1 Generalized Regression Estimator as a special case of Calibration Estimators

This is a summary of Section 1 of the article [2].

Let $U = \{1, 2, \ldots, N\}$ be a finite population. Let $y : U \longrightarrow \mathbb{R}$ be an $\mathbb{R}$-valued function defined on U (commonly called a "population parameter"). We will use the common notation y_i for $y(i)$. We wish to estimate $T_y := \sum_{i\in U} y_i$ via survey sampling. Let $p : \mathcal{S} \longrightarrow (0, 1]$ be our chosen sampling design, where $\mathcal{S} \subseteq \mathcal{P}(U)$ is the set of all possible samples in the design, and $\mathcal{P}(U)$ is the power set of U. For each $k \in U$, let $\pi_k := \sum_{s\ni k} p(s)$ be the inclusion probability of k under the sampling design p. We assume $\pi_k > 0$ for each $k \in U$. Then, the Horvitz-Thompson estimator

$$\widehat{T}_y^{\mathrm{HT}}(s) \;:=\; \sum_{k\in s} \frac{y_k}{\pi_k} \;=\; \sum_{k\in s} d_k y_k \;=\; \sum_{k\in U} I_{ks}\frac{y_k}{\pi_k}, \quad \text{where } d_k := \frac{1}{\pi_k} \text{ and } I_{ks} \;:=\; \begin{cases} 1, & \text{if } k \in s \\ 0, & \text{otherwise} \end{cases}$$

is well-defined and is known to be a design-unbiased estimator of T_y; in other words,

$$E_p\left[\,\widehat{T}_y^{\mathrm{HT}}\,\right] \;=\; \sum_{s\in\mathcal{S}} p(s)\cdot\widehat{T}_y^{\mathrm{HT}}(s) \;=\; \sum_{s\in\mathcal{S}} p(s)\cdot\left(\sum_{k\in U} I_{ks}\frac{y_k}{\pi_k}\right) \;=\; \sum_{k\in U}\frac{y_k}{\pi_k}\left(\sum_{s\in\mathcal{S}} p(s) I_{ks}\right) \;=\; \sum_{k\in U}\frac{y_k}{\pi_k}\,\pi_k \;=\; T_y.$$

We will call the d_k's above the *Horvitz-Thompson weights.*

Roughly, the generalized regression estimator for T_y is an estimator of the form:

$$\widehat{T}_y^{\mathrm{GREG}}(s) \;:=\; \sum_{k\in s} w_k(s) y_k,$$

where the sample-dependent "calibrated" weights $w_k(s)$ are the solution of a certain constrained minimization problem (see below) where the objective function depends on the $w_k(s)$'s and the Horvitz-Thompson weights d_k's, while the constraints involve the $w_k(s)$'s and auxiliary information. More precisely, the calibrated weights $w_k(s)$ solve the following constrained minimization problem:

Constrained Minimization Problem for the GREG calibrated weights

Conceptual framework: Let $\mathbf{x} : U \longrightarrow \mathbb{R}^{1\times J}$ be an $\mathbb{R}^{1\times J}$-valued function defined on U. We use the common notation $\mathbf{x}_k$ for $\mathbf{x}(k)$, for each $k \in U$.

Assumptions:

- The population total of $\mathbf{x}$

$$T_{\mathbf{x}} \;:=\; \sum_{k\in U} \mathbf{x}_k \in \mathbb{R}^{1\times J}$$

  is known.

- For each $s \in \mathcal{S}$, the value $(y_k, \mathbf{x}_k)$ can be observed for each $k \in s$ via the sampling procedure.

Constrained Minimization Problem: For each $k \in U$, let $q_k > 0$ be chosen. For each $s \in \mathcal{S}$, the calibrated weights $w_k(s)$, for $k \in s$, are obtained by minimizing the following objective function:

$$f_s(w_k(s)\,;\, d_k, q_k) \;:=\; \sum_{k\in s} \frac{(w_k(s) - d_k)^2}{d_k q_k}$$

subject to the (vectorial) constraint on $w_k(s)$:

$$\mathbf{h}(w_k(s)\,;\, \mathbf{x}_k, T_{\mathbf{x}}) \;:=\; -T_{\mathbf{x}} + \sum_{k\in s} w_k(s)\,\mathbf{x}_k \;=\; 0.$$

The above constrained minimization problem for the calibrated weights can be solved by the method of Lagrange Multipliers.

