Multilevel Regression and Poststratification

Yajuan Si
Research Assistant Professor
Survey Methodology Program, Survey Research Center
Institute for Social Research, University of Michigan

April 10, 2020

Acknowledgements

- Grant support from NSF-SES 1760133
- Comments and partial materials shared by
  - Andrew Gelman (Columbia University)
  - Lauren Kennedy (Columbia University)
  - Douglas Rivers (Stanford University)
- Views expressed are those of Yajuan Si and not those of Gelman, Rivers or Yougov

Outline

1. Overview and examples
2. Methodology and practice
3. Applications in survey research
4. Recent developments and challenges

- More detailed materials can be downloaded from Github: `https://github.com/yajuansi-sophie/MrP-presentations`

1. Overview and Examples

What is MRP?

Formally, **M**ultilevel **R**egression and **P**ost-stratificaiton

Informally, **Mr. P**

Behind MRP

Andrew Gelman

- Gelman proposed MRP (A. Gelman and Little 1997) and has demonstrated its success in public opinion research, especially on subgroup and trend analysis, e.g., Ghitza and Gelman (2013); Shirley and Gelman (2015).

- Stan made MRP generally accessible as an open source software project for statistical modeling and high-performance statistical computation.

Actually (per Gelman)

R. Little: a modeler's perspective of poststratification

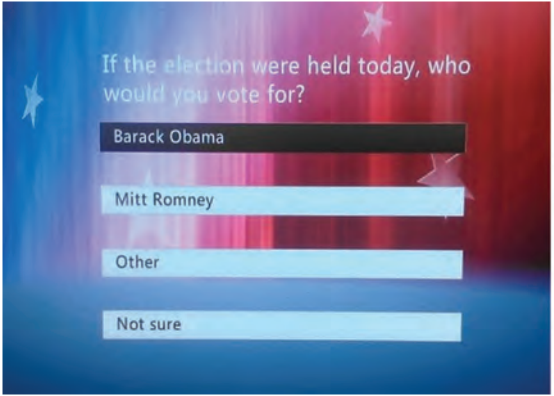
D. Rubin: multiple imputation

What problems does MRP address?

1. **DESIGN-based** *Poststratification* adjustment for selection bias. Correct for imbalances in sample composition, even when these are severe and can involve a large number of variables.

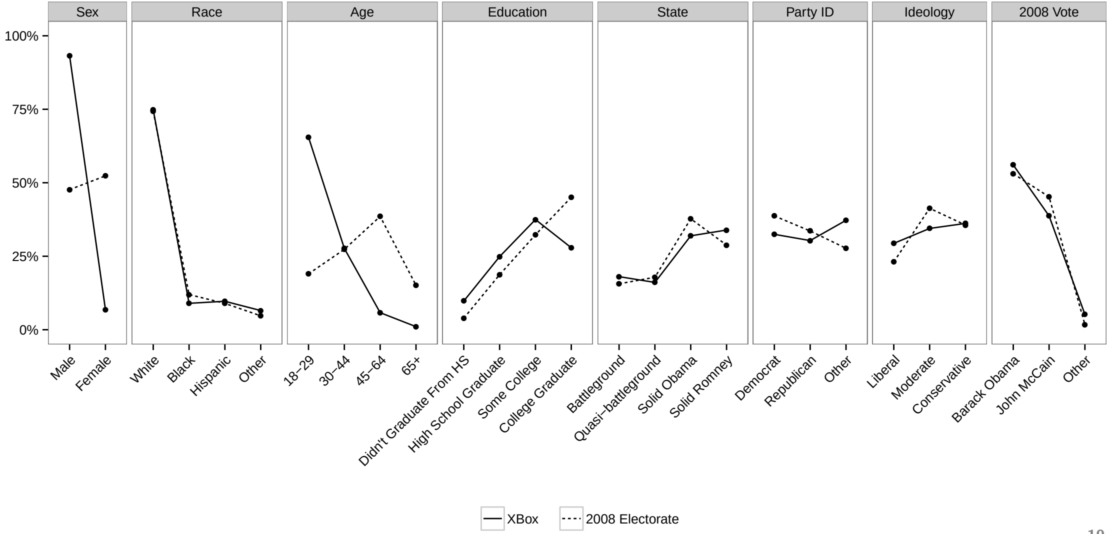
2. **MODEL-based** *Multilevel Regression* for small area estimation (SAE). Can provide stabilized estimates for subgroups over time (such as states, counties, *etc.*)

Example: the Xbox Poll



Wang et al. (2015) used MRP to obtain estimates of voting behavior in the 2012 US Presidential election based on a sample of 350,000 Xbox users, empaneled 45 days prior to the election.

