An improved approach to account for complex sampling designs in Bayesian analyses

Anne Cohen [1], Trivellore Raghunathan [1,2], Michael Elliott [1,2]

[1]Survey Methodology and Data Science, University of Michigan

[2]Department of Biostatistics, University of Michigan

May 15, 2026

Bayesian analysis

Bayesian modeling has gained popularity for its many advantages:

- Stabilize estimates by sharing information across multiple levels
- Account for missing, censored, or sparse data
- Incorporate prior information and uncertainty at all levels of modeling[1]
- Enables complex modeling where inference can be difficult to obtain using frequentist approaches

[1] Little, 2022.

Accounting for sampling designs in Bayesian analyses

- Complicated to account for sampling design in Bayesian methods
- Ignoring design can lead to biased estimation or underestimation of variance
- Some methods exist but have limitations[2]
  - Post-stratification methods often limited to simple summary statistics
  - Inclusion of design variables in model limits flexibility and increases computational complexity
  - Pseudo-likelihood approaches, in particular Williams and Savitsky, 2021, can be computationally complex and difficult to implement

[2] Wu and Stephenson, 2024.

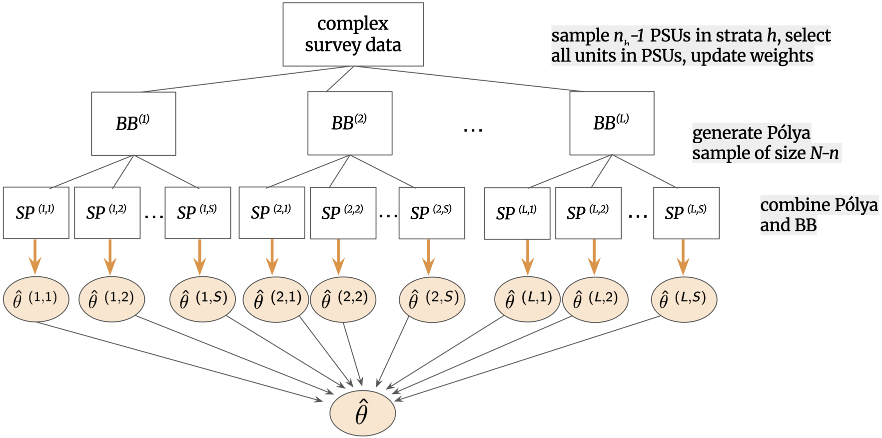
Finite population Bayesian bootstrap (FPBB)

- Method to generate “de-complexed” synthetic populations
- Treat unobserved population data as missing and generate synthetic populations using design information
- Run model on each synthetic population and combine results
    - Computationally inefficient, especially for complex Bayesian models[3]

Goal: Reduce computational complexity while maintaining proper inferential properties

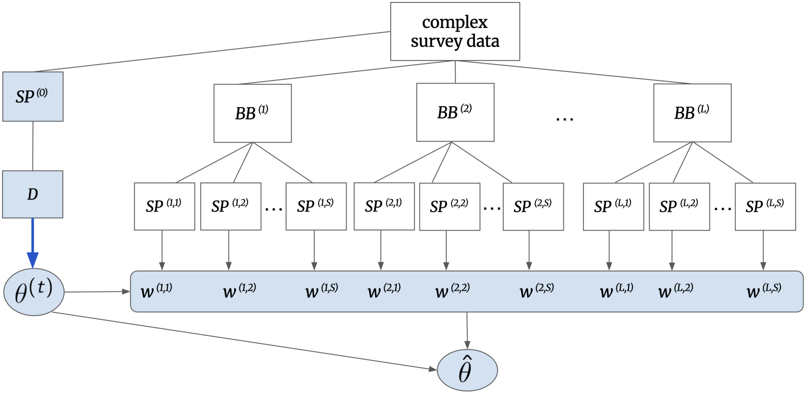
[3] Wu and Stephenson, 2024.

