Pseudo-Bayesian Inference for Complex Survey Data

Matt Williams[1] Terrance Savitsky[2]

[1]National Center for Science and Engineering Statistics
National Science Foundation
mrwillia@nsf.gov

[2]Office of Survey Methods Research
Bureau of Labor Statistics
Savitsky.Terrance@bls.gov

University of Michigan
April 8, 2020

Thank you!

- Terrance Savitsky for being a great collaborator and mentor.
- Brady West and Jennifer Sinibaldi for making this connection.
- Jill Esau for orchestrating.
- You all for sharing your time today!

Bio

1. Work
   - 9 years as mathematical statistical for federal government: USDA, HHS, NSF
   - Sample design, weighting, imputation, estimation, disclosure limitation (production and methods development)
2. Consulting
   - International surveys for agricultural production (USAID) and vaccination knowledge, attitudes, and behaviors (UNICEF)
3. Research (ORCID: 0000-0001-8894-1240)
   - Constrained Optimization for Survey Applications (weight adjustment, benchmarking model estimates)
   - Applying Bayesian inference methods to data from complex surveys.

Outline

1. Informative Sampling (Savitsky and Toth, 2016)
2. Theory and Examples
   - Consistency (Williams and Savitsky, 2020)
   - Uncertainty Quantification (Williams and Savitsky, in press)
3. Implementation Details
   - Model Fitting
   - Variance Estimation
4. Related and Current Works

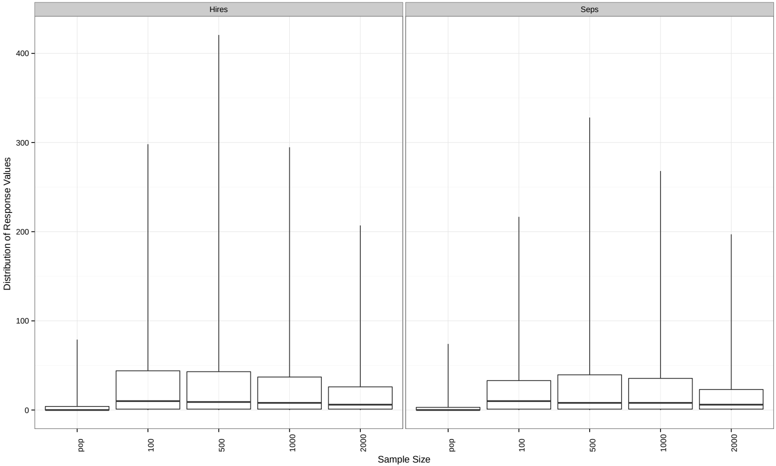
Outline

1. Informative Sampling (Savitsky and Toth, 2016)
2. Theory and Examples
   - Consistency (Williams and Savitsky, 2020)
   - Uncertainty Quantification (Williams and Savitsky, in press)
3. Implementation Details
   - Model Fitting
   - Variance Estimation
4. Related and Current Works

Example: Informative Sampling

- Take a sample from U.S. population of business establishments
- Single stage, fixed-size, pps sampling design
- $\mathbf{y} =$ (e.g., Hires, Separations)
- Size variable is total employment, x
- $y \not\perp x$.
- $B = 500$ Monte Carlo samples at each of $\mathbf{n}_\nu = (100, 500, 1500, 2500)$ establishments

Distributions of **y** in Informative Samples



Population Inference from Informative Samples

- **Goal**: perform inference about a finite population generated from an unknown model, $\mathbb{P}_{\theta_0}(\mathbf{y})$.
- **Data**: from under a complex sampling design distribution, $\mathbb{P}_{\nu}(\boldsymbol{\delta})$
  - Probabilities of inclusion $\pi_i = Pr(\delta_i = 1|\mathbf{y})$ are often associated with the variable of interest (purposefully)
  - Sampling designs are "informative": the balance of information in the sample $\neq$ balance in the population.
- **Biased Estimation**: estimate $\mathbb{P}_{\theta_0}(\mathbf{y})$ without accounting for $\mathbb{P}_{\nu}(\boldsymbol{\delta})$.
  - Use inverse probability weights $w_i = 1/\pi_i$ to mitigate bias.
- **Incorrect Uncertainty Quantification**:
  - Failure to account for dependence induced by $\mathbb{P}_{\nu}(\boldsymbol{\delta})$ leads to standard errors and confidence intervals that are the wrong size.

Outline

1 Informative Sampling (Savitsky and Toth, 2016)

2 Theory and Examples
- Consistency (Williams and Savitsky, 2020)
- Uncertainty Quantification (Williams and Savitsky, in press)

3 Implementation Details
- Model Fitting
- Variance Estimation

4 Related and Current Works

Why Bayes?