Selection bias in nonrepresentative the Xbox panel



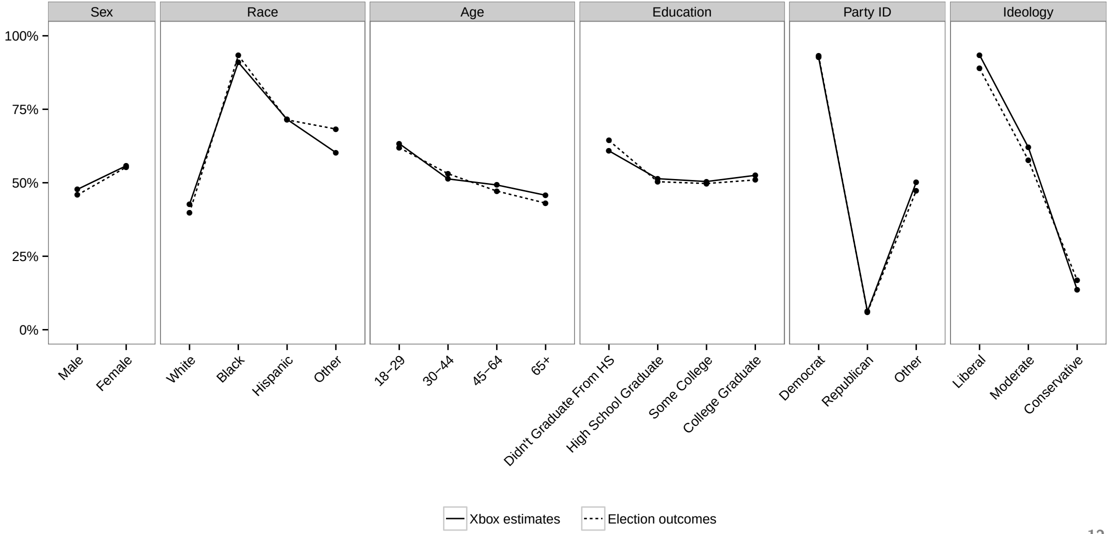
Apply MRP to big data

- Used detailed highly predictive covariates about voting behavior: sex, race, age, education, state, party ID, political ideology, and reported 2008 vote, resulting in 176,256 cells, 2 gender x 4 race x 4 age x 4 education x 4 party x 3 ideology x 50 states.
- Fit multilevel logistic regression:

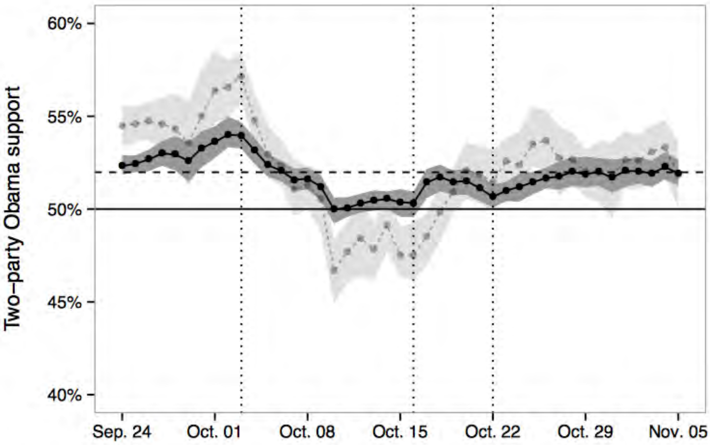
$$\Pr(Y_i = 1) = \mathrm{logit}^{-1}(\alpha_0 + \alpha_1 * sh + \alpha^{state}_{j[i]} + \alpha^{edu}_{j[i]} + \alpha^{sex}_{j[i]} + \alpha^{age}_{j[i]} + \alpha^{race}_{j[i]} + \alpha^{party}_{j[i]}),$$

  where $j[i]$ refers to the cell j that unit i belongs to.
- Introduce prior distributions $\alpha^{var}_{j[i]} \sim N(0, \sigma^2_{var})$, $\sigma^2_{var} \sim inv - \chi^2(\nu_0, \sigma^2_0)$.

MRP estimates of 2012 voting from Xbox panel



The power of poststratification adjustments



[1]

[1] The light gray line (with SEs) shows the result after adjusting for demographics; the dark gray line shows the estimates after also adjusting for day-to-day changes in the party identification of respondents. The vertical dotted lines show the dates of the presidential debates.

