty member at many Statistics Departments in the US, and was chair at the University of Iowa and Case Western Reserve University. He was the Applications and Case Studies Editor of JASA and founding co-editor of the Journal of Survey Statistics and Methodology. He has graduated twenty doctoral students who work in University, government and industry positions.

As the title, "Contributions to Inference from Survey Samples: In Honor of Joseph Sedransk," states, this session has highlighted Joe's contributions in survey sampling on topics such as selection and non-response bias, predictive inference, order-restricted inference, small area estimation, cluster sampling and data integration (combining two or more samples, one could be a non-probability sample). These are very hot and current areas of research in Survey Sampling. Joe's contributions have primarily been in the use of hierarchical Bayesian models for multi-stage surveys, small area estimation and cluster sampling coupled with selection and non-response bias issues. At 85 years Joe is formally retired, but still an active researcher. Just recently, he has worked

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

on a sole-authored paper on Bayesian informative sampling, and has many collaborators (different from his own students) on other topics in Survey Sampling as well. Our session highlights theories, methods and applications of these topics in Survey Sampling.

Two of the speakers are current collaborators, and one of them is a long-term friend of Joe. One speaker is the academic grand daughter of Joe, one of my PhD students. Another presenter is a collaborator of one of Joe's collaborators. I am a past student of Joe and a professor at Worcester Polytechnic Institute since my graduation in 1989 at the University of Iowa. Therefore, there are five speakers and a discussant, Balgobin Nandram.

At the beginning Guofen Yan, another student of Joe, was supposed to chair the session. She did all the ground work for the session, but she had to withdraw at the last moment, and had to return to Virginia one day before the session at the JSM2023 in Toronto. I am grateful to her for helping out this way. Therefore, Lu Chen, one of my firmer students chaired all the talks, except her talk that was chaired by one of my other students, Jiani Yin. The session went very well, and the attendance was good even though it was the last day of JSM2023.

The five talks in order of the presentations are listed here.

1. Yao, Y., Ogden, R. T., Zeng, C. and Chen, Q. (2023). Bivariate Hierarchical Bayesian Model for Combining Summary Measures and their Uncertainties from Multiple Sources, *Annals of Applied Statistics*, 17 (2), 1782-1800 (presented by Qixuan Chen, Columbia University, New York).

2. Chen, L. and Nandram, B. (2023), Model for State-Level Cash Rental Rates, *Joint Statistical Meetings*, pg. 1-12 (presented by Lu Chen, National Institute of Statistical Science).

3. Meeden, G. and Qureshi, N. (2023), Adaptive Cluster Sampling as Domain Estimation, *Joint Statistical Meetings*, pg. 1-28 (presented by Glen Meeden, University of Minnesota).

4. Thompson, M. E. (2023), The Bayesian-Frequentist Dialogue in Survey Methodology, *Joint Statistical Meetings*, pg 1-21 (presented by Mary Thompson, University of Waterloo)

5. Liao, X, Xu, X. and Meyer, M. C. (2023), Csurvey: Implementing Order Constraints in Survey Data Analysis, *Journal of Statistical Software* (to appear), pg. 1-20 (presented by Mary Meyer, Colorado State University).

I will now discuss the five papers in the order in which they were presented. On the way, I will highlight a couple of Joe's papers and some of my collaborations with him. I will describe how my collaborations with Joe form the basis of my own research.

This paper has six sections, including the current one. Each paper, under discussion, is described in a separate section and there is a section with some conclusions.

2. Qixuan Chen: Fay-Herriot with both parameters random

Qixuan gave three applications in her paper, but did not present the one on traffic safety that uses small area estimation. A sensible modeling of sample variances should include the area sample sizes (or effective sample sizes), but this was not part of the modeling. There are area level covariates as well.

I have an enormous amount of experience of the Fay-Herriot model at the National Agricultural Statistics Service (NASS), USDA. In fact, my collaborators and I at NASS have written many journal articles over the past five years. For example, see Erciulescu, Cruze and Nandram (2019), Nandram, Erciulescu and Cruze (2019), Chen, Nandram and Cruze (2022), Nandram, Cruze and Ericulescu (2023) and Nandram (2023), just to mention a few selected ones.

2.1 Bivariate model

For the application on traffic safety, Qixuan got

$$\rho_1 : .09\ (-.70, 0.80);\ \ \rho_2 = .03\ (-.65, .62)$$

It appears that ρ_1 and ρ_2 are the same, basically zero correlation. So I decided to look at this more deeply.

