A NEARLY PSEUDO-OPTIMAL METHOD FOR KEEPING CALIBRATION WEIGHTS FROM FALLING BELOW UNITY IN THE ABSENCE OF NONRESPONSE OR FRAME ERRORS

Phillip S. Kott
Senior Research Statistician, RTI International
6110 Research Blvd.; Rockville, MD 20852
Email: pkott@rti.org

ABSTRACT

Calibration weighting is a technique for adjusting randomization-based weights so that the estimator of a population total becomes unbiased under a linear prediction model. In the absence of nonresponse or frame errors, one set of calibration weights has been shown to be asymptotically optimal in some sense under Poisson sampling. Unfortunately, although it is desirable that each weight be at least one (so that the corresponding element “represents itself”), there is no guarantee that will be the case. We will see how to construct an asymptotically equivalent set of weights so that no weight is less than unity. One consequence is that it often will be a simple matter to construct a variance measure that simultaneously estimates the prediction variance and the randomization mean squared error of the estimator.

KEY WORDS:
Generalized exponential form, Prediction model, Randomization-based.

1. INTRODUCTION

Suppose we want to estimate the total a target variable \( y \) in a population \( U \) for based on data from a probability sample \( S \). If we know the selection probability, \( \pi_k \), for each sample element \( k \) in \( S \), then we could estimate \( T_y = \sum_{k \in U} y_k = \sum_{U} y_k \), with the expansion (Horvitz-Thompson) estimator

\[
E_t = \sum_{k \in S} y_k / \pi_k = \sum_{U} y_k I_k / \pi_k,
\]

where \( I_k = 1 \) when \( k \in S \) and 0 otherwise. In randomization-based inference, the \( I_k \) are random variables. It is easy to see that \( E_t(t_y) = T_y \). In other words, \( t_y \) is a randomization-unbiased estimator for \( T_y \).

We can also write \( t_y = \sum_{U} d_k y_k = \sum_{S} d_k y_k \), where \( d_k = I_k / \pi_k \). Deville and Särndal (1992) coined the term “calibration estimator” to describe an estimator of the form \( \sum_{S} w_k y_k \), where the set of calibration weights \( \{w_k\} \) are chosen to satisfy the calibration equation:

\[
\sum_{k \in S} w_k x_k = \sum_{k \in U} x_k = T_x
\]  

(1.1)
for some row vector of $P$ benchmark variables, $\mathbf{x}_k = (x_{1k}, ..., x_{Pk})$, about which $T_x$ is known. Another requirement is that $t_{y}^{CAL}$ be randomization consistent under mild conditions. This means that as the sample size grows arbitrarily large, the relative randomization mean squared error of $t_{y}^{CAL}$ must tend toward zero. For more on the randomization-based properties of the calibration estimator, see Kott (2008).

Since equation (1.1) holds, $t_{y}^{CAL}$ estimates $T_y$ perfectly when $y_k = \mathbf{x}_k \beta$ exactly. With that in mind, it is reasonable to expect $t_{y}^{CAL}$ to be a good estimator when $y_k$ and $\mathbf{x}_k \beta$ are close. This can be formalized by assuming the $y_k$ are random variables satisfying the linear prediction model:

$$y_k = \mathbf{x}_k \beta + \epsilon_k,$$

where $E(\epsilon_k | \{x_k, I_g; g \in U\}) = 0$ for all $k \in U$. Under this model, it is easy to see that $t_{y}^{CAL}$ is an unbiased estimator for $T_y$ in the sense that $E(t_{y}^{CAL} - T_y) = 0$ (suppressing the conditioning on $\mathbf{x}_k$ and $I_g$ for notational convenience).

Section 2 describes the general regression (GREG) estimator, which translates into the linear calibration estimator with $w_k = d_k (1 + q_k x_k g)$ for an appropriately defined vector $g$. Although each $q_k$ is often set to 1 in practice, the calibration estimator can be shown to be optimal in some sense under Poisson sampling when $q_k = (1 - \pi_k) / \pi_k$.

There is no guarantee with linear calibration that all the calibration weights will be nonnegative. Many computer programs cannot handle nonnegative weights.

Brewer (1999) argues that calibrated weights should never fall below unity, because when an element’s weight is less than 1, it “does not even represent itself.” Section 3 shows how the nonlinear general exponential form (Folsom and Singh 2000) can be combined with a particular set of instrumental variables to create calibration weights that never fall below unity but are asymptotically equivalent to “optimal” weights.

Section 4 provides a discussion of the simultaneous estimation of model variance and randomization mean squared error, which is aided by the absence of calibration weights less than 1.