Examples: MRP for public health, social science research

- CDC has recently been using MRP to produce county, city, and census tract-level disease prevalence estimates in the 500 cities project (`https://www.cdc.gov/500cities/`).
- A Case Study of Chronic Obstructive Pulmonary Disease Prevalence Using the Behavioral Risk Factor Surveillance System (Zhang et al. 2014; Zhang et al. 2015).
- MRP used the relationships between demography and vote choices to project state-level election results (`https://www.economist.com/graphic-detail/2019/07/06/if-everyone-had-voted-hillary-clinton-would-probably-be-president`).

MRP can also fail

Use MRP with caution

Statistical Modeling, Causal Inference, and Social Science

HOME BOOKS BLOGROLL SPONSORS AUTHORS FEED

« Scientific communication by press release

Nate Silver's website »

President of American Association of Buggy-Whip Manufacturers takes a strong stand against internal combustion engine, argues that the so-called "automobile" has "little grounding in theory" and that "results can vary widely based on the particular fuel that is used"

Posted by Andrew on 6 August 2014, 2:45 pm

Some people pointed me to this official statement signed by Michael Link, president of the American Association for Public Opinion Research (AAPOR). My colleague David Rothschild and I wrote a measured response to Link's statement which I posted on the sister blog. But then I made the mistake of actually reading what Link wrote, and it really upset me in that it reminded me of various anti-innovation attitudes in statistics I've encountered over the past few decades.

If you want to oppose innovation, fine: there are a lot of reasons why it can make sense to go with old methods and to play it slow. Better the devil you know etc. And on the other side there are reasons to go with the new. Open discussion and debate can be helpful in establishing the zones of application where different methods are more useful.

What I really *don't* like, though, is when someone takes a position and then just makes things up to support it, as if this is some kind of war of soundbites and it doesn't matter what you say as long as it sounds good. That's what Link did in his statement. He just made stuff up. AAPOR is a serious professional organization and this statement was a serious mistake on its part.

After reading Link's article, I wrote a long sarcastic post blasting it. But then I deleted my post: really, what was the point? Instead, I'll say things as directly as possible.

In his article, Link criticizes the recent decision of the New York Times to work with polling company YouGov to conduct an opt-in internet survey. Link states that "these methods have little grounding in theory and the results can vary widely based on the particular method used."

But he's just talking out his ass. Traditional surveys nowadays can have response rates in the 10% range. There's no "grounding in theory" that allows you to make statements about those missing 90% of respondents. Or, to put it another way, the "grounding in theory" that allows

Search

RECENT COMMENTS

2. Methodology and practice

Unify design-based and model-based inferences

- The underlying theory is grounded in survey inference: a combination of SAE (Rao and Molina 2015) and poststratification (D. Holt and Smith 1979).
- Motivated by R. Little (1993), a model-based perspective of poststratification.
- Suppose units in the population and the sample can be divided into J poststratification cells with population cell size N_j and sample cell size n_j for each cell $j = 1, \ldots, J$, with $N = \sum_{j=1}^{J} N_j$ and $n = \sum_{j=1}^{J} n_j$.
- Let $\overline{Y}_j$ be the population mean and $\bar{y}_j$ be the sample mean within cell j. The proposed MRP estimator is,

$$\tilde{\theta}^{\,\mathrm{mrp}} = \sum_{j=1}^{J} \frac{N_j}{N} \tilde{\theta}_j,$$

where $\tilde{\theta}_j$ is the model-based estimate of $\bar{Y}_j$ in cell j.

Compare with unweighted and weighted estimators

1. The unweighted estimator is the average of the sample cell means,

$$\bar{y}_s = \sum_{j=1}^{J} \frac{n_j}{n} \bar{y}_j. \tag{1}$$

2. The poststratification estimator accounts for the population cell sizes as a weighted average of the sample cell means,

$$\bar{y}_{ps} = \sum_{j=1}^{J} \frac{N_j}{N} \bar{y}_j. \tag{2}$$

Bias and variance

Let the poststratification cell inclusion probabilities, means for respondents and nonrespondents be ψ_j, $\bar{Y}_{jR}$ and $\bar{Y}_{jM}$, respectively.