Solution of the Constrained Minimization Problem for the Generalized Regression Estimator calibrated weights:

Let $s \in \mathcal{S}$ be fixed. We write the objective function as

$$f(\{w_k(s) \,:\, k \in s\}) \;=\; \sum_{k\in s} \frac{(w_k(s)-d_k)^2}{d_k q_k},$$

and we write the constraints on $w_k(s)$ as:

$$h_j(\{w_k(s) \,:\, k \in s\}) \;=\; \sum_{k\in s} w_k(s)\, x_{kj} - T_{x_j} = 0, \quad j = 1,2,\ldots,J.$$

By the Method of Lagrange Multipliers, if $\mathbf{w}_0 = \{w_k(s) \,:\, k \in s\}$ is a solution to the constrained minimization problem, then $\mathbf{w}_0$ satisfies:

$$\nabla_w f(\mathbf{w}_0) \;\in\; \mathrm{span}\{\, \nabla_w h_j(\mathbf{w}_0) \,:\, j = 1,2,\ldots,J\,\}.$$

Now,

$$\frac{\partial f}{\partial w_k(s)} \;=\; \frac{2(w_k(s)-d_k)}{d_k q_k} \quad \text{and} \quad \frac{\partial h_j}{\partial w_k(s)} \;=\; x_{kj}.$$

Thus, we seek $\lambda_1, \lambda_2, \ldots, \lambda_J$ such that

$$\frac{2(w_k(s)-d_k)}{d_k q_k} \;=\; \frac{\partial f}{\partial w_k(s)} \;=\; \sum_{j=1}^{J} 2\,\lambda_j \,\frac{\partial h_j}{\partial w_k(s)} \;=\; \sum_{j=1}^{J} 2\,\lambda_j\, x_{kj},$$

which immediately implies:

$$w_k(s) \;=\; d_k \left(1 + q_k \sum_{j=1}^{J} \lambda_j x_{kj}\right).$$

Substituting the above expression for $w_k(s)$ back into the constraints yields, for each $i = 1,2,\ldots,J$:

$$-T_{x_i} + \sum_{k\in s} d_k \left(1 + q_k \sum_{j=1}^{J} \lambda_j x_{kj}\right) x_{ki} = 0,$$

which can be rearranged to be:

$$\sum_{k\in s} d_k\, x_{ki} + \sum_{j=1}^{J} \left(\sum_{k\in s} d_k q_k x_{ki} x_{kj}\right) \lambda_j = T_{x_i}$$

The preceding equation can be rewritten in vectorial form:

$$\widehat{T}^{\mathrm{HT}}_{\mathbf{x}}(s) + \mathbf{A}(s)\cdot\lambda \;=\; T_{\mathbf{x}},$$

where $\mathbf{A}(s) \in \mathbb{R}^{J\times J}$ is the symmetric matrix with entries:

$$\mathbf{A}(s)_{ij} = \sum_{k\in s} d_k q_k x_{ki} x_{kj}.$$

Assuming the matrix $\mathbf{A}(s)$ is invertible, the vector λ of Lagrange multipliers is given by:

$$\lambda \;=\; \mathbf{A}(s)^{-1}\left(T_{\mathbf{x}} \,-\, \widehat{T}^{\mathrm{HT}}_{x}(s)\right).$$

Hence, the generalized regression estimator $\widehat{T}_y^{\mathrm{GREG}}(s)$ is given by:

$$\begin{aligned}
\widehat{T}_y^{\mathrm{GREG}}(s) &= \sum_{k\in s} w_k(s) y_k = \sum_{k\in s} d_k(1+q_k\,\mathbf{x}_k^T\,\lambda)\,y_k = \sum_{k\in s} d_k y_k + \sum_{k\in s} d_k q_k(\mathbf{x}_k^T\cdot\lambda)\,y_k \\
&= \widehat{T}_y^{\mathrm{HT}}(s) + \left(\sum_{k\in s} d_k q_k y_k \cdot \mathbf{x}_k^T\right)\cdot\lambda \\
&= \widehat{T}_y^{\mathrm{HT}}(s) + \left(\sum_{k\in s} d_k q_k y_k \cdot \mathbf{x}_k^T\right)\cdot \mathbf{A}(s)^{-1}\cdot\Big(T_{\mathbf{x}} - \widehat{T}_x^{\mathrm{HT}}(s)\Big).
\end{aligned}$$

□

Example: Ratio estimator as a special case of GREG estimator (hence also of calibration estimator)

We first give the definition of the Ratio Estimator.