For $i = 1, \dots, n$, define $t_i = \log(s_i)$ and $\phi_i = \log(\sigma_i)$ Then, we have the very complicated model,

$$\begin{bmatrix} y_i \\ t_i \end{bmatrix} \overset{ind}{\sim} \text{Normal}\left(\begin{bmatrix} \theta_i \\ \phi_i \end{bmatrix}, \begin{bmatrix} e^{2\phi_i} & \rho_1 e^{\phi_i} \sigma_{s_i} \\ \rho_1 e^{\phi_i} \sigma_{s_i} & \sigma_{s_i}^2 \end{bmatrix} \right)$$

$$\begin{bmatrix} \theta_i \\ \phi_i \end{bmatrix} \overset{ind}{\sim} \text{Normal}\left(\begin{bmatrix} \mu_\theta \\ \mu_\sigma \end{bmatrix}, \begin{bmatrix} r_\theta^2 & \rho_2 r_\theta r_\sigma \\ \rho_2 r_\theta r_\sigma & r_\sigma^2 \end{bmatrix} \right)$$

To proceed, replacing ϕ_i in the covariance matrix by μ_σ, and integrating out (θ_i, ϕ_i),

$$\begin{bmatrix} y_i \\ t_i \end{bmatrix} \overset{ind}{\sim} \text{Normal}\left(\begin{bmatrix} \mu_\theta \\ \mu_\sigma \end{bmatrix}, \begin{bmatrix} e^{2\mu_\sigma} + r_\theta^2 & \rho_1 e^{\mu_\sigma} \sigma_{s_i} + \rho_2 r_\theta r_\sigma \\ \rho_1 e^{\mu_\sigma} \sigma_{s_i} + \rho_2 r_\theta r_\sigma & \sigma_{s_i}^2 + r_\sigma^2 \end{bmatrix} \right)$$

That is, ρ_1 and ρ_2 may not be identifiable. So I presented a simpler solution.

2.2 Alternative Solution

We have

$$\hat{\theta}_i \mid \theta_i, \hat{\sigma}_i^2 \overset{ind}{\sim} \text{Normal}(\theta_i, \hat{\sigma}_i^2), i = 1, \dots, \ell.$$

Define $\hat{\kappa}_i = \frac{\hat{\sigma}_i^2}{GM}$, where GM is the geometric mean of the $\hat{\sigma}_i^2$. Define $a_o = \min(\hat{\kappa}_i)$ and $b_o = \max(\hat{\kappa}_i)$, and we will assume that a_o and b_o are known or (a_o, b_o) or can otherwise be specified. Now, define $\hat{\psi}_i = \frac{\hat{\kappa}_i - a_o}{b_o - a_o}, i = 1, \dots, \ell$.

Then, our new assumption is

$$\hat{\theta}_i \mid \hat{\psi}_i, \theta_i, \sigma^2 \overset{ind}{\sim} \text{Normal}\left\{ \theta_i, \frac{\sigma^2}{(1 - \hat{\psi}_i) a_o + \hat{\psi}_i b_o} \right\}$$

and

$$\hat{\psi}_i \overset{ind}{\sim} \text{Beta}\{\mu\frac{\gamma}{1-\gamma}, (1-\mu)\frac{\gamma}{1-\gamma}\}, i = 1, \dots, \ell.$$

The rest of the Bayesian Fay-Herriot model remains the same,

$$\theta_i \mid \underset{\sim}{\beta}, \rho, \sigma^2 \overset{ind}{\sim} \text{Normal}(\underset{\sim}{x}_i' \underset{\sim}{\beta}, \frac{\rho}{1-\rho}\sigma^2),$$

$$\underset{\sim}{\beta} \sim \text{Normal}(\underset{\sim}{\beta}_o, \sigma^2 \Sigma_o),$$

$$\pi(\mu, \gamma, \rho, \sigma^2) \propto \frac{1}{\sigma^2}, 0 < \mu, \gamma, \rho < 1.$$

2.3 Proportions

Finally, I point out that estimation of proportions is a bit tricky. Let us look at the sampling process,

$$\hat{\theta}_i \mid \theta_i, \hat{\sigma}_i^2 \overset{ind}{\sim} \text{Normal}(\theta_i, \hat{\sigma}_i^2), i = 1, \dots, \ell.$$

We want to input correlations in the pairs $(\hat{\theta}_i, \hat{\sigma}_i^2)$. Let us consider the case of proportions,

$$\hat{p}_i \overset{ind}{\sim} \text{Normal}\{p_i, \frac{\hat{p}_i(1-\hat{p}_i)}{n_i}\}, i = 1, \dots, \ell.$$