$$\mathrm{bias}(\bar{y}_s) = \sum \frac{\frac{N_j}{N}\bar{Y}_{jR}(\psi_j - \bar{\psi})}{\bar{\psi}} + \sum \frac{N_j}{N}(1-\psi_j)(\bar{Y}_{jR} - \bar{Y}_{jM}) \doteq A + B$$

$$\mathrm{bias}(\bar{y}_{ps}) = \sum \frac{N_j}{N}(1-\psi_j)(\bar{Y}_{jR} - \bar{Y}_{jM}) = B$$

$$Var(\bar{y}_s|\vec{n}) = \sum_j \frac{n_j}{n^2} S_j^2$$

$$Var(\bar{y}_{ps}|\vec{n}) = \sum_j \frac{N_j^2}{N^2}(1 - n_j/N_j)\frac{S_j^2}{n_j}$$

Partial pooling with MRP

- Introduce the exchangable prior, $\theta_j \sim N(\mu, \sigma_\theta^2)$.

- The approximated MRP estimator is given by

$$\tilde{\theta}^{\mathrm{mrp}} = \sum_{j=1}^{J} \frac{N_j}{N} \frac{\bar{y}_j + \delta_j \bar{y}_s}{1 + \delta_j}, \text{ where } \delta_j = \frac{\sigma_j^2}{n_j \sigma_\theta^2}, \tag{3}$$

as a weighted combination of $\bar{y}_s$ and $\bar{y}_{ps}$, where the weight is controlled by $(n_j, \sigma_\theta^2, \sigma_j^2)$.

- The bias and variance trade-off for the MRP estimator (Si, in preparation)

The key steps

1. **Multilevel regression** Fit a model relating the survey outcome to covariates across poststratification cells to estimate θ_j;
2. **Poststratification** Average the cell estimates weighted by the population cell count N_j; or **Prediction** Impute the survey outcomes for all population units.

Ingredients for MRP and the running example

Survey Pew Research Organization's *October 2016 Political Survey* (2,583 interviews, conducted October 20-25, 2016.)

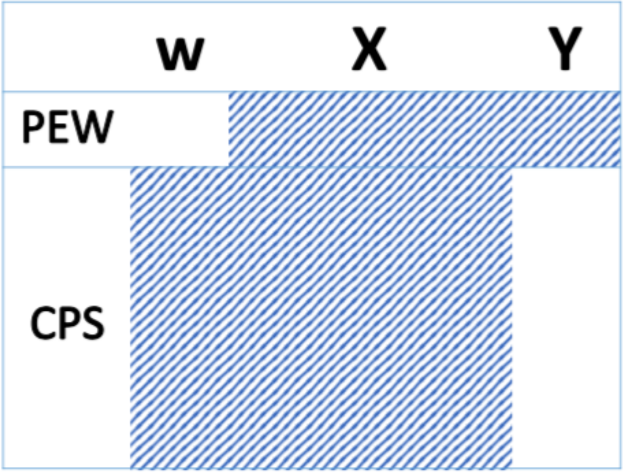
Survey variable 2016 Presidential voting intention

Covariates Individual characteristics (from the survey) and group level predictors (2012 state vote)

Post-strata Age x Gender x Race x Education x State

Stratum counts from the November 2016 Voting and Registration Supplement to the *Current Population Survey*

Data structure



The easy way with `rstanarm`

- `Rstanarm` is an R package that writes and executes Stan code for you.
- It uses the same notation as `lme4` for specifying multilevel models.

```
library(rstanarm)
fit <- stan_glmer(demvote ~ 1 + age4 + gender + race3 + educ4 +
    region + qlogis(obama12) + (1 | state), data = pew, family = binomial)
```

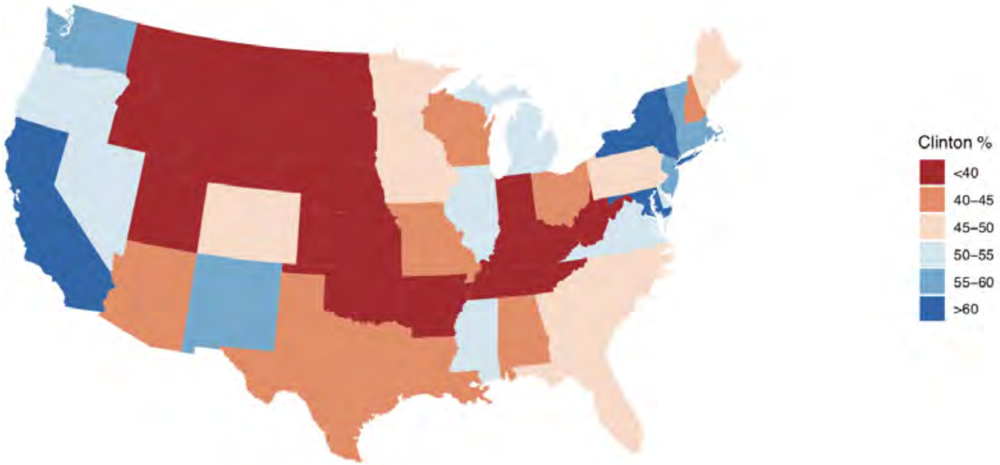
- The function `posterior_predict` in `rstanarm` substitutes for the usual `predict` function in R:

```
imputations <- posterior_predict(fit, draws = 500,
  newdata = select(cps, age4, gender, race3, educ4, region, obama12, state))
```