- Allows more complex, non-parametric (semi-supervised) models
- Use hierarchical modeling to capture rich dependence in data
- Have small sample properties from posterior distribution
- Full uncertainty quantification
- Gold standard for imputation

Pseudo Posterior

- Pseudo posterior $\propto$ Pseudo Likelihood $\times$ Prior

$$p^{\pi}\left(\boldsymbol{\theta}|\mathbf{y},\tilde{\mathbf{w}}\right) \propto \left[\prod_{i=1}^{n} p\left(y_i|\boldsymbol{\theta}\right)^{\tilde{w}_i}\right] p\left(\boldsymbol{\theta}\right)$$

$$w_i := \frac{1}{\pi_i}$$

$$\tilde{w}_i = \frac{w_i}{\frac{\sum w_i}{n}}, \;\; i = 1, \ldots, n$$

Similar Posterior Geometry

$$\mathcal{N}_P\left(\mathbf{y}_i|\boldsymbol{\mu}_i, \boldsymbol{\Phi}^{-1}\right)^{w_i} \propto \mathcal{N}_P\left(\mathbf{y}_i|\boldsymbol{\mu}_i, [w_i\boldsymbol{\Phi}]^{-1}\right)$$

- normalize weights, $\sum_{i=1}^{n} w_i = n$, to scale posterior

Outline

1 Informative Sampling (Savitsky and Toth, 2016)

2 Theory and Examples
- Consistency (Williams and Savitsky, 2020)
- Uncertainty Quantification (Williams and Savitsky, in press)

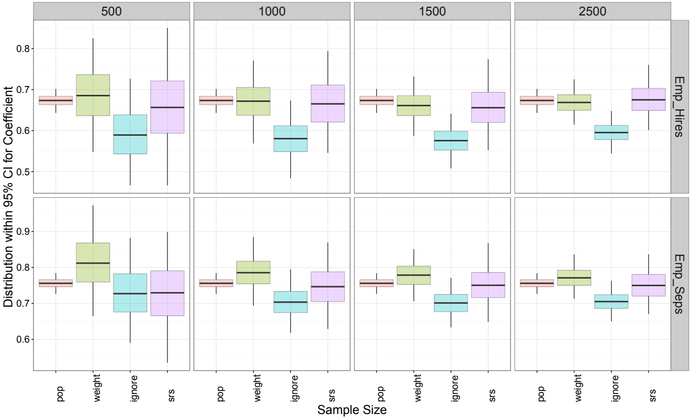
3 Implementation Details
- Model Fitting
- Variance Estimation

4 Related and Current Works

Pseudo Posterior Contraction - Count Data

$$y_{id} \overset{\text{ind}}{\sim} \text{Pois}\left(\exp\left(\psi_{id}\right)\right)$$

$$\underset{}{\overset{N\times D}{\Psi}} \sim \overset{N\times P}{\mathbf{X}}\ \overset{P\times D}{\mathbf{B}} + \mathcal{N}_{N\times D}\left(\mathbb{I}_N, \overset{D\times D}{\Lambda^{-1}}\right)$$



Frequentist Consistency of a (Pseudo) Posterior

- Estimated distribution $p^{\pi}\left(\boldsymbol{\theta}|\mathbf{y},\tilde{\mathbf{w}}\right)$ collapses around generating parameter θ_0 with increasing population N_ν and sample n_ν sizes.
  - Evaluated with respect to joint distribution of population generation $\mathbb{P}_{\theta_0}(\mathbf{y})$ and the sample inclusion indicators $\mathbb{P}_\nu(\boldsymbol{\delta})$.
- Conditions on the model $\mathbb{P}_{\theta_0}(\mathbf{y})$ (standard)
  - Complexity of the model limited by sample size
  - Prior distribution not too restrictive (e.g. point mass)
- Conditions on the sampling design $\mathbb{P}_\nu(\boldsymbol{\delta})$ (new)
  - Every unit in population has non-zero probability of inclusion $\Longrightarrow$ finite weights
  - Dependence restricted to countable blocks of bounded size $\Longrightarrow$ arbitrary dependence within clusters, but approximate independence between clusters.

Simulation Example: Three-Stage Sample

Area (PPS), Household (Systematic, sorting by Size), Individual (PPS)

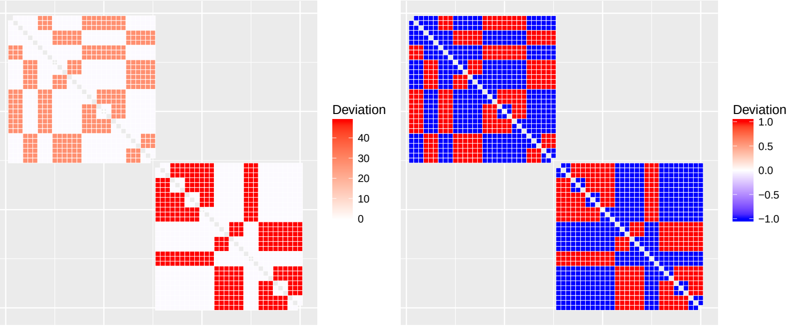


Figure: Factorization matrix $(\pi_{ij}/(\pi_i \pi_j) - 1)$ for two PSU's. Magnitude (left) and Sign (right). Systematic Sampling $(\pi_{ij} = 0)$. Clustering and PPS sampling $(\pi_{ij} > \pi_i \pi_j)$. Independent first stage sample $(\pi_{ij} = \pi_i \pi_j)$

