

# Rake Weighting Survey Subsamples

Shiro Kuriwaki and Soichiro Yamauchi

November 2019

We first discuss two traditional approaches: (rake) weighting and regression-based propensity score weights. These traditional approaches each have known limitations. Rake weighting matches sample moments exactly but cannot match on categories that are too small. Propensity scores smooth over covariates but are not guaranteed to match on moments. After highlighting them, we outline our own proposal, a doubly-robust estimator with balancing properties. Our running example is estimating state-level weights from the Cooperative Congressional Election Study, where only national weights are provided.

## 1 Introduction to Rake Weights (Iterated Proportional Fitting)

Post-stratification is the process of generating weights to coerce the weighted proportion of the sample to equal any given target distribution (it is called “post” because it occurs after the survey was fielded, instead of over-sampling low-propensity individuals during the survey). The method is simple and was outlined in Deming and Stephan (1940). For a recent overview of the practical literature, see Kolenikov (2014).

Index individuals by  $i \in \{1, \dots, n\}$ . Each individual has a scalar weight at iteration  $t$  as  $w_i^{(t)}$ , and  $j \in \{1, \dots, K\}$  categorical covariates  $\mathbf{X}_i$ . For example, one grouping could be “education”, which is a discrete variable with four levels (high school or less, 2-year college, college, or post-graduate). Denote the levels of group  $j$  by  $\ell \in \{1, \dots, L_j\}$ .

For each grouping, there is a target distribution

$$\boldsymbol{\pi}_j = \{\pi_{j1}, \dots, \pi_{jL_j}, \sum_{\ell=1}^{L_j} \pi_{j\ell} = 1\}$$

that we would like to hit. The corresponding actual proportion from the survey is

$$\hat{\boldsymbol{\pi}}_j = \{\hat{\pi}_{j1}, \dots, \hat{\pi}_{jL_j}, \sum_{\ell=1}^{L_j} \hat{\pi}_{j\ell} = 1\}.$$

These estimates are weighted averages of binary random variables, where the normalized weights at iteration  $t$  are applied as simple weighted means:

$$\hat{\pi}_{j\ell}^{(t)} = \frac{1}{n} \sum_{i=1}^n w_i^{(t)} \mathbf{1}(X_{ji} = \ell)$$

The algorithm consists of two nested loops. The outer loop is indexed by  $t$ . The inner loop within each  $t$ , which we index by  $j$ , iterates within the  $K$  covariates. At each group  $j$ , we update the weights simply by computing the ratio of target over actual for each level  $\ell$ ,

$$r_{j\ell}^{(t+1)} \leftarrow \frac{\pi_{j\ell}}{\hat{\pi}_{j\ell}^{(t)}}. \quad (1.1)$$

This ratio is the crux of rake weighting, which is the same across packages.<sup>1</sup> This is why rake weights are also known as iterative *proportional* weighting. To interpret this ratio as weights, we apply these factors equally to all respondents for which  $X_{ji} = \ell$ , with some normalizing so that the sum of weights is  $n$ ,

$$w_i^{(t+1)} \leftarrow \left( \sum_{\ell=1}^{L_j} \mathbf{1}(X_{ji} = \ell) r_{j\ell}^{(t+1)} \right) / \underbrace{\left( \frac{1}{n} \sum_{i=1}^n \sum_{\ell=1}^{L_j} \mathbf{1}(X_{ji} = \ell) r_{j\ell}^{(t+1)} \right)}_{\text{Normalizing factor}} \quad (1.2)$$

this will always set the proportion to the target because now for all  $\ell \in \{1, \dots, L_j\}$ ,

$$\frac{1}{n} \sum_{i=1}^n w_i^{(t+1)} \mathbf{1}(X_{ji} = \ell) = \frac{\pi_{j\ell}}{\hat{\pi}_{j\ell}^{(t+1)}} \frac{1}{n} \sum_{i=1}^n \mathbf{1}(X_{ji} = \ell) = \pi_{j\ell} \quad (1.3)$$

**Convergence** However, the inner loop does not necessarily guarantee that the weight  $w^{(t+1)}$  balances the sample moments for other groups  $j'$ .

At the each of iteration, the algorithm assess total balance on typically a squared loss:

$$SS^{(t+1)} = \sum_{j=1}^K \sum_{\ell=1}^{L_j} \left( \frac{1}{n} \sum_{i=1}^n (w_i^{(t+1)} \mathbf{1}(X_{ji} = \ell)) - \pi_{j\ell} \right)^2$$

and stops if the sums of square is below some threshold, e.g.  $SS^{(t+1)} < 10^{-5}$ . Other algorithms such as what Kolenikov describes for Stata give the condition that all  $K$  categories each meet its own threshold.

**Limitations** There are two limitations to this algorithm. The first is that users typically set a cap on the weights, e.g.  $w_i^{(t+1)} = \max\{\tilde{w}_i^{(t)}, 10\}$ . The CCES caps weights at 15 for the common content ( $n = 60,000$ ) and at 7 for team modules ( $n = 1,000$ ). The second is easily expected from the fact that the main function has  $\hat{p}_{j\ell}$  in the denominator, meaning it cannot be zero. That is,

$$\sum_{i=1}^n \mathbf{1}(X_{ji} = \ell) > 0, \forall j, \ell$$

i.e. that there is at least one observation for each level, in all groupings. This is essentially a limit on the number of dimensions one can match on. For example, if there are no Asian Americans over 65 years old in a sample of Vermont respondents, then the rake weighting cannot weight to race-age bin interactions.

<sup>1</sup> See, e.g., the relevant lines in the source code of [anesrake](#), [iterake](#), and [survey](#).

## 2 Implementation of Rim Weights

We use the 2017 ACS ( $N = 3,210,525$ ). This includes 267,971 non-citizens, but we keep them in for now because they are part of the ACS calibrated counts. The covariates are those mentioned at the beginning of this memo: gender, age, race, and education.

We separate states into three tiers based on a data-availability basis:

1. Large states, where all six pairwise interaction of the categories has no zero-cells in the CCES. These include the top seven largest states in the CCES. These states are California, Texas, Florida, New York, Ohio, Pennsylvania, and Illinois. We calculate rim weights by the *marginals and the interactions*.
2. Medium states, where at least pairwise interaction has at least one zero cell, but where all the marginals are populated in the CCES. We calculate rim weights here by the *marginals only*.
3. Small states, where even some states have missing cells. This happens only in race, and those states are Alaska, Delaware, and North Dakota. We calculate the rim weights here by *marginals only, ignoring the zero cell altogether*.

We conducted this using the ACS with sampling weights and the ACS without sampling weights. The resulting rim weights for each state are called `weight_st_wacs` and `weight_st_uacs`, respectively.