(This creates a matrix `imputations` of dimension `draws` x `nrow(newdata)`.)

- Extract the estimates using `get_state_estimates`.

What the map looks like

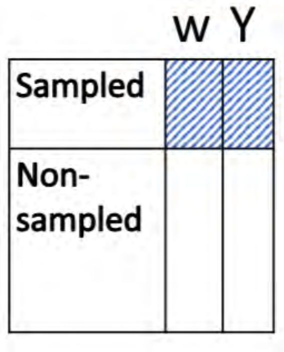


3. Applications in survey research

A unified MRP framework

- "Survey weighting is a mess" (A. Gelman 2007).
- It depends on the goal of weighting adjustments (Bell and Cohen 2007; Breidt and Opsomer 2007; R. J. A. Little 2007; Lohr 2007; Pfeffermann 2007)
- MY goal is to unify design-based and model-based inference approaches as *data integration* to
  - Combine weighting and prediction
  - Unify inferences from probability- and nonprobability-based samples
- **Key quantities** : $j = 1, \ldots, J$, θ_j and N_j

Bayesian Nonparametric Weighted Sampling Inference (Si, Pillai, and Gelman 2015)



- Consider independent sampling with unequal inclusion probabilities.
- The externally-supplied weight is the only information available.
- **Assume the unique values of unit weights determine the poststratification cells via a 1-1 mapping.**
- Simultaneously predict $w_{j[i]}$'s and y_i's for $N - n$ nonsampled units.

Incorporate weights into modeling

1. We assume n_j's follow a multinomial distribution conditional on n,

$$\vec{n} = (n_1, \dots, n_J) \sim \text{Multinomial}\left(n; \frac{N_1/w_1}{\sum_{j=1}^J N_j/w_j}, \dots, \frac{N_J/w_J}{\sum_{j=1}^J N_j/w_j}\right).$$

   Here N_j's are unknown parameters.

2. Let $x_j = \log w_j$. For a continuous survey response y, by default

$$y_i \sim \mathrm{N}(\mu(x_{j[i]}), \sigma^2),$$

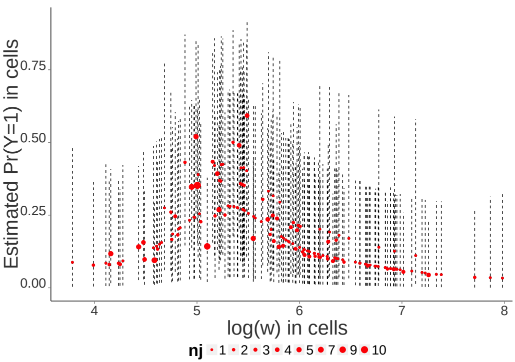
   where $\mu(x_j)$ is a mean function of x_j.

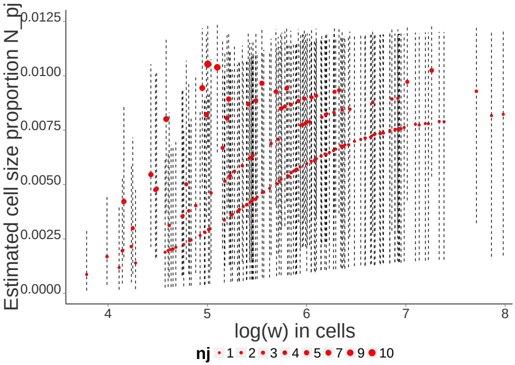
3. We introduce a Gaussian process (GP) prior for $\mu(\cdot)$

$$\mu(x) \sim \mathrm{GP}(x\beta, \Sigma_{xx}),$$

   where Σ_{xx} denotes the covariance function of the distances for any $x_j, x_{j'}$.