Simulation Examples: Logistic Regression

- $$y_i \mid \mu_i \overset{\text{ind}}{\sim} \text{Bern}\left(F_l(\mu_i)\right), \; i = 1, \ldots, N$$
- $$\boldsymbol{\mu} = -1.88 + 1.0\boldsymbol{x}_1 + 0.5\boldsymbol{x}_2$$
- The $\mathbf{x}_1$ and $\mathbf{x}_2$ distributions are $\mathcal{N}(0,1)$ and $\mathcal{E}(r = 1/5)$ with rate r
- Size measure used for sample selection is $\tilde{\boldsymbol{x}}_2 = \boldsymbol{x}_2 - \min(\boldsymbol{x}_2) + 1$, *but* neither $\tilde{\boldsymbol{x}}_2$ or $\boldsymbol{x}_2$ are available to the analyst.
- Intercept chosen so median of $\boldsymbol{\mu} \approx 0 \rightarrow$ median of $F_l(\boldsymbol{\mu}) \approx 0.5$.

Simulation Example: Three-Stage Sample (Cont)

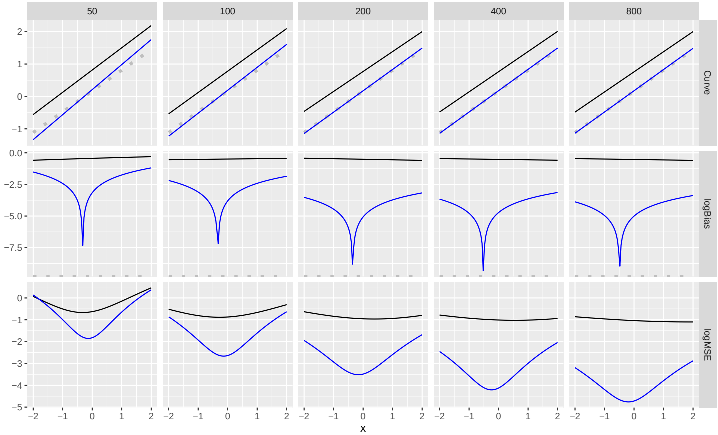


Figure: The marginal estimate of $\mu = f(x_1)$. **population curve**, sample with **equal weights**, and **inverse probability weights**. Top to bottom: estimated curve, log of BIAS, log MSE. Left to right: sample size (50 to 800).

Outline

1 Informative Sampling (Savitsky and Toth, 2016)

2 Theory and Examples
- Consistency (Williams and Savitsky, 2020)
- Uncertainty Quantification (Williams and Savitsky, in press)

3 Implementation Details
- Model Fitting
- Variance Estimation

4 Related and Current Works

Asymptotic Variances

- Let $\ell_\theta(\boldsymbol{y}) = \log p(\boldsymbol{y}|\theta)$.
- Rely on the variance and expected curvature of the score function $\dot{\ell}_{\theta_0} = \frac{\partial \ell}{\partial \theta}|_{\theta=\theta_0}$ with $\ddot{\ell}_{\theta_0} = \frac{\partial^2 \ell}{\partial^2 \theta}|_{\theta=\theta_0}$
- $H_{\theta_0} = -\frac{1}{N_\nu} \sum_{i \in U_\nu} \mathbb{E}_{P_{\theta_0}} \ddot{\ell}_{\theta_0}(\mathbf{y}_{\nu i})$
- $J_{\theta_0} = \frac{1}{N_\nu} \sum_{i \in U_\nu} \mathbb{E}_{P_{\theta_0}} \dot{\ell}_{\theta_0}(\mathbf{y}_{\nu i}) \dot{\ell}_{\theta_0}(\mathbf{y}_{\nu i})^T$
- Under correctly specified models:
  - $H_{\theta_0} = J_{\theta_0}$ (Bartlett's second identity)
  - Posterior variance $N_\nu \mathbb{V}(\theta|\boldsymbol{y}) = H_{\theta_0}^{-1}$ same as variance of MLE (Bernstein-von Mises)

Scaling and Warping of Pseudo MLE

- Mispecified (under-specified) full joint sampling distribution $\mathbb{P}_\nu(\boldsymbol{\delta})$.
- Failure of Bartlett's Second Identity for composite likelihood
- Asymptotic Covariance: $H_{\theta_0}^{-1} J_{\theta_0}^{\pi} H_{\theta_0}^{-1}$
- Simple Random Sampling: $J_{\theta_0}^{\pi} = J_{\theta_0}$
- Unequal weighting: $J_{\theta_0}^{\pi} \geq J_{\theta_0}$

$$J_{\theta_0}^{\pi} = J_{\theta_0} + \frac{1}{N_\nu} \sum_{i=1}^{N_\nu} \mathbb{E}_{P_{\theta_0}} \left\{ \left[\frac{1}{\pi_{\nu i}} - 1 \right] \dot{\ell}_{\theta_0}(\mathbf{y}_{\nu i}) \dot{\ell}_{\theta_0}(\mathbf{y}_{\nu i})^T \right\}$$