Define $\hat{Q}_i = \hat{p}_i(1-\hat{p}_i)$, a quadratic function, symmetric about $\frac{1}{2}$ with a maximum of $\frac{1}{4}$ at $\frac{1}{2}$. So

$$\text{Cor}(\hat{p}_i, \hat{Q}_i) \approx 0.$$

If we know more, things can be different. If $0 < \hat{p}_i < \frac{1}{2}$, then the correlation will be strongly positive, and if $\frac{1}{2} < \hat{p}_i < 1$, the correlation will be strongly negative. It is necessary to include such constraints in the modeling of proportions.

3. Lu Chen: Mixture, discounting and robustness

Because I am a collaborator of Lu Chen, I decided to present three extensions that we can work on.

3.1 Mixture model (historical data)

$$f(y_i \mid \underset{\sim}{\beta}, p, q, \rho, \gamma) = (1-p-q)\text{Normal}_{y_i}(\underset{\sim}{x}_i' \underset{\sim}{\beta}, \frac{\rho\gamma\sigma^2}{a})$$

$$+p\text{Normal}_{y_i}(\underset{\sim}{x}_i' \underset{\sim}{\beta}, \frac{\gamma\sigma^2}{a}) + q\text{Normal}_{y_i}(\underset{\sim}{x}_i' \underset{\sim}{\beta}, \frac{\sigma^2}{a}), i = 1, \dots, n,$$

$p > q$ and $0 < p, q, p+q, \rho, \gamma < 1$; priors on all parameters (mild and severe outliers).

3.2 Different a_i

There may be "births" and "deaths" for counties. The discounting factor is only weakly identified so one needs to be careful; see Ibrahim and Chen (2000). Thus, we use a logit model,

$$a_i = \frac{e^{\gamma_0+\gamma_1 \log(r_i)}}{1+e^{\gamma_0+\gamma_1 \log(r_i)}}, i = 1,\ldots,n,$$

and

$$a_i = \frac{\phi_0 r_i^{\frac{\phi_1}{1-\phi_1}}}{1-\phi_0+\phi_0 r_i^{\frac{\phi_1}{1-\phi_1}}}, i = 1,\ldots,n,$$

where r_i are the number reports.

3.3 Clustering (Ishwaran and James, 2001)

Assume

$$\underset{\sim}{\phi}_t \overset{ind}{\sim} \text{Normal}\{\underset{\sim}{\mu}, \delta^2(R-\gamma W)^{-1}\}, t = 1,2,$$

$$\mu_i \overset{iid}{\sim} \sum_{s=1}^{\ell_o} p_s \text{Normal}(z_s, \kappa_1^2), \ell_o \le \ell, i = 1,\ldots,\ell, \ell_o \le \ell,$$

$$z_s \overset{ind}{\sim} \text{Normal}(0, \kappa_2^2), s = 1,\ldots,\ell_o,$$

$$p_1 = \nu_1, p_2 = \nu_2(1-\nu_1),\ldots, p_{\ell_o} = \prod_{s=1}^{\ell_o-1}(1-\nu_s),$$

and

$$\nu_s \overset{ind}{\sim} \text{Beta}\{1-\delta_1, \frac{1-\delta_2}{\delta_2}+(s-1)\delta_1\}, s = 1,\ldots,\ell_o.$$

Remark: For NASS problems on county estimates, it is now my strong believe that speculative states should be modeled separately from non-speculative states. But for coherence, the two models should be connected into a single one. I now believe that these CAR models are not appropriate. States, which are geographically closed, may not be so with respect to the study variable. A speculative state may be a geographical neighbor of a non-speculative state, but they may be very different in terms of the study variable. So it is sensible to separate the states into speculative and non-speculative states; the two clusters are very different.

4. Glen Meeden: Adaptive cluster sampling as domain estimation

Glen introduced an approach different from Thompson (1990) for domain estimation. There are three issues (**sparseness**, clustering, informative sampling). The clustering and the informative nature of the design were not used in Glen's method. However, his approach is nonparametric Bayesian, and the key idea is the Polya posterior.