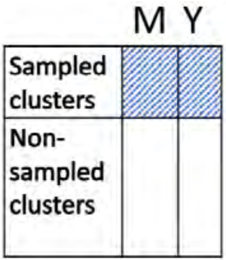
Estimates of cell means and cell size proportions





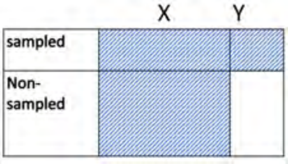
Proportion estimation of individuals with public support based on the Fragile Families and Child Wellbeing Study.

Bayesian inference under cluster sampling with probability proportional to size (Makela, Si, and Gelman 2018)



- Bayesian cluster sampling inference is essentially outcome prediction for nonsampled units in the sampled clusters and all units in the nonsampled clusters.
- However, the design information of nonsampled clusters is missing, such as the measure size under PPS.
- Predict the unknown measure sizes using Bayesian bootstrap and size-biased distribution assumptions.
- Account for the cluster sampling structure by incorporation of the measure sizes as covariates in the multilevel model for the survey outcome.

Bayesian hierarchical weighting adjustment and survey inference (Si et al. 2020)



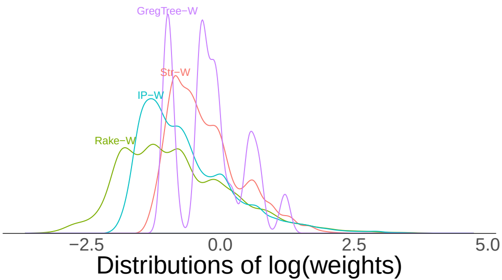
- Handle deep interactions among weighting variables
- The population cell mean θ_j is modeled as

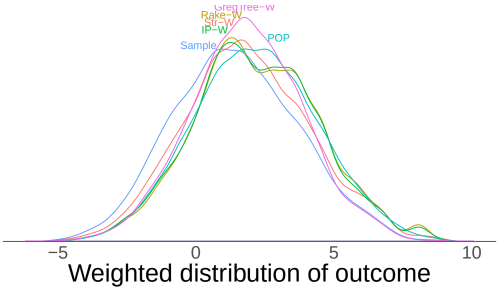
$$\theta_j = \alpha_0 + \sum_{k \in S^{(1)}} \alpha_{j,k}^{(1)} + \sum_{k \in S^{(2)}} \alpha_{j,k}^{(2)} + \cdots + \sum_{k \in S^{(q)}} \alpha_{j,k}^{(q)}, \tag{4}$$

  where $S^{(l)}$ is the set of all possible l-way interaction terms, and $\alpha_{j,k}^{(l)}$ represents the kth of the l-way interaction terms in the set $S^{(l)}$ for cell j.
- Introduce structured prior distribution to account for the hierarchical structure and improve MrP under unbalanced and sparse cell structure.
- Derive the equivalent unit weights in cell j that can be used classically

$$w_j \approx \frac{n_j/\sigma_y^2}{n_j/\sigma_y^2 + 1/\sigma_\theta^2} \cdot \frac{N_j/N}{n_j/n} + \frac{1/\sigma_\theta^2}{n_j/\sigma_y^2 + 1/\sigma_\theta^2} \cdot 1, \tag{5}$$

Model-based weights and predictions





The model-based weights are stable and yield efficient inference. Predictions perform better than weighting with the capability to recover empty cells.[2]

[2]Greg-tree is based on the tree-based method in McConville and Toth (2017)

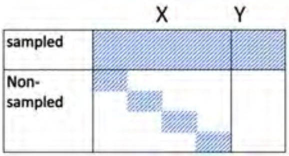
Stan fitting under structured prior in rstanarm

```
fit <-stan_glmer(formula =
      Y ~ 1 + (1 | age) + (1 | eth) + (1 | edu) + (1 | inc) +
      (1 | age:eth) + (1 | age:edu) + (1 | age:inc) +
      (1 | eth:edu) + (1 | eth:inc) +
      (1 | age:eth:edu) + (1 | age:eth:inc),
    data = dat_rstanarm,  iter = 1000, chains = 4, cores = 4,
    prior_covariance =
      rstanarm::mrp_structured(
        cell_size = dat_rstanarm$n,
        cell_sd = dat_rstanarm$sd_cell,
        group_level_scale = 1,
        group_level_df = 1
      ),
    seed = 123,
    prior_aux = cauchy(0, 5),
    prior_intercept = normal(0, 100, autoscale = FALSE),
    adapt_delta = 0.99
  )
```

