Ralph Folsom's 1991 Social Statistics Proceedings Paper: A Seminal Contribution to Calibration Weighting for Nonresponse You Likely Never Heard of

Jeremy Aldworth[1], Philip S. Kott[2]
[1]RTI International, 3040 Cornwallis Road, Research Triangle Park, NC 27709
[2]RTI International, 6110 Executive Boulevard, Suite 900, Rockville, MD 20852

Abstract
In their celebrated 1992 JASA paper, Deville and Särndal introduced the term "calibration weighting" to describe mild adjustments of the design weights in a survey sample that, in the absence of unit nonresponse, force weighted estimates for a vector of calibration variables to equal known population totals. Their paper showed that linear-regression weights and raking weights were examples of calibration weights.

Few know, however, that in a 1991 paper in the ASA PROCEEDINGS OF THE SECTION ON SOCIAL STATISTICS, Ralph Folsom employed what would later be called calibration equations to adjust for unit nonresponse. He showed that using raking weights leads to nearly unbiased estimated totals when the log of the probability of a unit response is a linear function of the raking variables. He also showed how to create (not-yet-called calibration weights when the probability of unit response is a logistic function of the calibration variables. Folsom and collaborators would later expand his 1991 work and develop the GEM (general exponential model) stand-alone software, which later became the basis of the WTADJUST procedure in SUDAAN.

Key Words: Raking, Linear-regression weighting, logistic response model

1. Introduction

1.1 Outline of Paper

In Section 2, we introduce Deville and Särndal's (1992) seminal "calibration weighting" paper. They describe an approach in which the design weights of a sample may be (mildly) adjusted so that under the new calibrated weights, weighted totals of auxiliary variables equal their known population totals. Henceforth, we will refer to the resulting adjustments made to the design weights as poststratification (PS) adjustments. Their approach assumes the absence of unit nonresponse (NR).

In Section 3, we discuss how *one year earlier*, Folsom (1991) described a similar approach, in which raking equations were constructed to achieve PS adjustments to the design weights based on auxiliary variables with known population totals. Folsom (1991) extended and modified this approach so that calibration equations could be constructed to satisfy NR adjustments as well.

Finally, in Section 4, we show how Folsom and collaborators expanded his original work, together with the work done by Deville and Särndal (1992), to develop the general

exponential model (GEM), which is a unified approach that allows adjustments for PS, NR, and (they claimed) extreme values (EVs). Fitting the GEM model is the basis for the WTADJUST procedure in SUDAAN.

1.2 Notational Preliminaries

Consider a *population* $U = \{1, \dots, k, \dots, N\}$, and *sample* $S \subseteq U$. The *sampling design* $p(\cdot)$ is a probability mass function defined on the set of all 2^N subsets of U, and is used to select a particular sample, s (i.e., $p(s) = P(S = s)$). Define the *sample membership indicator* $I_k = 1$ if $k \in s$ and $I_k = 0$ otherwise. Assume that for each unit $k \in U$, the *first-order inclusion probability* $\pi_k = E(I_k) = P(k \in s) = \sum_{s \subset U: k \in s} p(s)$ is positive, and known for each $k \in s$. The *design weights* are defined as $d_k = 1/\pi_k\,;\ k \in s$.

Denote y_k as the *study variable*, and suppose there also exists a p-vector of *auxiliary variables* $\mathbf{x}_k = (x_{k1}, \dots, x_{kp})'$, for which there exists a vector $\mathbf{c}$ such that $\mathbf{c}'\mathbf{x}_k = 1$ for all $k \in s$. Assume that $\mathbf{x}_k$ is observed for all $k \in s$. In the absence of unit nonresponse, y_k would also be observed for all $k \in s$.

Denote the population totals of y_k and $\mathbf{x}_k$ as $t_y = \sum_U y_k$ and $\mathbf{t}_x = \sum_U \mathbf{x}_k$, respectively, where $\sum_A z_k$ is used as shorthand for $\sum_{k \in A} z_k$. Their corresponding π-estimators are defined as $\hat{t}_{y\pi} = \sum_s d_k y_k$ and $\hat{\mathbf{t}}_{x\pi} = \sum_s d_k \mathbf{x}_k$, respectively.

2. Deville and Särndal's (1992) "Calibration Weighting" Approach

2.1 Overview

The objective of the Deville and Särndal (1992) [henceforth referred to as D&S] approach is to create new "calibrated" w_k weights, based on (mildly) adjusted d_k weights, to satisfy the following constraint, hereinafter referred to as a *PS constraint*:

$$\sum_s w_k \mathbf{x}_k = \sum_U \mathbf{x}_k = \mathbf{t}_x. \qquad (1)$$

Note that PS constraint (1) requires a couple of assumptions to be met. We have already mentioned one of them: $\mathbf{x}_k$ is observed for all $k \in s$. In addition, we require the population totals $\mathbf{t}_x$ to be known (e.g., from an external source like the US Census Bureau).