FPBB



→ Simplifies modeling process, but computationally inefficient to run complicated models on each synthetic population

Proposed improvement: IS-FPBB



Use **importance sampling theory** to weight draws from a model given a sample D from a single synthetic population to approximate many synthetic populations

Benefits of importance sampling (IS)

- Importance sampling allows estimation of parameter θ from a distribution $\mathcal{P}$ using weighted draws from a similar and approximate distribution $\mathcal{D}$
- Generate draws only once
- Model is applied to a sample $\mathcal{D}$ from a single synthetic population
  - Synthetic population is generated deterministically by sampling weight
  - $\mathcal{D}$ generated through systematic sampling to ensure sample spans SP

IS-FPBB weighting & inference procedure

1. Generate $L \times S$ SPs using the FPBB and 1 SP using a deterministic approach
2. Take a systematic sample (SS) from the first SP
3. Draw values $\theta^{(t)}$ from the posterior using the SS
4. Weight draws by importance sampling weights $w(\theta^{(t)})$ that use the SS to approximate the remaining $L \times S$ SPs
5. Weighted estimates $\hat{\theta}^{(\ell,s)}$ averaged across S replicates for each ℓ to get $\hat{\theta}^{(\ell)}$
6. Variance of estimates $\hat{\theta}^{(\ell)}$ computed across ℓ as $(1 + 1/L)var(\hat{\theta}^{(\ell)})$
7. Credible intervals constructed with degrees of freedom approximation

Simulation study

1. Create a stratified and clustered population ($N = 15000$) with varying stratum sizes and associations between X, Y, and random intercept Z
2. Generate complex sample ($n = 200$) with unequal probabilities of selection between strata due to varying cluster sizes
3. Apply IS-FPBB approach with a linear random intercept model
4. Repeat Steps 2 and 3 500 times
5. Compare results to unweighted (UW) and a pseudo-likelihood approach (WS)

Simulation study results

	Intercept (β_0=0.75)			**Slope** (β_1=0.998)		
	IS-FPBB	WS	UW	IS-FPBB	WS	UW
Bias	**-0.032**	**-0.025**	**-0.38**	**0.011**	**0.009**	**0.36**
Mean Posterior SD	0.28	0.26	0.25	0.094	0.073	0.072
Empirical SD	0.14	0.13	0.13	0.10	0.10	0.097
Coverage (%)	100	100	72.2	90.2	85.2	0.80
Time (sec)	5.76	94.83	3.30	5.76	94.83	3.30

- IS-FPBB and WS produce negligible bias

Simulation study results

	Intercept (β_0=0.75)			Slope (β_1=0.998)		
	IS-FPBB	WS	UW	IS-FPBB	WS	UW
Bias	-0.032	-0.025	-0.38	0.011	0.009	0.36
Mean Posterior SD	**0.28**	**0.26**	**0.25**	0.094	0.073	0.072
Empirical SD	**0.14**	**0.13**	**0.13**	0.10	0.10	0.097
Coverage (%)	100	100	72.2	90.2	85.2	0.80
Time (sec)	5.76	94.83	3.30	5.76	94.83	3.30

- Mean posterior SDs overestimate empirical SDs for intercept

Simulation study results

	Intercept (β_0=0.75)			**Slope** (β_1=0.998)		
	IS-FPBB	WS	UW	IS-FPBB	WS	UW
Bias	-0.032	-0.025	-0.38	0.011	0.009	0.36
Mean Posterior SD	0.28	0.26	0.25	0.094	0.073	0.072
Empirical SD	0.14	0.13	0.13	0.10	0.10	0.097
Coverage (%)	**100**	**100**	**72.2**	90.2	85.2	0.80
Time (sec)	5.76	94.83	3.30	5.76	94.83	3.30

- Overcoverage for IS-FPBB and WS
- Undercoverage for UW

Simulation study results

	Intercept (β_0=0.75)			**Slope** (β_1=0.998)		
	IS-FPBB	WS	UW	IS-FPBB	WS	UW
Bias	-0.032	-0.025	-0.38	0.011	0.009	0.36
Mean Posterior SD	0.28	0.26	0.25	**0.094**	**0.073**	**0.072**
Empirical SD	0.14	0.13	0.13	**0.10**	**0.10**	**0.097**
Coverage (%)	100	100	72.2	90.2	85.2	0.80
Time (sec)	5.76	94.83	3.30	5.76	94.83	3.30

- Mean posterior SD slightly underestimates empirical SD

Simulation study results

	Intercept (β_0=0.75)			**Slope** (β_1=0.998)		
	IS-FPBB	WS	UW	IS-FPBB	WS	UW
Bias	-0.032	-0.025	-0.38	0.011	0.009	0.36
Mean Posterior SD	0.28	0.26	0.25	0.094	0.073	0.072
Empirical SD	0.14	0.13	0.13	0.10	0.10	0.097
Coverage (%)	100	100	72.2	**90.2**	**85.2**	**0.80**
Time (sec)	5.76	94.83	3.30	5.76	94.83	3.30