It is pertinent to review the Polya posterior. Have n values, $y_1, \ldots, y_n$ from a finite population of size, N, and we want to sample posterior density, $\pi(\underset{\sim}{y}_{ns} \mid \underset{\sim}{y}_s)$, to provide a Bayesian predictive inference, when nothing is known about the population (sparse and clustered). For $k = 1, \ldots, N-n$,

$$y_{n+k} = \begin{cases} y_1 & k_1 \\ y_2 & k_2 \\ . & . \\ . & . \\ . & . \\ y_n & k_n, \end{cases}$$

where we assume the counts, $k_i = 1$, at start (all values distinct, but this is not necessary). The idea is to sample one of $y_1, \ldots, y_n$ and replace that value by two values which are same as the one drawn. Continue this process until there are N values, including the first n values.

The population consists of N squares (grid cells), and $y_i, i = 1, \ldots, N$, denote the counts in the squares most of them are zeros (sparseness). Inference is needed for $T_D = \sum_{i \in D} y_i$, where D is domain of non-empty grid cells. In Glen's approach θ (unknown) is the cardinality of D.

Glen has a simple random sample of n squares, but actually $m \geq n$ is the sample size in the ADC sample (Population Thompson: $n = 10$, $m = 45$, apparently m is not revealed by Glen); Glen took a random sample of 40 squares to perform the simulation.

He considered two estimators, ADC (double counting) and BAY (sparseness, quasi-Bayes, use Polya posterior).

Table 1: Glen's Simulation of Thompson population for inference about T_D

	Est	Rbias	Abserr	Lowbd	Len	Freqcov
ADC	211	0.112	87.9	118	341	0.921
BAY	157	-.173	45.7	118	274	1.000

NOTE: $N = 400, T_D = 190, 3$ networks, $\frac{\theta}{N} = 21/400 = 0.0525$. Simulated 1000 simple random samples of size 40.

Remark: BAY (quasi-Bayes) underestimates variability. For the simulation study, for ADC the range is $(118, 341)$ and for BAY the range is $(118, 274)$ much shorter.

4.1 Glen's Bayesian approach

Have x positive squares of the m squares (samples). Therefore, we have $\theta - x$ positive squares out of $N - m$ (nonsamples). Then, assume a hypergeometric distribution for x? That is,

$$f(x \mid \theta) = H_x(\theta), x \leq \theta \leq \min(N, x + N - m).$$

Glen actually used this hypergeometric distribution (private communication), but it was not stated.

Table 2: Glen's prior distribution for θ

θ	k_1	k_2	$\ldots$	k_G
$\pi(\theta)$	p_1	p_2	$\ldots$	p_G

NOTE: $\theta/N \sim \text{Beta}(1,90)$ discretized.

Glen constructed a prior based on simulation (he said) and the Beta distribution, discretized, for θ/N, where sparseness (many zeros) of the cells is taken into consideration, but not the informative nature of the design (selection probabilities)?

Posterior distribution of θ

$$Pr(\theta = k_j \mid x) = q_j = \frac{p_j H_x(k_j)}{\sum_{j'=1}^{G} p_{j'} H_x(k_{j'})}, j = 1, \ldots, G.$$

Bayesian Analysis is straight forward.

a. Draw θ from its posterior.

b. Have $y_1, \ldots, y_x$ (seen); need $y_{x+1}, \ldots, y_\theta$ (useen) and use Polya posterior to get them. Compute $T = \sum_{i=1}^{x} y_i + \sum_{i=x+1}^{\theta} y_i$.

c. Repeat (a) and (b) to estimate the posterior distribution of T.

4.2 Another Bayesian approach

Rapley and Welch (Bayesian Analysis, 2008) provided a Bayesian model with selection bias.

Let N denote the number of points; M the number of grid cells (squares); x the number of non-empty grid cells and p the number of nonempty networks. We have $p \leq x \leq M$, and an initial simple random sample (without replacement), $S = (i_1, \ldots, i_m)$.

For i^{th} network, let y_i denote number of non-empty grid cells (squares) and N_i the number of points. Note that $x = \sum_{i=1}^{p} y_i$.

Hierarchical Bayesian model

We have the following assumptions,

$$[x] \text{ (truncated Binomial)},$$

$$[p \mid x] \text{ (truncated Binomial)},$$

$$[\underset{\sim}{y} \mid p, x] \text{ (truncated multinomial)},$$

$$Pr(S \mid x, p, \underset{\sim}{y}) \text{ (PPS sampling of non-empty networks)},$$

$$[\underset{\sim}{N} \mid \underset{\sim}{y}, p] = \prod_{i=1}^{p} [N_i \mid y_i] \text{ (truncated Poisson)}.$$

Posterior inference is required for

$$N = \sum_{i=1}^{p} N_i.$$

Remark: Selection bias is taken care of in the fourth line of the model.