Their approach is to determine a *distance function* between the d_k and w_k weights, and then to minimize the distance function (summed over the sample) subject to PS constraint (1). This can be viewed as a constrained optimization problem using Lagrange multipliers.

2.2 Example

D&S illustrate their approach with an example using the distance function $(w_k - d_k)^2/d_k$. Minimization of this distance function (summed over the sample) subject to PS constraint (1) results in the calibration weights $w_k = d_k(1 + \mathbf{x}_k'\boldsymbol{\lambda})$, where $\boldsymbol{\lambda} = (\sum_s d_k \mathbf{x}_k \mathbf{x}_k')^{-1}(\mathbf{t}_x - \hat{\mathbf{t}}_{x\pi})$. Note that the *adjustment factor* $w_k/d_k = a_k(\boldsymbol{\lambda}) = 1 + \mathbf{x}_k'\boldsymbol{\lambda}$ is linear in $\boldsymbol{\lambda}$.

The resulting estimator of t_y is the familiar *generalized regression estimator* $\hat{t}_{ygreg} = \sum_s w_k y_k = \hat{t}_{y\pi} + (\mathbf{t}_x - \hat{\mathbf{t}}_{x\pi})\widehat{\mathbf{B}}_s$, where $\widehat{\mathbf{B}}_s = (\sum_s d_k \mathbf{x}_k \mathbf{x}_k')^{-1} \sum_s d_k \mathbf{x}_k y_k$.

2.3 Generality of D&S Approach

The power of the D&S approach is its generality, because different distance functions yield different calibration weights. The authors provide seven case examples of distance functions, and the corresponding adjustment factors that they yield. In other words, the D&S approach gives rise to a flexible family of estimators of totals satisfying PS constraints based on auxiliary variables.

However, the D&S approach assumes there is no unit nonresponse.

3. Folsom's (1991) Raking Approach

3.1 Overview

Folsom (1991) does not appear to be available anyway online, and we only managed to find a yellowed hardcopy of the paper in the US Census Bureau library! Folsom uses similar raking techniques for both PS and NR adjustments to the weights.

3.2 Folsom's Raking Approach to PS Adjustments

Folsom's raking algorithm is set up to find an adjustment factor $a_k(\boldsymbol{\lambda})$ so that it satisfies the following calibration equation based on auxiliary variables:

$$\sum_s w_k \mathbf{x}_k \equiv \sum_s d_k a_k(\boldsymbol{\lambda}) \mathbf{x}_k = \sum_U \mathbf{x}_k = \mathbf{t}_x. \quad (2)$$

Note that calibration equation (2) (Folsom called it a "raking equation") is identical to PS constraint (1), but here the relationship $w_k = d_k a_k(\boldsymbol{\lambda})$ is explicitly expressed. Using justification from earlier work (which he references), Folsom chooses the form of the adjustment factor to be an exponential function,

$$a_k(\boldsymbol{\lambda}) = \exp(\mathbf{x}_k' \boldsymbol{\lambda}), \quad (3)$$

which is nonlinear in $\boldsymbol{\lambda}$. Folsom solved the calibration equations to determine $a_k(\boldsymbol{\lambda})$ by iteratively reweighted least squares.

D&S derive an identical exponential adjustment factor (Case 2 in their paper). But they note that the range of $\exp(\mathbf{x}_k' \boldsymbol{\lambda})$ is $(0, \infty)$! A range this large has the potential to cause an adjusted weight to become an EV.

D&S therefore suggest the following modification to trim the range of this adjustment factor (Case 6):

$$a_k(\boldsymbol{\lambda}) = \frac{l(u-1)+u(1-l)\exp(A\mathbf{x}_k'\boldsymbol{\lambda})}{(u-1)+(1-l)\exp(A\mathbf{x}_k'\boldsymbol{\lambda})}, \quad (4)$$

where $A = (u-l)/((1-l)(u-1))$, for l and u user-specified *lower* and *upper* bounds, respectively, with $0 < l < 1 < u$. It can be shown that adjustment factor (4) has the user-specified range: $l < a_k(\boldsymbol{\lambda}) < u$, which may be useful for preventing EVs after weight adjustments have been applied.

In Section 4, we show how Folsom and collaborators use a similar modification to D&S's modified adjustment factor (4) in their development of the GEM.