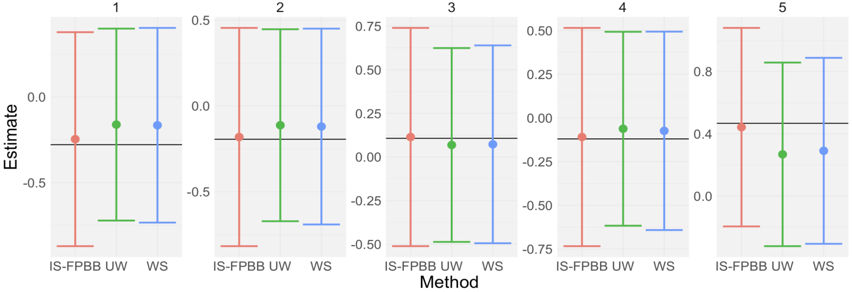
- Leading to undercoverage of varying degrees

Simulation study results

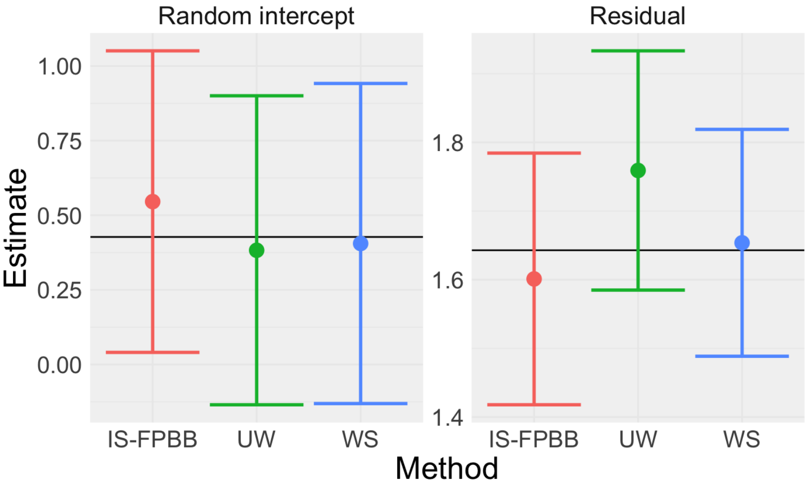
	Intercept (β_0=0.75)			**Slope** (β_1=0.998)		
	IS-FPBB	WS	UW	IS-FPBB	WS	UW
Bias	-0.032	-0.025	-0.38	0.011	0.009	0.36
Mean Posterior SD	0.28	0.26	0.25	0.094	0.073	0.072
Empirical SD	0.14	0.13	0.13	0.10	0.10	0.097
Coverage (%)	100	100	72.2	90.2	85.2	0.80
Time (sec)	**5.76**	**94.83**	3.30	5.76	94.83	3.30

- IS-FPBB is 16× faster than WS

Simulation study results: Random effects



Simulation study results: Random effect variance



Data application: NHANES

- National Health and Nutrition Examination Survey (NHANES) 2017-2018 ($n = 2,961$)
- Four-stage sampling design with clustering and stratification
- Oversampling by race, ethnicity, socioeconomic status, age
- **Goal:** study association between smoking cigarettes and blood pressure (BP) across and within sociodemographic domains
    - Smoking and increased BP are serious risk factors of cardiovascular disease
    - Racial, ethnic, education, and gender disparities in smoking and BP[4]
    - Mixed evidence of a long-term relationship between smoking cigarettes and BP[5]

[4] Al Kibria, 2019; Alexander et al., 2016; Cao et al., 2023; Hertz et al., 2005; Shiffman and Paton, 1999; Zacher, 2023.

[5] Berglund and Wilhelmsen, 1975; Omvik, 1996; Primatesta et al., 2001.

Data application: Methods

$$Y_{ij} = \beta_0 + \alpha_j + (\beta_1 + b_{1j})X_{1ij} + \beta_2 X_{2ij} + \beta_3 X_{3ij} + \epsilon_{ij}$$

$$\begin{pmatrix} b_{0j} \\ b_{1j} \end{pmatrix} \sim MVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma \right), \ \Sigma = \begin{bmatrix} \sigma_{b_0} & 0 \\ 0 & \sigma_{b_1} \end{bmatrix} \Omega \begin{bmatrix} \sigma_{b_0} & 0 \\ 0 & \sigma_{b_1} \end{bmatrix}$$

$$\Omega \sim LKJcorr(2), \beta_0, \beta_1, \beta_2, \beta_3 \sim N(0, 10),$$

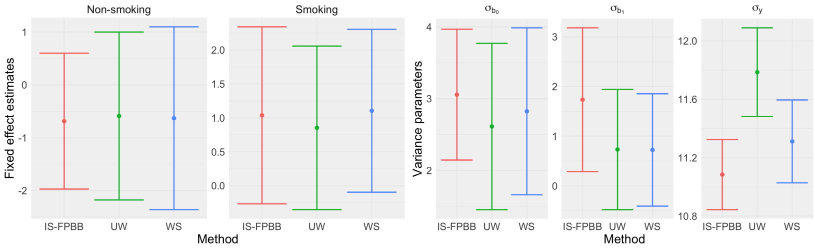
$$\epsilon_{ij} \sim N(0, \sigma_y), \sigma_{b_0}, \sigma_{b_1}, \sigma_y \sim Exp(1).$$