5. Mary Meyer: Domain estimation with order restrictions

Mary provided an algorithm to do domain estimation. Apparently, she used only the domain averages with the appropriate survey weights.

5.1 Software

The population is partitioned into D domains, and inference is required for the finite population mean.

Have survey data with survey weights, W, and study variable y.

Mary finds the order restricted estimates, $\hat{\underset{\sim}{\theta}}$, such that

$$\min_{\underset{\sim}{\theta}} (\underset{\sim}{\bar{y}} - \underset{\sim}{\theta})' W (\underset{\sim}{\bar{y}} - \underset{\sim}{\theta}), A\hat{\underset{\sim}{\theta}} \geq 0,$$

where A specifies any linear constraint.

Comments

1. The R software, **Csurvey**, works for all kinds of linear constraints;
2. Nonparametric procedure (least square estimators);
3. Covariates can be included;
4. Confidence intervals and tests for empty domains or domains with small number of observations;

5. Unfortunately, it appears that within domain variability is lost.

$$\min_{\underset{\sim}{\theta}} \sum_{i=1}^{D} \sum_{j=1}^{n_i} W_{ij}(y_{ij} - \theta_i)^2, A\hat{\underset{\sim}{\theta}} \geq 0.$$

Note that

$$\sum_{i=1}^{D} \sum_{j=1}^{n_i} W_{ij}(y_{ij} - \theta_i)^2 = \sum_{i=1}^{D} \sum_{j=1}^{n_i} W_{ij}(y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^{D} W_{i\cdot}(\bar{y}_i - \theta_i)^2,$$

where $\bar{y}_i = \frac{\sum_{j=1}^{n_i} W_{ij} y_{ij}}{\sum_{j=1}^{n_i} W_{ij}}$ and $W_{i\cdot} = \sum_{j=1}^{n_i} W_{ij}$. The first term on the right-hand side disappears because it is not a function of $\underset{\sim}{\theta}$. Therefore, the variability is underestimated but the point estimators should be fine.

5.2 Bayesian example (normal means)

Have survey data $(w_{ij}, y_{ij}), j = 1, \ldots, n_i, i = 1, \ldots, D$ (w_{ij} are adjusted survey weights)

$$y_{ij} \mid \theta_i, \overset{ind}{\sim} \text{Normal}\{\theta_i, \frac{\sigma^2}{w_{ij}}\},$$

$$\pi(\underset{\sim}{\theta}, \sigma^2) = \frac{1}{\sigma^2}, \theta_1 \leq \ldots \leq \theta_D.$$

Using Bayes' theorem, the joint posterior density is

$$\pi(\underset{\sim}{\theta}, \sigma^2 \mid \underset{\sim}{y}) \propto (\frac{1}{\sigma^2})^{n/2+1} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{D} \sum_{j=1}^{n_i} w_{ij}(y_{ij} - \bar{y}_i)^2} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{D} w_{i\cdot} \sum_{j=1}^{n_i} (\bar{y}_i - \theta_i)^2}, \theta_1 \leq \ldots \leq \theta_D,$$

where $\bar{y}_i = \frac{\sum_{j=1}^{n_i} w_{ij} y_{ij}}{\sum_{j=1}^{n_i} w_{ij}}$ and $w_{i\cdot} = \sum_{j=1}^{n_i} w_{ij}, i = 1, \ldots, D$.

1. The joint posterior density of $(\underset{\sim}{\theta}, \sigma^2)$ is proper.
2. It is easy to use the Gibbs sampler to draw samples from the the CPD σ^2, inverse gamma, and the CPDs of the θ_i, truncated normal distributions.
3. If we assume domain d is missing (i.e., no data for μ_d using only domain d), we can simply eliminate the part of the posterior without domain data, but keep the support.
4. In a similar manner, we can do any other distribution (e.g., Bernoulli for binary data); Euclidean distance is not appropriate though.

5.3 Comments on order restrictions

A fundamental paper on all my work on Bayesian order restricted inference is Sedransk, Monahan and Chiu (1985). I actually presented it in a seminar when I was a PhD student at the University of Iowa.

There are many other papers based on this work. Nandram, Sedransk and Smith (1997) wrote a paper on the aging of fish, which were stratified by length. There is a unimodal order on cell proportions within each stratum and there is also a unimodal across the modes. It is interesting that uncertainty of the modal positions was added to the multinomial-Dirichlet model. Sometime later I gave a talk on this topic at the University of Iowa, and one member in the audience told me that we should call it the umbrella ordering. We have written several papers later (e.g., Nandram and Peiris, 2018), and I have recently advised a PhD dissertation (Xinyu, 2021) on this topic, and Joe was one of the Dissertation Committee members.