Generated model-based weights

```
cell_table <- fit$data[,c("N","n")]
weights <- model_based_cell_weights(fit, cell_table)
weights <- data.frame(w_unit = colMeans(weights),
                      cell_id = fit$data[["cell_id"]],
                      Y = fit$data[["Y"]],
                      n = fit$data[["n"]]) %>%
           mutate(w = w_unit / sum(n / sum(n) * w_unit), # model-based weights
             Y_w = Y * w
           )
```

Bayesian raking estimation (Si and Zhou 2020)



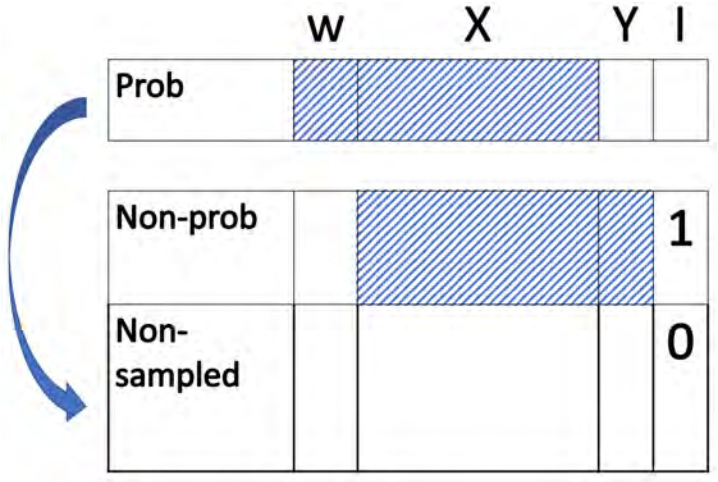
- Often the margins of weighting variables are available, rather than the crosstabulated distribution
- The iterative proportional fitting algorithm suffers from convergence problem with a large number of cells with sparse structure
- Incorporate the marginal constraints via modeling
- Integrate into the Bayesian paradigm, elicit informative prior distributions, and simultaneously estimate the population quantity of interest

4. Recent developments and challenges

Structural, spatial, temporal prior specification

- We developed structured prior distributions to reflect the hierarchy in deep interactions (Si et al. 2020)
- Sparse MRP with LassoPLUS (Goplerud et al. 2018)
- Use Gaussian Markov random fields as a prior distribution to model certain structure of the underlying categorical covariate (Gao et al. 2019)
- Using Multilevel Regression and Poststratification to Estimate Dynamic Public Opinion (A. Gelman et al. 2019)

Introduce design to big data



MRP framework for data integration (Si, in preparation)

1. Under the **quasi-randomization** approach, we assume the respondents within poststratum h are treated as a random sample of the population stratum cases,

$$\vec{n} = (n_1, \ldots, n_J)' \sim \text{Multinomial}((cN_1\psi_1, \ldots, cN_J\psi_J), n), \tag{6}$$

where $c = 1/\sum_j N_j\psi_j$, and the poststratification cell inclusion probabilities $\psi_j = g^{-1}(Z_j\alpha)$. With noninformative prior distributions, this will be equivalent to Bayesian bootstratp.

2. Under the **super-population modeling**, we assume the outcome follows a normal distribution with cell-specific mean and variance values, and the mean functions are assigned with a flexible class of prior distributions

$$\begin{aligned} y_{ij} &\sim N(\theta_j(\psi_j), \sigma_j^2) \\ \theta_j(\psi_j) &\sim f(\mu(\psi_j), \Sigma_\Psi) \end{aligned} \tag{7}$$

Work in progress

- Noncensus variables in poststratification
- Compare MRP estimator with doubly robust estimators
- Adjust for selection bias in analytic modeling
- Causal inference
- . . . .

MRP is a statistical method

HOME BOOKS BLOGROLL SPONSORS

MRP (multilevel regression and poststratification; Mister P): Clearing up misunderstandings about

Someone pointed me to this thread where I noticed some issues I'd like to clear up:

David Shor: "MRP itself is like, a 2009-era methodology."

Nope. The first paper on MRP was from 1997. And, even then, the component pieces were not new: we were just basically combining two existing ideas from survey sampling: regression estimation and small-area estimation. It would be more accurate to call MRP a methodology from the 1990s, or even the 1970s.

Will Cubbison: "that MRP isn't a magic fix for poor sampling seems rather obvious to me?"

Yep. We need to work on both fronts: better data collection and better post-sampling adjustment. In practice, neither alone will be enough.