- ***i***: individual; ***j***: domain defined as the intersection of gender, race, ethnicity, and education
- ***Y***: BP as standardized mean arterial blood pressure (MAP), $MAP = DiastolicBP + \frac{1}{3}(SystolicBP - DiastolicBP)$ $Y = MAP - \overline{MAP}$
- $\boldsymbol{X_1}$: smoking status (1 for current smoking, 0 else)
- $\boldsymbol{X_2}$: age; $\boldsymbol{X_3}$: taking blood pressure medication or not
- Results compared across our method IS-FPBB, WS, and UW

Data application: Substantive findings

- Individuals who smoke have a higher average BP but the difference is not statistically significant
- Non-smoking NH Black individuals and Hispanic males with higher education demonstrate *increased* BP
- Non-smoking Hispanic and NH White females have *lower* BP
- No significant subgroup differences among individuals who smoke

Data application: Methodological results



- IS-FPBB and WS have more similar fixed effect estimates than UW
- Random effect estimates consistent across methods
- Our method has a larger random slope variance and smaller residual variance than the other two
- Our method is more than 3× faster than WS

Discussion

Conclusions

- Flexible and computationally efficient approach to account for sampling designs in Bayesian survey inference
- Simulation study demonstrates ability to remove bias and appropriate posterior SDs
- Data application provides adequate fixed and random effect estimation
  - Current estimation of random effect variance produces some overestimation
- Simulation study and data application show improvement in speed of estimation compared to alternative

Future work

- Extend the method for random effect models with non-normal outcomes and improve variance component estimation

Thank you

Thanks for listening! Please reach out if you have any questions, feedback, or ideas:

Annie Cohen
anniecoh@umich.edu

References I

Al Kibria, G. M. (2019). Racial/ethnic disparities in prevalence, treatment, and control of hypertension among us adults following application of the 2017 american college of cardiology/american heart association guideline. *Preventive Medicine Reports*, *14*, 100850. https://doi.org/10.1016/j.pmedr.2019.100850

Alexander, L. A., Trinidad, D. R., Sakuma, K.-L. K., Pokhrel, P., Herzog, T. A., Clanton, M. S., Moolchan, E. T., & Fagan, P. (2016). Why we must continue to investigate menthol's role in the african american smoking paradox. *Nicotine & Tobacco Research*, *18*(Suppl 1), S91–S101. https://doi.org/10.1093/ntr/ntv209

References II

Berglund, G., & Wilhelmsen, L. (1975). Factors related to blood pressure in a general population sample of swedish men. *Acta Medica Scandinavica*, *198*(1-6), 291–298. https://doi.org/https://doi.org/10.1111/j.0954-6820.1975.tb19543.x

Cao, P., Jeon, J., Tam, J., Fleischer, N. L., Levy, D. T., Holford, T. R., & Meza, R. (2023). Smoking disparities by level of educational attainment and birth cohort in the u.s.. *American Journal of Preventive Medicine*, *64*(4), S22–S31. https://doi.org/10.1016/j.amepre.2022.06.021

References III

Hertz, R. P., Unger, A. N., Cornell, J. A., & Saunders, E. (2005). Racial disparities in hypertension prevalence, awareness, and management. *Archives of Internal Medicine*, *165*(18), 2098–2104. https://doi.org/10.1001/archinte.165.18.2098

Little, R. (2022). Bayes, buttressed by design-based ideas, is the best overarching paradigm for sample survey inference. *Survey Methodology*. https://www150.statcan.gc.ca/n1/pub/12-001-x/2022002/article/00001-eng.htm

Omvik, P. (1996). How smoking affects blood pressure. *Blood Pressure*, *5*(2), 71–77. https://doi.org/10.3109/08037059609062111