3.3 Folsom's Approach to NR Adjustments

Recall that PS constraint (1) (equivalently, calibration equation (2)) requires that $\mathbf{x}_k$ is observed for all $k \in s$. Folsom shows that this requirement also needs to be met for the calibration equation he developed to adjust weights for NR. However, in the presence of NR (i.e., where y_k is not observed for some units in s), how do we meet this requirement?

Folsom suggests that in the presence of NR, knowledge about $\mathbf{x}_k$ for all $k \in s$ may be available in the equivalent of a double sampling design. An example of a double sampling design is one in which the Phase 1 sample is a large screening sample of households where detailed roster information is collected. The Phase 2 sample is a smaller subsample of single persons selected from respondents of the Phase 1 sample (and serves as our sample s).

Folsom assumes that information about $\mathbf{x}_k$ for all $k \in s$ would be available from what is equivalent to the Phase 1 sample of a double sampling design. He then showed that we can use this information, as well as modeled information about the NR mechanism itself, to develop an appropriate calibration equation for NR.

To characterize the NR mechanism, we first need to define a *response indicator variable*:

$$r_k = \begin{cases} 1, & \text{if } k \in s \text{ responds;} \\ 0, & \text{if } k \in s \text{ responds not.} \end{cases}$$

If the binary variable r_k is associated with any of the available auxiliary variables, then we can create a logistic regression model formalizing those associations in what is usually called a *response propensity model*:

$$P(r_k = 1) = (1 + \exp(-\mathbf{x}_k'\boldsymbol{\lambda}))^{-1}; \ \ k \in s. \tag{5}$$

Now we have all the pieces to create the NR calibration equation. We need to find an adjustment factor $a_k(\boldsymbol{\lambda})$ so that it satisfies the following calibration equation based on auxiliary variables:

$$\sum_s r_k w_k \mathbf{x}_k \equiv \sum_s r_k d_k a_k(\boldsymbol{\lambda}) \mathbf{x}_k = \sum_s d_k \mathbf{x}_k \equiv \hat{\mathbf{t}}_{x\pi}. \tag{6}$$

Note that the control totals in calibration equation (6) are *random*, because we have all the information we need to create them. We may not have access to fixed (i.e., population) control totals for the auxiliary variables in question.

Examination of calibration equation (6) suggests that an intuitively obvious adjustment factor would be the *inverse* of the response propensity function (5). In other words, the adjustment factor should have the form,

$$a_k(\boldsymbol{\lambda}) = [P(r_k = 1)]^{-1} = (1 + \exp(-\mathbf{x}_k'\boldsymbol{\lambda})) \geq 1. \tag{7}$$

As indicated, the NR adjustment factor (7) has the essential property that it cannot be less than 1. In contrast, the PS exponential adjustment factor (3) has 0 as a lower bound, and hence would not be as suitable for NR adjustments. Equation (5) assumes that the probability of response is a logistic function of $\mathbf{x}_k$. Folsom also considered the possibility that the probability of response had the form:

$$P(r_k = 1) = \exp(-\mathbf{x}_k'\boldsymbol{\lambda})\,;\ k \in s.$$

This leads to the less-suitable raking adjustment factor in equation (3).

In the PS adjustment approach, you select your own auxiliary variables (obviously, from among those where $\mathbf{x}_k$ is observed for all $k \in s$, and the population totals $\mathbf{t}_x$ are known). In contrast, in the NR adjustment approach, the response propensity model selects the auxiliary variables (obviously, from among those where $\mathbf{x}_k$ is observed for all $k \in s$, taken from the Phase 1 sample). Thus, the NR adjustment approach implicitly assumes that NR is dependent on the auxiliary variables in the response propensity model, but independent of the response (i.e., study) variable. In other words, the approach is based on a missing-at-random (MAR) assumption, which is much less heroic than a missing-completely-at-random (MCAR) assumption.

In addition, Folsom derives variance estimators that capture the impact of his calibration methods on the estimators of totals using the adjusted weights. We will not discuss them here, but will point out that Folsom correctly accounted for his estimators using information outside the respondent sample in variance estimation. Often when raking or another calibration weighting technique is employed, users mistakenly treat the calibration weights as design weights in variance estimation.

4. Folsom and Collaborators' Generalized Exponential Model (GEM)

4.1 Overview

Folsom and collaborators (Folsom and Singh, 2000; Chen, Penne, and Singh, 2000; Singh and Folsom, 2000; and Vaish, Gordek, and Singh, 2000) propose a generalized weight-adjustment approach that can simultaneously address the following sampling concerns:

- NR bias – by using NR adjustments
- Coverage bias and variance inflation – by using PS adjustments
- Variance inflation in small domains due to EVs – by setting appropriate lower and upper bounds to adjustments.