  - Shape of asymptotic distribution warped by unequal weighting $\propto \frac{1}{\pi_{\nu i}}$
- If less efficient (cluster) sampling design : $J_{\theta_0}^{\pi} \geq J_{\theta_0}$
- If more efficient (stratified) sampling design : $J_{\theta_0}^{\pi} \leq J_{\theta_0}$

NCSES

BLS

Asymptotic Covariances Different

- Pseudo MLE: $H_{\theta_0}^{-1} J_{\theta_0}^{\pi} H_{\theta_0}^{-1}$ (Robust)
- Pseudo Posterior: $H_{\theta_0}^{-1}$ (Model-based)
- The un-adjusted pseudo-posterior will give the wrong coverage of uncertainty regions.

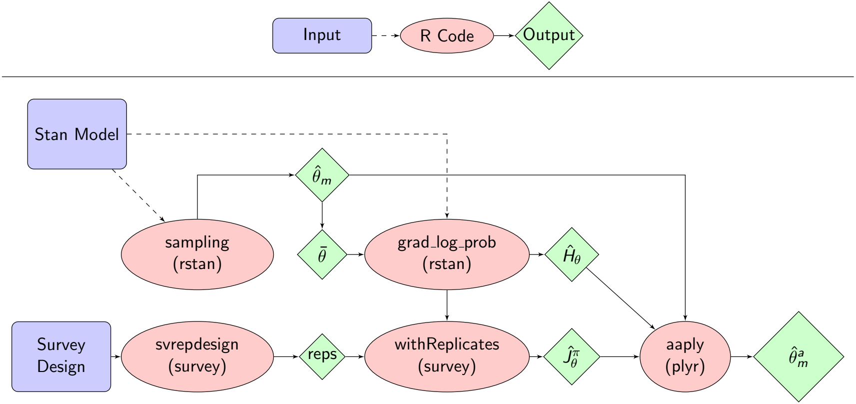
Adjust Pseudo Posterior draws to Sandwich

- $\hat{\theta}_m \equiv$ sample pseudo posterior for $m = 1, \ldots, M$ draws with mean $\bar{\theta}$
- $\hat{\theta}^a_m = \left(\hat{\theta}_m - \bar{\theta}\right) R_2^{-1} R_1 + \bar{\theta}$
- where $R_1' R_1 = H_{\theta_0}^{-1} J_{\theta_0}^{\pi} H_{\theta_0}^{-1}$
- $R_2' R_2 = H_{\theta_0}^{-1}$

Adjustment Procedure

- Procedure to compute adjustment, $\hat{\theta}_m^a$
  - Input $\hat{\theta}_m$ drawn from single run of MCMC
  - Re-sample data under sampling design
  - Draw PSUs (clusters) without replacement
  - Compute $\hat{H}_{\theta_0}$ and $\hat{J}_{\theta_0}^{\pi}$
- Expectations with respect to P_{θ_0}, P_{ν}
  - Let $\mathbb{P}_{N_\nu}^{\pi} = \frac{1}{N_\nu} \sum_{i=1}^{N_\nu} \frac{\delta_{\nu i}}{\pi_{\nu i}} \delta\left(\mathbf{y}_{\nu i}\right)$
  - $J_{\theta_0}^{\pi} = \text{Var}_{P_{\theta_0}, P_\nu}\left[\mathbb{P}_{N_\nu}^{\pi} \dot{\ell}_{\theta_0}\right]$
  - $H_{\theta_0}^{\pi} = -\mathbb{E}_{P_{\theta_0}, P_\nu}\left[\mathbb{P}_{N_\nu}^{\pi} \ddot{\ell}_{\theta_0}\right] = H_{\theta_0}$