As Mary mentioned shape restrictions, I finally noted in my discussions that problems with logconcave and unimodal restrictions are interesting. Logconcave densities are a special sub-class of the wider class of unimodal densities; logconcave densities are more docile. These provide more flexible (i.e., nonparametric) models, and order restrictions are used to study these.

6. Mary Thompson: Bayesian-frequentist dialogue

Mary raised the very interesting topic of the Bayesian-frequentist dialogue. I want to point out that superpopulation theory holds the key to this dialogue. In this school, the finite population is a random sample from a parametric model (the superpopulation) and a probability sample is taken from this finite population. Unlike design base analysis, this permits all the finite population values to be treated as random variables. Therefore, this also permits using parametric (and nonparametric) models, a break through for complex applications. For example, small area estimation needs models; see Nandram and Choi (2010) and Molina, Nandram and Rao (2014). This was initiated by many papers such as Ericson (1969), Scott and Smith (1969) and Royal (1970). In the era of Markov chain Monte Carlo methods, the hierarchical Bayesian models are the workhorse for complex applications. But I differ with Mary a bit because she wrote “the probability sampling design unrelated to y might or might not play a role” in the analysis. It is now well-known that informative (selection is related to y) probability sample design does play a crucial role in Bayesian analysis, especially when data integration of a probability sample and a non-probability sample, an emerging area in survey sampling and statistical methods, is needed; see Nandram and Rao (2021, 2023) and Nandram, Choi and Liu (2021), which use surrogate sample (Nandram 2007) to do Bayesian predictive inference.

For Mary, I addressed the issue of equivalent sample sizes and normalization constants when survey weights are used. We note that survey weights are attributes of the entire ensemble, not just an individual unit (e.g., Gelman 2007). Survey weights are different from regular covariates. Besides there is hardly going to be any simple relation between the study variable and the survey weights.

6.1 Equivalent sample size

We have $(W_i, Y_i), i = 1, \ldots, n$, where W_i are original survey weights and Y_i are responses. We assume (normality is not required)

$$\underset{\sim}{Y} \sim \text{Normal}\left\{\theta \underset{\sim}{j}, \{(1-\rho)I + \rho J\}\sigma^2\right\},$$

where $0 < \rho < 1$ is an intra-class correlation.

Consider estimation of θ using the Horvitz-Thompson estimator (optimal or sub-optimal),

$$\hat{\theta} = \frac{\sum_{i=1}^n W_i Y_i}{\sum_{i=1}^n W_i}.$$

Under the model,

$$\hat{\theta} \sim \text{Normal}\left\{\theta, \{(1-\rho)\frac{W'W}{(\underset{\sim}{j}'\underset{\sim}{W})^2} + \rho\}\sigma^2\right\}.$$

Here $\hat{\theta}$ is both design unbiased and model unbiased.

Then, the effective sample size (Kish, 1965) is

$$n_{\rho e} = \frac{n_e}{(1-\rho) + \rho n_e} = \frac{n_e}{1 + (n_e - 1)\rho},$$

independent of σ^2, where $n_e = \frac{(\sum_{i=1}^n W_i)^2}{\sum_{i=1}^n W_i^2}$. Note that $1 \le n_e \le n$; see also Potthoff, Woodbury and Manton (1992).

Remark: When survey weights (not design weights) are used, they should be included in a normalized composite likelihood not as covariates. If they are used as covariates, they will increase precision (small or large) but when they are used in a composite likelihood, they will decrease precision (mitigate bias) and this is what is required.

6.2 Normalization constants

We have the data, $(W_i, y_i), i = 1, \ldots, n$, where W_i are original survey weights and y_i are study variable. If we assume that y_i are correlated; we use composite likelihood with independence, of course. Let $g(y \mid \underset{\sim}{\theta})$ be a density for the y_i.