References IV

- Primatesta, P., Falaschetti, E., Gupta, S., Marmot, M. G., & Poulter, N. R. (2001).Association between smoking and blood pressure. *Hypertension*, *37*(2), 187–193. https://doi.org/10.1161/01.HYP.37.2.187
- Raghunathan, T., Reiter, J. P., & Rubin, D. (2003).Multiple Imputation for Statistical Disclosure Limitation. *Journal of Official Statistics*, *19*, 1–16. Retrieved September 6, 2023, from http://www2.stat.duke.edu/~jerry/Papers/jos03.pdf
- Shiffman, S., & Paton, S. M. (1999).Individual differences in smoking: Gender and nicotine addiction. *Nicotine & Tobacco Research: Official Journal of the Society for Research on Nicotine and Tobacco*, *1 Suppl 2*, S153–157, discussion S165–166. https://doi.org/10.1080/14622299050011991

References V

Williams, M. R., & Savitsky, T. D. (2021). Uncertainty estimation for pseudo-bayesian inference under complex sampling. *International Statistical Review*, *89*(1), 72–107. https://doi.org/10.1111/insr.12376

Wu, S. M., & Stephenson, B. J. K. (2024). Bayesian estimation methods for survey data with potential applications to health disparities research. *WIREs Computational Statistics*, *16*(1), e1633. https://doi.org/10.1002/wics.1633

Zacher, M. (2023). Educational disparities in hypertension prevalence and blood pressure percentiles in the health and retirement study. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *78*(9), 1535–1544. https://doi.org/10.1093/geronb/gbad084

FPBB approach

1. For each stratum h, select an SRSWR of $n_h - 1$ PSUs and select all elements within each sampled PSU
2. Update *selected* element level weights $w_{hji}^{(\ell)} = (\frac{n_h}{n_h-1})w_{hji} := w_i^{(\ell)}$
3. Pólya sample of size $N - n$ with $P(\delta_{ip}^{(\ell,s)} = 1) = \frac{w_i^{(\ell)} - 1 + r_{i,p-1}^{(\ell,s)} * (N-n)/n}{N-n+(p-1)*(N-n)/n}$.
4. Form the FPBB SP $\mathcal{P}_{syn}^{(\ell,s)}$ by combining the Pólya and the sample from Step 1 and extract $r_i^{(\ell,s)}$ for all $i = 1, ..., n$.
5. Repeat steps 3-4 S times.
6. Repeat all steps L times.

Notation

ℓ	Bayesian bootstrap $\ell = 1, ..., L$
s	replicate of ℓ $s = 1, ..., S$
i	complex sample unit indicator
p	SP unit indicator
$\delta_{ip}^{(\ell,s)}$	selection of unit i for p^{th} unit of the $(\ell, s)^{th}$ SP
$w_{hji}^{(\ell)}$	sampling weight for unit i in the j^{th} PSU of the h^{th} stratum in the ℓ^{th} bootstrap
$r_i^{(\ell,s)}$	number of times unit i in the $(\ell, s)^{th}$ SP

Degrees of freedom approximations

$\theta|\mathcal{D}$ is t-distributed but degrees of freedom are unknown[6]

Options:

- Same degrees of freedom as $\mathcal{P}$: $L-1$, will slightly underestimate coverage because of importance sampling approximation
- Matching the observed moments of $\theta|\mathcal{D}$ with theoretical moments of $\theta|\mathcal{P}$
  - Either second or fourth moments
  - Conservative assumption that $\mathcal{D}$ is an SRS from $\mathcal{P}$
  - Varying levels of overestimation

[6] Raghunathan et al., 2003.

Importance Sampling (IS)

Allows estimation of parameter θ from a distribution $\mathcal{P}_{syn}^{(\ell,s)}$ using weighted draws from a similar and approximate distribution $\mathcal{D}$

$$\hat{\theta}^{(\ell,s)} = E(\theta|\mathcal{P}_{syn}^{(\ell,s)}) = \int \theta P(\theta|\mathcal{P}_{syn}^{(\ell,s)})d\theta \approx \frac{\sum_{t=1}^{T} w(\theta^{(t)})^{(\ell,s)}\theta^{(t)}}{\sum_{t=1}^{T} w(\theta^{(t)})^{(\ell,s)}},$$

$$\text{where } w(\theta^{(t)})^{(\ell,s)} = \frac{P(\mathcal{P}_{syn}^{(\ell,s)}|\theta)}{P(\mathcal{D}|\theta)} = \prod_{i=1}^{n} f(y_i|\theta^{(t)})^{r_i^{(\ell,s)}-d_i}$$

Notation

θ	parameter of interest
t	posterior draw indicator, $t = 1, ..., T$
d_i	number of times unit i of the complex sample is in $\mathcal{D}$
$r_i^{(\ell,s)}$	number of times unit i in the $(\ell, s)^{th}$ SP