R Code Schematic



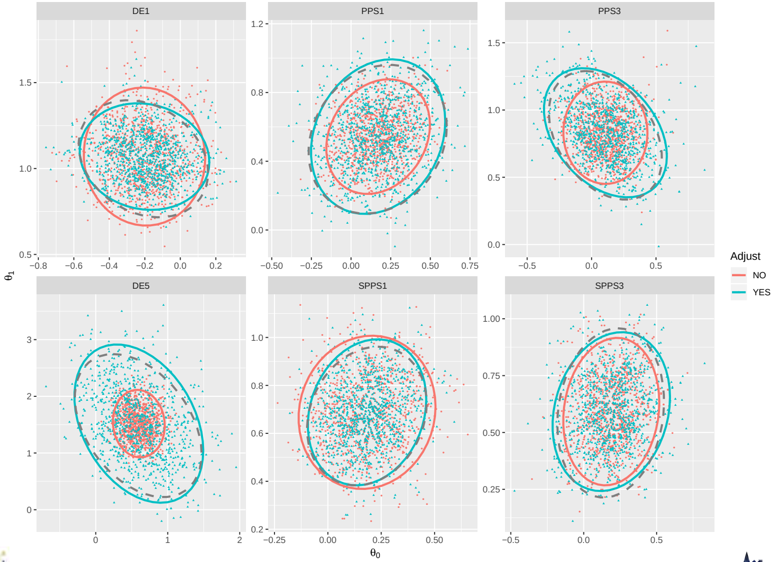
Simulation Study - Generate Population

- Binary Response: $\boldsymbol{y} \in \{0, 1\}$
- Two predictors: $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$
- Cluster designs: cluster level effect $\boldsymbol{z}_2 \rightarrow$ within cluster correlation
- Size measure used for sample selection is $\tilde{\boldsymbol{x}}_2 = \boldsymbol{x}_2 - \min(\boldsymbol{x}_2) + 1$, *but* neither $\tilde{\boldsymbol{x}}_2$ or $\boldsymbol{x}_2$ are available to the analyst.
- Intercept chosen so median of $\boldsymbol{\mu} \approx 0 \rightarrow$ median of $F_l(\boldsymbol{\mu}) \approx 0.5$. About 50/50 for 0's, 1's.

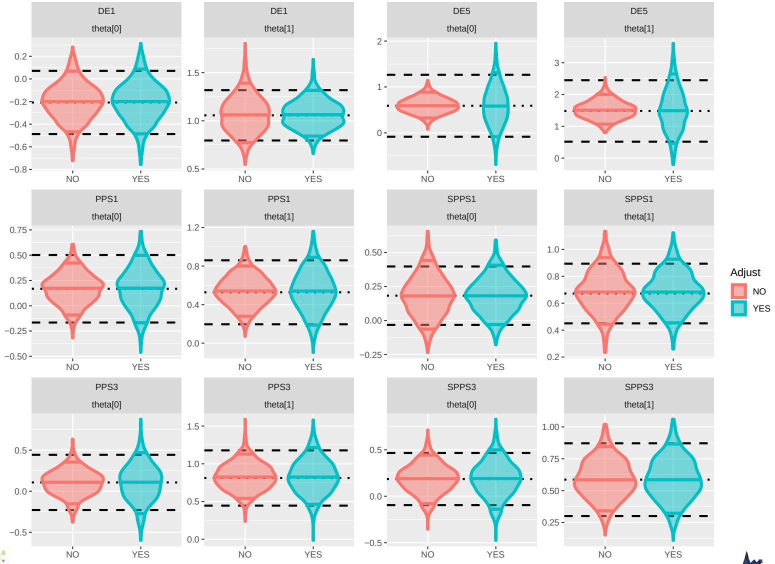
Simulation Study - Six Sample Designs

- Weak vs. Strong within cluster dependence: DE1 and DE5 equally-weighted. DE5 replicates units within PSU.
- One Stage PPS design with/out strata: PPS1 single stage unequally-weighted. SPPS1 is stratified
- Three-Stage PPS design with/out strata: PPS3 is 3-stage. SPPS3 is stratified. Sample 40 of 200 PSUs, 5 of 10 HHs/PSU, 1 of 3 units/HH
- Sample size $n = 200$.

Joint Distribution



Marginal Distributions



Coverage Results for 90% Target Nominal Coverage

Scenario	Marginal θ_0		Marginal θ_1		Joint θ_0, θ_1		Width θ_0		Width θ_1	
	$\hat{\theta}_m$	$\hat{\theta}^a_m$	$\hat{\theta}_m$	$\hat{\theta}^a_m$	$\hat{\theta}_m$	$\hat{\theta}^a_m$	$\hat{\theta}_m$	$\hat{\theta}^a_m$	$\hat{\theta}_m$	$\hat{\theta}^a_m$
DE1	0.89	0.86	0.89	0.90	0.93	0.87	0.52	0.51	0.64	0.63
DE5	0.43	0.81	0.56	0.94	0.32	0.88	0.55	1.24	0.70	1.60
PPS1	0.77	0.88	0.83	0.91	0.74	0.93	0.50	0.69	0.55	0.70
SPPS1	0.91	0.84	0.96	0.96	0.99	0.88	0.49	0.41	0.54	0.55
PPS3	0.74	0.91	0.79	0.87	0.75	0.86	0.51	0.75	0.57	0.75
SPPS3	0.77	0.95	0.80	0.87	0.74	0.87	0.51	0.73	0.56	0.71

Outline

1. Informative Sampling (Savitsky and Toth, 2016)
2. Theory and Examples
   - Consistency (Williams and Savitsky, 2020)
   - Uncertainty Quantification (Williams and Savitsky, in press)
3. Implementation Details
   - Model Fitting
   - Variance Estimation
4. Related and Current Works