We use adjusted weights, not original weights, $w_i = n_e \frac{W_i}{\sum_{i=1}^n W_i}$. Then

$$\prod_{i=1}^n g(y_i \mid \underset{\sim}{\theta})^{w_i}$$

is not a density. The correct density is

$$f(\underset{\sim}{y} \mid \underset{\sim}{\theta}) = \frac{\prod_{i=1}^n g(y_i \mid \underset{\sim}{\theta})^{w_i}}{\prod_{i=1}^n \{\int g(y_i \mid \underset{\sim}{\theta})^{w_i} dy_i\}}$$

With respect to Bayesian-frequentist dialogue, there are two issues.

a. The normalization constant should be used because if the denominator is a function of $\underset{\sim}{\theta}$, the likelihood without it will be incorrect. Both non-Bayesians and Bayesians drop the normalization constant and this is clearly incorrect. At least Bayesians should keep the normalization constant. But, in general, it is difficult to work with the normalization constant.

b. For all practical applications, there are more 2-tuples than 1-tuples, so that when a pairwise composite likelihood is used, there will be an artificial gain in precision. This is true for 3-tuples and so on. (For example, for a sample of size n, $\binom{n}{2} > \binom{n}{1}, n \geq 4$.) See Rao, Verret and Hidiroglou (2014), Ribatet, Cooley and Davison (2012) and Varrin, Reid and Firth (2011) for discussion on composite likelihoods.

6.3 Comments on Scott and Smith (1969)

Scott-Smith model was introduced to analyze data from two-stage cluster sampling, but it can be used just as well for small area estimation with different inferential objectives. This is the reason why Battese, Harter and Fuller (BHF, 1988) did not recognize this model; the Scott-Smith model does not accommodate covariates and it is a special case of the BHF model. Indeed, my PhD dissertation, advised by Joe Sedransk, at the University of Iowa started with the Scott-Smith model. I am very pleased that Mary mentioned the Scott-Smith model. Bayesian analyses of the BHF model are given by Molina, Nandram and Rao (2014) and Toto and Nandram (2010).

I have written many papers in which the Scott-Smith model. Nandram and Sedransk (1993) on the analysis of binary data from cluster sampling has been the basic starting point of many of my papers. Nandram, Toto and Choi (2011) is a prominent example, where a full Bayesian analysis is discussed with computations. Lockwood (2023) is a PhD dissertation that uses the Scott-Smith model extensively. She used the Scott-Smith model to express no relation between the study variable and the covariates. Similar ideas are used in current work of Nandram on non-probability samples.

7. Concluding Remarks

I am very honored to be a student of Professor Joe Sedransk. In 1984, he practically brought me and my family (Minwantie and Nankumarie) to the United States to study under him at SUNY, Albany. After a year, I moved with him to the University of Iowa, where he became chair in 1985, and I earned my PhD in 1989. Joe is an outstanding professor and researcher. I owed at lot to my collaboration with him. Joe is very personable and a great friend.

It was a pleasure for me to organize this invited session on his 85^{th} birthday. Although, it was extremely hard work, it was fun to work for the discussion of the five papers. It took about one month's of my summer time (four months) to prepare for this auspicious occasion, and it is an event that I will always, always, remember.

Acknowledgments

Balgobin Nandram used his Professional Development Account to fully cover his cost to attend the JSM2023 meeting in Toronto Convention Center. Guofen Yan treated the invited speakers with lunch at the JSM one day before the session. I had all other lunches with my PhD students (Lu Chen, Yuan Yu and Jiani Yin), past and present, at the JSM2023.

REFERENCES

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

Chen, X. (2021), *Constraint Bayesian Inference for Count Data from Small Areas*, PhD Dissertation, Department of Mathematical Sciences, Worcester Polytechnic Institute, pg. 1-122.

Chen, L., Nandram, B., and Cruze, N. B. (2022), Hierarchical Bayesian model with inequality constraints for us county estimates, *Journal of Official Statistics*, 38 (3), 709–732.

Erciulescu, A. L., Cruze, N. B. and Nandram, B. (2019), Model-based county level crop estimates incorporating auxiliary sources of information, *Journal of the Royal Statistical Society*, 182 (1), 283-303.

Ericson, W. A. (1969), Subjective Bayesian models in sampling finite populations, *Journal of the Royal Statistical Society*, Series B 31 (2), 195-233.

Ibrahim, J. G. and Chen, M-H. (2000), Power Prior Distributions for Regression Models, *Statistical Science*, 15 (1), 46-60.

Ishwaran, H. and James, L. F. (2001), Gibbs Sampling Methods for Stick-breaking Priors, *Journal of the American Statistical Association*, 96, 161-173.

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

Gelman, A. (2007), Struggles with Survey Weighting and Regression Modeling, *Statistical Science*, 22 (2), 153-164.

Lockwood, A. (2023), *Bayesian Predictive Inference for a Study Variable Without Specifying a Link to the Covariates*, PhD Dissertation, Department of Mathematical Sciences, Worcester Polytechnic Institute, pg. 1-110.