David Shor: 2012 seems like a perfect example of how focusing on correcting non-response bias and collecting as much data as you can is going to do better than messing around with MRP.

There's a misconception here. "Correcting non-response bias" is not an alternative to MRP; rather, MRP is a method for correcting non-response bias. The whole point of the "multilevel" (more generally, "regularization") in MRP is that it allows us to adjust for more factors that could drive nonresponse bias. And of course we used MRP in our paper where we showed the importance of adjusting for non-response bias in 2012.

And "collecting as much data as you can" is something you'll want to do no matter what. Yair used MRP with tons of data to understand the 2018 election. MRP (or, more generally, RRP) is a great way to correct for non-response bias using as much data as you can.

Also, I'm not quite clear what was meant by "messing around" with MRP. MRP is a statistical method. We use it, we don't "mess around" with it, any more than we "mess around" with any other statistical method. Any method for correcting non-response bias is going to require some "messing around."

In short, MRP is a method for adjusting for nonresponse bias and data sparsity to get better survey estimates. There are other ways of getting to basically the same answer. It's important to adjust for as many factors as possible and, if you're going for small-area estimation with sparse data, that you use good group-level predictors.

MRP is a 1970s-era method that still works. That's fine. Least squares regression is a 1790s-era method, and it still works too! In both cases, we continue to do research to improve and better understand what we're doing.

2

Two key assumptions under MRP

1. Equal inclusion probabilities of the individuals within cells.

2. The included individuals are similar to those excluded within cells.

Challenges

- Robust model specification for complicated data
- Multiple (types of) survey variables
- Missing not at random/non-ignorable/informative selection
- External validation
- Incorporate substantive knowledge

Thank you

yajuan@umich.edu

References

Bell, Robert M., and Michael L. Cohen. 2007. "Comment: Struggles with Survey Weighting and Regression Modeling." *Statistical Science* 22 (2): 165–67.

Breidt, F. Jay, and Jean D. Opsomer. 2007. "Comment: Struggles with Survey Weighting and Regression Modeling." *Statistical Science* 22 (2): 168–70.

Gao, Yuxiang, Lauren Kennedy, Daniel Simpson, and Andrew Gelman. 2019. "Improving Multilevel Regression and Poststratification with Structured Priors." https://arxiv.org/abs/1908.06716.

Gelman, Andrew. 2007. "Struggles with Survey Weighting and Regression Modeling." *Statistical Science* 22 (2): 153–64.

Gelman, Andrew, and Thomas C. Little. 1997. "Poststratifcation into Many Cateogiries Using Hierarchical Logistic Regression." *Survey Methodology* 23: 127–35.

Gelman, Andrew, Jeffrey Lax, Justin Phillips, Jonah Gabry, and Robert Trangucci. 2019. "Using Multilevel Regression and Poststratification to Estimate Dynamic Public Opinion." http://stat.columbia.edu/ gelman/research/unpublished/MRT(1).pdf.

Ghitza, Y., and A. Gelman. 2013. "Deep Interactions with MRP: Election Turnout and Voting Patterns Among Small Electoral Subgroups." *American Journal of Political Science* 57: 762–76.

Goplerud, Max, Shiro Kuriwaki, Marc Ratkovic, and Dustin Tingley. 2018. "Sparse Multilevel Regression and Poststratification." https://scholar.harvard.edu/files/dtingley/files/sparsemultilevel.pdf.

Holt, D., and T. M. F. Smith. 1979. "Post Stratification." *Journal of the Royal Statistical Society Series A* 142 (1): 33–46.

Little, R. J. A. 1993. "Post-Stratification: A Modeler's Perspective." *Journal of the American Statistical Association* 88: 1001–12.

Little, Roderick J. A. 2007. "Comment: Struggles with Survey Weighting and Regression Modeling." *Statistical Science* 22 (2): 171–74.

Lohr, Sharon L. 2007. "Comment: Struggles with Survey Weighting and Regression Modeling." *Statistical Science* 22 (2): 175–78.

Makela, Susanna, Yajuan Si, and Andrew Gelman. 2018. "Bayesian Inference Under Cluster Sampling with Probability Proportional to Size." *Statistics in Medicine* 37 (26): 3849–68.

McConville, Kelly S., and Daniell Toth. 2017. "Automated Selection of Post-Strata Using a Model-Assisted Regression Tree Estimator." https://arxiv.org/abs/1712.05708.

Pfeffermann, Danny. 2007. "Comment: Struggles with Survey Weighting and Regression Modeling." *Statistical Science* 22 (2): 179–83.