Model Fitting Via Stan

- Stan is a platform for statistical modeling and computation (Stan Development Team, 2016)
  - Users specify log density functions
  - Stan provides MCMC sampling, variational inference, or maximum likelihood optimization
  - Stan interfaces with several languages, including R (Rstan)
    - Requires Rtools, for compiling of C++ code.
- Two examples using Stan
  - survey weighted logistic regression (Williams and Savitsky, 2020)
  - survey weighted quantile regression with penalized splines (Williams and Savitsky, 2018)

Outline

1 Informative Sampling (Savitsky and Toth, 2016)

2 Theory and Examples
- Consistency (Williams and Savitsky, 2020)
- Uncertainty Quantification (Williams and Savitsky, in press)

3 Implementation Details
- Model Fitting
- Variance Estimation

4 Related and Current Works

Variance Estimation

- The de-facto approach:
  - approximate sampling independence of the primary sampling units (Heeringa et al., 2010).
  - within-cluster dependence treated as nuisance
- Two common methods:
  - Taylor linearization and replication based methods.
  - A variety of implementations are available (Binder, 1996; Rao et al., 1992).

Taylor Linearization

Let y_{ij} and w_{ij} be the observed data for individual i in cluster j of the sample. Assume the parameter θ is a vector of dimension d with population model value θ_0.

1. Approximate an estimate $\hat{\theta}$, or a 'residual' $(\hat{\theta} - \theta_0)$, as a weighted sum: $\hat{\theta} \approx \sum_{i,j} w_{ij} z_{ij}(\theta)$ where z_{ij} is a function evaluated at the current values of y_{ij}, and $\hat{\theta}$ (e.g. $z_i(\hat{\theta}) = H_{\theta_0}^{-1} \dot{\ell}_{\hat{\theta}}(\mathbf{y}_i)$).
2. Compute the weighted components for each cluster (e.g., primary sampling units (PSUs)): $\hat{\theta}_j = \sum_i w_{ij} z_{ij}(\theta)$.
3. Compute the variance between clusters: $\widehat{Var(\hat{\theta})} = \frac{1}{J-d} \sum_{j=1}^{J} (\hat{\theta} - \hat{\theta}_j)(\hat{\theta} - \hat{\theta}_j)^T$
4. For stratified designs, compute $\hat{\theta}_s$ and $\widehat{Var(\hat{\theta}_s)}$ within strata and sum $\widehat{Var(\hat{\theta})} = \sum_s \widehat{Var(\hat{\theta}_s)}$.

Replication

Let y_{ij} and w_{ij} be the observed data for individual i in cluster j of the sample. Assume the parameter θ is a vector of dimension d with population model value θ_0.

1. Through randomization (bootstrap), leave-one-out (jackknife), or orthogonal contrasts (balanced repeated replicates), create a set of K replicate weights $(w_i)_k$ for all $i \in S$ and for every $k = 1, \ldots, K$.
2. Each set of weights has a modified value (usually 0) for a subset of clusters, and typically has a weight adjustment to the other clusters to compensate: $\sum_{i \in S} (w_i)_k = \sum_{i \in S} w_i$ for every k.
3. Estimate $\hat{\theta}_k$ for each replicate $k \in 1, \ldots, K$.
4. Compute the variance between replicates: $\widehat{Var}(\hat{\theta}) = \frac{1}{K-d} \sum_{k=1}^{K} (\hat{\theta} - \hat{\theta}_k)(\hat{\theta} - \hat{\theta}_k)^T$.
5. For stratified designs, generate replicates such that each strata is represented in every replicate.

Challenges

There are two notable trade-offs associated with these methods:

- Taylor linearization: value $\hat{\theta}$ computed once then used in a plug in for $z_i(\theta)$.
  - Replication methods: estimate $\hat{\theta}_k$ computed K times.
  - Sizable differences in computational effort
- Replication methods: no derivatives are needed.
  - Taylor linearization: requires the calculation of a gradient to derive the analytical form of the first order approximation $z_i(\theta)$.
  - This poses significant analytical challenges for all but the simplest models.

Some Improvements

- Abstraction of Derivatives (less analytic work for Taylor Linearization)
  - Recent advances in algorithmic differentiation (Margossian, 2018), allows us to specify the model as a log density but only treat the gradient in the abstract without specifying it analytically.
  - Implemented in Stan and Rstan (Carpenter, 2015; Stan Development Team, 2016)
- Hybrid Approach or Taylor Linearization for replicate designs (less computation for Replication approaches)
  - Survey package (Lumley, 2016) to calculate replication variance of gradient $\dot{\ell}_\theta$
  - Use plug in for θ, only estimate once

$$(\hat{\psi} - \psi_0) = H_{\theta_0}(\hat{\theta} - \theta_0) \approx \sum_{i \in S} w_i \dot{\ell}_{\hat{\theta}}(\mathbf{y}_i) = \sum_{i \in S} w_i z_i(\hat{\theta}),$$

with $\text{Var}_{P_{\theta_0}, P_\nu}(\hat{\psi} - \psi_0) = J^{\pi}_{\theta_0}$.