2. THE GENERAL REGRESSION ESTIMATION

The general regression or GREG estimator has the form:

$$t_{y}^{GREG} = t_{y}^{E} + \left( T_x - \sum_{k \in S} d_k x_k \right) \left( \sum_{k \in S} q_k d_k x_k' x_k \right)^{-1} \sum_{k \in S} q_k d_k x_k' y_k,$$

This assumes that the matrix $\sum_{k \in S} q_k d_k x_k' x_k$ is invertible, which we do.

The GREG estimator in equation (2.1) can be put in calibration form as $t_{y}^{GREG} = \sum_{k \in S} w_k y_k$, by setting
\[ w_k = d_k + \left( T_x - \sum_{j \in S} d_j x_j \right) \left( \sum_{j \in S} q_j d_j x_j' x_j \right)^{-1} q_k d_k x_k' \]

Strictly speaking, the \( w_k \) are functions of the realized sample, \( S \), and the \( q_k \), but we suppress that in the notation for convenience.

To choose values for the \( q_k \) that are optimal in some sense, we follow the reasoning in Rao (1994) and first consider randomization unbiased estimators of the form

\[ t_y^{(b)} = t_y^E + \left( T_x - \sum_{k \in S} d_k x_k \right) b, \]

where \( b \) is a known matrix. It is not hard to see that the choice of \( b \) that minimizes the randomization variance of \( t_y^{(b)} \) is \( b = \left[ \text{Var}_I(\Sigma_S d_k x_k) \right]^{-1} \text{Cov}_I(t_y^E, \Sigma_S d_k x_k) \). Although this \( b \) is generally unknown, both \( \text{Var}_I(.) \) and \( \text{Cov}_I(.) \) can be estimated from the sample. Under Poisson sampling, plugging unbiased and consistent estimates for \( \text{Var}_I(.) \) and \( \text{Cov}_I(.) \) into \( b \) would yield the calibration estimator in equation (1.1) with \( q_k = (1 - \pi_k) / \pi_k = d_k - 1 \). That calibration estimator is thus asymptotically optimal in some sense under Poisson sampling. Bankier (2002) named it “pseudo-optimal,” observing that it is asymptotically optimal under stratified simple random sampling when the strata are large and their indicator functions are components of \( x_k \).

Using different reasoning, Brewer (1999) suggests setting \( q_k = (1 - \pi_k) / z_k \), where \( z_k \) is some measure of size. With this setting, when element \( k \) is a certainty selection, which means \( d_k = 1 \), \( w_k \) is forced to equal 1 as well. In practice, this tends to limit the number of weights falling below unity under linear calibration. It does not guarantee their elimination, however. When one or more \( w_k \) falls below 1 after linear calibration, Brewer advises setting those \( w_k \) to 1, removing the corresponding elements from the sample and population in equation (1.1), and running linear calibration again on the remainder of the sample. This can be done iteratively if necessary.

3. THE GENERALIZED EXPONENTIAL FORM

In an earlier work, Brewer (1995) generalizes the GREG by replacing each \( q_j x_j \) with the instrumental vector \( h_j \). This generalization assumes that \( \sum_S d_k h_k' x_k \) is invertible, which we do here. After the replacements are made, one can write the resulting estimator in calibration form with \( w_k = d_k (1 + h_k g) \), where \( g = \left( \sum_S d_k x_k' h_k \right)^{-1} (T_x - \sum_S d_k x_k)' \). As the sample size grows infinitely large, note that \( g \) and so each \( h_k g \) converge to zero under mild conditions.

Folsom and Singh (2000) suggest using a series of linear approximation to find – if possible – a vector \( g \) such that weights of the form \( w_k = d_k f_k(h_k g) \), where

\[ f_k(h_k g) = \frac{\ell_k (u_k - c_k) + u_k (c_k - \ell_k) \exp(A_k h_k g)}{(u_k - c_k) + (c_k - \ell_k) \exp(A_k h_k g)}, \quad (3.1) \]
A Nearly Pseudo-optimal method for keeping calibration... 394

\[ A_k = \frac{u_k - \ell_k}{(c_k - \ell_k)(u_k - c_k)}, \]

and \(\infty \geq u_k > c_k > \ell_k \geq 0\),

satisfy the calibration equation in (1.1). Note that each \(w_k\) in this formulation is bounded between \(d_k \ell_k\) and \(d_k u_k\) and so cannot be negative. The upper bound on \(w_k\) can be arbitrarily large. In fact, \(u_k\) can be infinity.