4.2 GEM Adjustment Factor

Folsom and collaborators use adjustment factor (4) (i.e., D&S Case 6 adjustment factor) as the basis for their generalization. Recall that for adjustment factor (4), users can specify the lower and upper bounds of the range of the adjustment factor: $l < a_k(\boldsymbol{\lambda}) < u$, where $0 < l < 1 < u$. Note that the fixed "centering constant" is equal to 1.

Folsom and collaborators extend adjustment factor (4) to obtain the GEM adjustment factor as follows:

$$a_k(\boldsymbol{\lambda}) = \frac{l_k(u_k - c_k) + u_k(c_k - l_k)\exp\,(A_k\mathbf{x}_k'\boldsymbol{\lambda})}{(u_k - c_k) + (c_k - l_k)\exp\,(A_k\mathbf{x}_k'\boldsymbol{\lambda})}, \qquad (8)$$

where $A_k = (u_k - l_k)/((c_k - l_k)(u_k - c_k))$, with $0 < l_k < c_k < u_k$, for each $k \in s$.

The GEM adjustment factor (8) differs from the D&S adjustment factor (4) in that it has *lower* and *upper* bounds (l_k and u_k, respectively) that can be specified by users for *each unit* $k \in s$; in addition, a variable *centering constant* c_k replaces the fixed constant 1 (and

can be specified for each unit $k \in s$). It can be shown that adjustment factor (8) has the user-specified range: $l_k < a_k(\boldsymbol{\lambda}) < u_k$, for each $k \in s$.

Folsom and collaborators also demonstrate that the GEM-calibrated estimator of t_y is asymptotically design-consistent under mild assumptions. Their estimated variance captures the impact of GEM modeling.

4.3 GEM Example

An example is provided to illustrate how GEM can be used to address the three sampling concerns described in Section 4.1.

We first need to determine appropriate lower and upper bounds to prevent the adjustments from creating EVs among the adjusted weights. The GEM adjustment factor has the flexibility to specify different lower and upper bounds for each unit in the sample, but in practice Folsom and his collaborators argue that it will usually be sufficient to specify the bounds within each of three weight domains (i.e., high EV, non-extreme, and low EV). Then set the lower and upper bounds appropriately for each of the three domains, separately for NR and PS adjustments (discussed below). For example, the domain consisting of non-extreme weights (which should be the largest domain, by far) would probably not require any further adjustments to the lower and upper bounds already set for NR and PS adjustments.

NR and PS adjustments are typically applied in two steps. First, NR adjustments are applied to the design weights, then PS adjustments are applied to the NR-adjusted weights.

For NR adjustments, all lower bounds must be set to be at least 1 (this ensures that the adjustment factor itself has a lower bound of 1, which is required for NR adjustments, as indicated in Section 3.3). The centering constant always needs to be greater than the lower bound (a value of 2 or the inverse of the response propensity for each sampling unit would work – in practice, it does not really matter which of the two is used). Finally, set the upper bound appropriately for each of the three EV weight domains. (One of us does not like creating EV weight domains for NR adjustments; Kott and Liao, 2012, p. 108.)

For PS adjustments, the centering constant always needs to be set equal to 1. Set lower and upper bounds appropriately for each of the three EV weight domains.

References

Chen, P., Penne, M. A., and Singh, A. C. (2000). Experience with the generalized exponential model for weight calibration for the National Household Survey on Drug Abuse, *ASA Proc. Surv. Res. Meth. Sec.*, 604-609.

Deville, J. -C. and Särndal, C. E. (1992). Calibration estimation in survey sampling, *JASA*, 87, 376-382.

Folsom, R. E. (1991). Exponential and logistic weight adjustments for sampling and nonresponse error reduction, *ASA Proc. Soc. Statist. Sec.*, 197-202.

Folsom, R. E. and Singh, A. C. (2000). A generalized exponential model for sampling weight calibration for extreme values, non-response and poststratification, *ASA Proc. Surv. Res. Meth. Sec.*, 598-603.

Kott, P. S. and Liao, D. (2012). Providing double protection for unit nonresponse with a nonlinear calibration-weighting routine, *Surv. Res. Meth.*, 105-111.

Singh, A. C. and Folsom, R. E. (2000). Bias corrected estimating function approach for variance estimation adjusted for post-stratification, *ASA Proc. Surv. Res. Meth. Sec.*, 610-615.

Vaish, A. K., Gordek, H., and Singh, A. C. (2000). Variance estimation adjusted for weight calibration via the generalized exponential model with application to the National Household Survey on Drug Abuse, *ASA Proc. Surv. Res. Meth. Sec.*, 616-621.