NCSES BLS

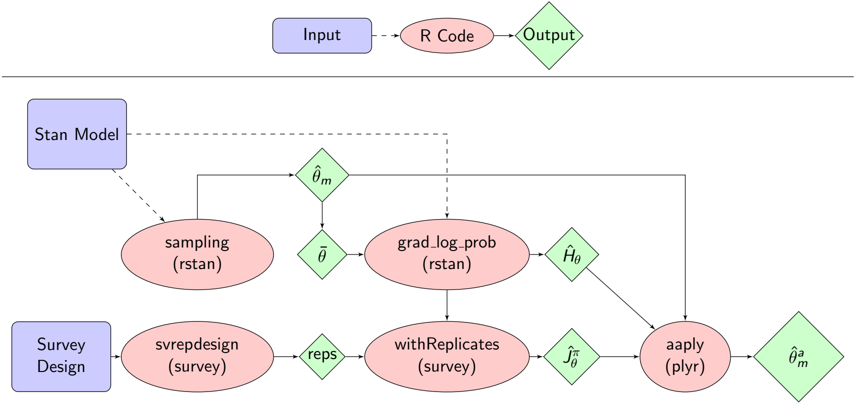
Example: Design Effect for Survey-Weighted Bayes

- Pseudo posterior $\propto$ Pseudo Likelihood $\times$ Prior

$$p^{\pi}\left(\boldsymbol{\theta}|\mathbf{y},\tilde{\mathbf{w}}\right) \propto \left[\prod_{i=1}^{n} p\left(y_i|\boldsymbol{\theta}\right)^{\tilde{w}_i}\right] p\left(\boldsymbol{\theta}\right)$$

- Variances Differ:
  - Weighted MLE: $H_{\theta_0}^{-1} J_{\theta_0}^{\pi} H_{\theta_0}^{-1}$ (Robust)
  - Weighted Posterior: $H_{\theta_0}^{-1}$ (Model-Based)
- Adjust for Design Effect: $R_2^{-1} R_1$
  - $\hat{\theta}_m \equiv$ sample pseudo posterior for $m = 1, \ldots, M$ draws with mean $\bar{\theta}$
  - $\hat{\theta}_m^a = \left(\hat{\theta}_m - \bar{\theta}\right) R_2^{-1} R_1 + \bar{\theta}$
  - where $R_1' R_1 = H_{\theta_0}^{-1} J_{\theta_0}^{\pi} H_{\theta_0}^{-1}$
  - $R_2' R_2 = H_{\theta_0}^{-1}$

R Code Schematic



Outline

1 Informative Sampling (Savitsky and Toth, 2016)

2 Theory and Examples
- Consistency (Williams and Savitsky, 2020)
- Uncertainty Quantification (Williams and Savitsky, in press)

3 Implementation Details
- Model Fitting
- Variance Estimation

4 Related and Current Works

Related Papers

- Consistency of the Pseudo-Posterior (Savitsky and Toth, 2016)
- Extension to multistage surveys (Williams and Savitsky, 2020)
- Extension to pairwise weights and outcomes (Williams and Savitsky, 2018)
- Extension to Divide and Conquer computational methods (Savitsky and Srivastava, 2018)
- Correction of asymptotic coverage (Williams and Savitsky, in press)
- Joint modeling of Outcome and Weights (León-Novelo and Savitsky, 2019)

Current Work

1. Collaboration with State Department on International Polls
   - BigSurv 2020
   - Multinomial response - election polls
2. Mixed Models for Survey Data
   - Invited Session at JSM 2020
   - Savitsky and Williams (2019)
3. Pseudo-Posterior for Differential Privacy
   - Invited Session at JSM 2020
   - Savitsky et al. (2019)

References I

Binder, D. A. (1996), 'Linearization methods for single phase and two-phase samples: a cookbook approach', *Survey Methodology* **22**, 17–22.

Carpenter, B. (2015), 'Stan: A probabilistic programming language', *Journal of Statistical Software* .

Heeringa, S. G., West, B. T. and Berglund, P. A. (2010), *Applied Survey Data Analysis*, Chapman and Hall/CRC.

León-Novelo, L. G. and Savitsky, T. D. (2019), 'Fully bayesian estimation under informative sampling', *Electron. J. Statist.* **13**(1), 1608–1645.
URL: *https://doi.org/10.1214/19-EJS1538*

Lumley, T. (2016), 'survey: analysis of complex survey samples'. R package version 3.32.

Margossian, C. C. (2018), 'A review of automatic differentiation and its efficient implementation', *CoRR* **abs/1811.05031**.
URL: *http://arxiv.org/abs/1811.05031*

Rao, J. N. K., Wu, C. F. J. and Yue, K. (1992), 'Some recent work on resampling methods for complex surveys', *Survey Methodology* **18**, 209–217.