Molina, I., Nandram, B. and Rao, J. N. K., (2014), Small Area Estimation of General Parameters with Application to Poverty indicators: A hierarchical Bayes Approach, *The Annals of Applied Statistics*, 8 (2), 852-885.

Nandram, B. (2023), Overcoming Challenges Associated with Early Bayesian State Estimation of Planted Acres in the United States *Special Proceedings: Society of Statistics, Computing and Applications*, 25, 51-78

Nandram, B. (2007), Bayesian Predictive Inference Under Informative Sampling Via Surrogate Samples, In *Bayesian Statistics and Its Applications, Eds. S.K. Upadhyay, Umesh Singh and Dipak K. Dey*, Anamaya, New Delhi, Chapter 25, 356-374.

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

Nandram, B., Choi, J. W. and Liu, Y. (2021), Integration of Nonprobability and Probability Samples via Survey Weights, *International Journal of Statistics and Probability*, 10 (6) (in press).

Nandram, B. and Rao, J. N. K (2021), A Bayesian Approach for Integrating a Small Probability Sample with a Nonprobability Sample, *Proceedings of the American Statistical Association, Survey Research Methods Section*, 1568-1603.

Nandram, B. and Rao, J. N. K (2023), Bayesian Predictive Inference when Integrating a Nonprobability Sample and a Probability Sample. *arXiv:2305.08997v1 [Stat.ME], 15 May 2023*, pg. 1-35.

Nandram, B. and Sedransk, J. (1993), Bayesian Predictive Inference for a Finite Population Proportion: Two-Stage Cluster Sampling, *Journal of the Royal Statistical Society*, Series B, 55 (2), 399-408.

Nandram, B., Sedransk, J.and Smith, S. J. (1997), Order-Restricted Bayesian Estimation of the Age Composition of a Population of Atlantic Cod, *Journal of the American Statistical Association*, 92 (437), 33-40.

Nandram, B. and Peiris, T. B. (2018), Bayesian Analysis of a ROC Curve for Categorical Data Using a Skew-Binormal Model, *Statistics and Its Interface*, 11 (2), 369-384.

Nandram, B., Cruze, N. and Erciulescu, A. (2023), Bayesian Small Area Models under Inequality Constraints with Benchmarking and Double Shrinkage *Survey Methodology* 49 (2) (in press).

Nandram, B., Erciulescu, A. L. and NB Cruze, N. B. (2019), Bayesian benchmarking of the Fay-Herriot model using random deletion Survey Methodology 45 (2), 365-391

Potthoff, R. F., Woodbury, M. A. and Manton, K. G. (1992), "Equivalent Sample Size" and "Equivalent Degrees of Freedom" Refinements for Inference Using Survey Weights Under Superpopulation Models, *Journal of the American Statistical Association*, 87 (418), 383-396.

Rao, J. N. K., Verret, F. and Hidiroglou, M. A. (2014), A Weighted Composite Likelihood Approach to Inference for Two-level Models from Survey Data, *Survey Methodology*, 39 (2), 263-282.

Rapley, V. E. and Welsh, A. H. (2008), Model-based inferences from adaptive cluster sampling, *Bayesian Analysis*, 3 (4), 717-736.

Ribatet, M., Cooley, D. and Davison, A. C. (2012), Bayesian Inference from Composite Likelihoods, With an Application to Spatial Extremes, *Statistica Sinica*, 22, 813-845.

Varin, C., Reid, N. and Firth, D. (2011), An Overview of Composite Likelihood Methods, *Statistica Sinica*, 21 (1), 5-42.

Royal, R. M. (1970), On finite population sampling theory under certain linear regression models, *Biometrika*, 57, 377-387.

Scott, A. and Smith, T. M. F. (1969), Estimation in Multi-Stage Surveys, *Journal of the American Statistical Association*, 64 (327), 830-840.

Sedransk, J., Monahan, J. and Chiu, H. Y. (1985), Bayesian Estimation of Finite Population Parameters in Categorical Data Models Incorporating Order Restrictions, *Journal of the Royal Statistical Society*, Series B, 47 (3), 519-527.

Thompson, S. K. (1990), Adaptive Cluster Sampling, *Journal of the American Statistical Association* 85 (412), 1050-1059.

Toto, M. C. S. and Nandram, B. (2010), "A Bayesian Predictive Inference for Small Area Means Incorporating Covariates and Sampling Weights," *Journal of Statistical Planning and Inference*, 140, 2963-2979.