Folsom and Singh observe that with this generalized exponential form:

\[
\frac{df_k(z)}{dz} = \frac{(u_k - f_k(z))(f_k(z) - \ell_k)}{(u_k - c_k)(c_k - \ell_k)}.
\]

Consequently, when all \(c_k = 1, f_k(0)\) and its derivative, \(df_k(0)/dz\), are unity for all \(k\). As a result, when all \(c_k = 1\), calibration weights having the form \(w_k = d_k f_k(h_k g)\) are asymptotically equivalent to linear calibration weights of the form \(w_k = d_k (1 + h_k g)\) (recall that as the sample size grows, the \(h_k g\) converge to zero). This suggests the following method for producing nearly pseudo-optimal calibration weights that are never below unity.

When possible, set \(w_k = d_k f_k(h_k g)\), where the \(f_k(.)\) are defined in equation (3.1), \(h_k = (d_k - 1)x_k, c_k = 1,\) and \(\ell_k = 1/d_k\). This setting for \(\ell_k\) ensures that no calibration weight is less than unity. Since \(\ell_k\) must be smaller than \(c_k\) for the generalized exponential form in equation (3.1) to be used, certainties (i.e., elements such that \(d_k = 1\)) need special handling. Thus, as with the pseudo-optimal weights, when \(d_k = 1\), set \(w_k = 1\).

An equivalent scheme removes certainties from both the sample and the population and sets their calibration weights to unity before computing the calibration weights for the rest of the sample with the generalized exponential form as described above. The calibration equation effectively becomes \(\sum_{S-C} w_k x_k = \sum_{U-C} x_k\), where \(C\) denote the set of certainties.

This fitting of the generalized exponential form can be done with the SUDAAN® (RTI International, 2012) procedure WTADJX by labeling the components of \(x_k\) the “calibration variables” in the CALVARS statement, and the components of \(h_k = (d_k - 1)x_k\) the “model variables” in the MODEL statement. The ADJUST=POST setting is used with the values of \(\sum_{U-C} x_k\) being the population totals in the POSTWT statement. The CENTER is set to 1 for all elements, while LOWER is \(1/d_k\). UPPER can be any value greater than 1 or it can be left unspecified.

4. DISCUSSION

Consider the prediction model in equation (1.2). If the \(\ell_k\) are uncorrelated, each with variance \(\sigma^2\), then, because equation (1.1) is satisfied, the model variance of the calibration estimator for \(T_y\) is
As Kott and Brewer (2001) point out, there are reasons to ignore the last term, \( \sum_{k \in S} w_k \sigma_k^2 - \sum_{k \in U} \sigma_k^2 \).

It is exactly zero when \( \sigma_k^2 = x_k \zeta \) for some \( \zeta \). Even if that equality does not hold, \( \sum_{k \in S} w_k \sigma_k^2 - \sum_{k \in U} \sigma_k^2 \) will be relatively small compared to \( \sum_{k \in S} w_k \sigma_k^2 \) when for almost all \( k \):

\[
w_k^2 \gg w_k,
\]
as is often the case. Thus, \( \sum_k(w_k^2 - w_k)\sigma_k^2 \) is frequently a good approximation of the model variance of \( \hat{t}_y^{CAL} \).

To estimate the model variance of \( \hat{t}_y^{CAL} \), one can replace each \( \sigma_k^2 \) with \( e_k^2 \), where \( e_k = y_k - x_k (\sum_{j \in S} h_j' x_j)^{-1} \sum_{j \in S} h_j' y_j \). Under Poisson sampling, Kott (2008) discusses why \( v_1 = \sum_k(w_k^2 - w_k)e_k^2 \) is also a good estimator of the randomization mean squared error of \( \hat{t}_y^{CAL} \). Notice that replacing the \( d_j \) in \( e_k \) by \( w_j \) (or by \( d_j [df(h, g)/dz] \) as Folsom and Singh (2000) would do) has no asymptotic impact on \( v_1 \) since \( w_j/d_j \) (and \( df(h, g)/dz \)) converges to 1 as the sample size grows arbitrarily large. Observe that to guarantee that \( v_1 \) is nonnegative, no positive \( w_k \) can be less than unity. Using the calibration procedure described in the previous section avoids that possibility.

Kott (2008) also discusses a simultaneous estimator for the model variance and randomization mean squared error of a calibration estimator under stratified simple random sampling, where \( n_\alpha \) is the sample size in stratum \( S_\alpha \):

\[
v_2 = \frac{n_\alpha}{n_\alpha - 1} \sum_{k \in S_\alpha} \left( w_k^2 - w_k \right) e_k^2 \left( \frac{\sum_{k \in S_\alpha} \left( w_k^2 - w_k \right)^{1/2} \ e_k}{n_\alpha} \right)^2 \frac{\left( \sum_{k \in S_\alpha} \left( w_k^2 - w_k \right)^{1/2} \ e_k \right)^2}{n_\alpha}.
\]

Note that for this estimator even to be defined, no positive \( w_k \) can be less than unity. Again, using the calibration procedure described here avoids that possibility.
REFERENCES