Savitsky, T. D. and Srivastava, S. (2018), 'Scalable bayes under informative sampling', *Scandinavian Journal of Statistics* **45**(3), 534–556. 10.1111/sjos.12312.
URL: *http://dx.doi.org/10.1111/sjos.12312*

References II

Savitsky, T. D. and Toth, D. (2016), 'Bayesian Estimation Under Informative Sampling', *Electronic Journal of Statistics* **10**(1), 1677–1708.

Savitsky, T. D. and Williams, M. R. (2019), 'Bayesian Mixed Effects Model Estimation under Informative Sampling', *arXiv e-prints* p. arXiv:1904.07680.

Savitsky, T. D., Williams, M. R. and Hu, J. (2019), 'Bayesian pseudo posterior mechanism under differential privacy', *arXiv:1909.11796* .

Stan Development Team (2016), 'RStan: the R interface to Stan'. R package version 2.14.1.
URL: *http://mc-stan.org/*

Williams, M. R. and Savitsky, T. D. (2018), 'Bayesian pairwise estimation under dependent informative sampling', *Electron. J. Statist.* **12**(1), 1631–1661.

Williams, M. R. and Savitsky, T. D. (2020), 'Bayesian estimation under informative sampling with unattenuated dependence', *Bayesian Anal.* **15**(1), 57–77.
URL: *https://doi.org/10.1214/18-BA1143*

Williams, M. R. and Savitsky, T. D. (in press), 'Uncertainty Estimation for Pseudo-Bayesian Inference Under Complex Sampling', *International Statistical Review* .
URL: *https://doi.org/10.1111/insr.12376*

Bonus Slides

- Stan syntax examples
- Quantile Regression Example

Stan: Files

R file (.R)

```
library(rstan)
# compile stan code
mod  = stan_model('wt_logistic.stan')
#sample stan model, given data, other inputs
sampling(object = mod, data = ...)
```

Stan file (.stan)

```
functions{ }
data{ }
parameters{ }
transformed parameters{ }
model{ }
```

Stan File: survey weighted logistic regression

```
functions{
real wt_bin_lpmf(int[] y, vector mu, vector weights, int n){
    real check_term;
    check_term  = 0.0;
    for( i in 1:n )
    {
check_term    = check_term +
weights[i] * bernoulli_logit_lpmf(y[i] | mu[i]);
    }
    return check_term;
  }}


model{
  /*improper prior on theta in (-inf,inf)*/
  /* directly update the log-probability for sampling */
  target          += wt_bin_lpmf(y | mu, weights, n);
}
```

Stan File: survey weighted quantile regression with splines

```
functions{
real penalize_spline_lpdf(vector theta, matrix Q,
real tau_theta, int num_bases, int degree) {
  return 0.5 * ( (num_bases-degree) * log(tau_theta) -
   tau_theta * quad_form(Q, theta) );  }
real rho_p(real p, real u){
        return .5 * (fabs(u) + (2*p - 1)*u); }
real ald_lpdf(vector y, vector mu, vector weights, real tau, real p, int n){
    real w_tot;
    real log_terms;
    real check_term;
    w_tot       = sum( weights );
    log_terms   = w_tot * (log(tau) + log(p) + log(1-p));
    check_term  = 0.0;
    for( i in 1:n )
    {
      check_term    = check_term + weights[i] * rho_p( p, (y[i]-mu[i]) );
    }
    check_term = tau * check_term;
    return log_terms - check_term; }}
```

Stan File: survey weighted quantile regression with splines

```
model{
  tau_theta       ~ gamma( 1.0, 1.0 );
  tau             ~ gamma( 1.0, 1.0 );
  theta           ~ penalize_spline(Q, tau_theta, num_knots+degree, degree);
  /* directly update the log-probability for sampling */
  target          += ald_lpdf(y | mu, weights, tau, p, n);
}
```

Example: Sampling and Analyzing Spouse Pairs

Let δ_i and δ_j be indicators that individuals i and j are in the sample. Then the joint indicator $\delta_{ij} = \delta_i \delta_j$.

- Marginal weight $w_i = \delta_i / P\{\delta_i = 1\}$
- Pairwise weight $\tilde{w}_i = \sum_{i \neq j \in D} \left(\delta_{ij} / P\{\delta_{ij} = 1\}\right) / (N_D - 1)$
- For spouses, $N_D = 2$, so 'multiplicity' $(N_D - 1) = 1$.
- For marginal models (anyone with a spouse), use w_i
- For conditional models (both spouses in the sample), use $\tilde{w}_i$

Comparing Conditional Behaviors of Spouses by Age

2014 National Survey on Drug Use and Health

- Median alchohol use (days in past month)
- By Age
- By Use of Spouse
  - solid : spouse ≥ 1
  - dash : spouse $= 0$
- Compare Weights
  - equal, marginal, pairwise