Definition (Ratio Estimator of the population total T_y of a population characteristic y with respect to that of another characteristic x) [See Section 5.6, [3], p.176; see also Chapter 6, [1].]
Let $U = \{1,2,\ldots,N\}$ be a finite population. Let $x, y : U \longrightarrow \mathbb{R}$ be two population characteristics. Suppose the population total $T_x := \sum_{k=1}^N x_k$ of x is known. Let $p : \mathcal{S} \subset \mathcal{P}(U) \longrightarrow (0,1]$ be a sampling design such that the inclusion probability $\pi_k := \sum_{s\ni k} p(s) > 0$, for each $k \in U$. Hence, $\widehat{T}_y^{\mathrm{HT}}(s)$ and $\widehat{T}_x^{\mathrm{HT}}(s)$ are well-defined for each sample $s \in \mathcal{S}$. The **ratio estimator**, $\widehat{T}_y^{\mathrm{R}} : \mathcal{S} \longrightarrow \mathbb{R}$, of the population T_y of y is, by definition,

$$\widehat{T}_y^{\mathrm{R}}(s) := T_x \cdot \frac{\widehat{T}_y^{\mathrm{HT}}(s)}{\widehat{T}_x^{\mathrm{HT}}(s)}, \quad \text{for each } s \in \mathcal{S}.$$

Now, we make the following:

Observation: $\widehat{T}_y^{\mathrm{GREG}} = \widehat{T}_y^{\mathrm{R}}$**, under the choice** $d_i = 1/\pi_i$ **and** $q_k = 1/x_k$
Indeed, $\mathbf{A}(s)$ is now a scalar, and we write $A(s)$, and

$$A(s) = \sum_{k\in s} d_k q_k x_k^2 = \sum_{k\in s} \frac{1}{\pi_k}\frac{1}{x_k}x_k^2 = \sum_{k\in s}\frac{x_k}{\pi_k} = \widehat{T}_x^{\mathrm{HT}}(s).$$

Next, the Lagrange multiplier $\lambda = \lambda(s)$ is now given by:

$$\lambda = \lambda(s) = \frac{1}{A(s)}\Big(T_x - \widehat{T}_x^{\mathrm{HT}}(s)\Big) = \frac{1}{\widehat{T}_x^{\mathrm{HT}}(s)}\Big[T_x - \widehat{T}_x^{\mathrm{HT}}(s)\Big] = \frac{T_x}{\widehat{T}_x^{\mathrm{HT}}(s)} - 1$$

Thus, the Generalized Regression Estimator $\widehat{T}_y^{\mathrm{GREG}}$ of T_y is given by:

$$\begin{aligned}
\widehat{T}_y^{\mathrm{GREG}}(s) &= \widehat{T}_y^{\mathrm{HT}}(s) + \left(\sum_{k\in s}\frac{1}{\pi_k}\frac{1}{x_k} y_k x_k\right)\lambda = \widehat{T}_y^{\mathrm{HT}}(s) + \widehat{T}_y^{\mathrm{HT}}(s)\left(\frac{T_x}{\widehat{T}_x^{\mathrm{HT}}(s)} - 1\right) \\
&= T_x\cdot\frac{\widehat{T}_y^{\mathrm{HT}}(s)}{\widehat{T}_x^{\mathrm{HT}}(s)} \\
&=: \widehat{T}_y^{\mathrm{R}}(s),
\end{aligned}$$

as required. □

2 Calibration Estimators

The general calibration estimator $\widehat{T}_y^{\mathrm{Cal}}$ is very similar to the generalized regression estimator $\widehat{T}_y^{\mathrm{GREG}}$, in that $\widehat{T}_y^{\mathrm{Cal}}$ is also the solution to a constrained minimization problem. The difference is that the objection function in the case of $\widehat{T}_y^{\mathrm{Cal}}$ has a more general form.

Constrained Minimization Problem for Weights of Calibration Estimators

Conceptual framework: Let $\mathbf{x} : U \longrightarrow \mathbb{R}^{1\times J}$ be an $\mathbb{R}^{1\times J}$-valued function defined on U. We use the common notation $\mathbf{x}_k$ for $\mathbf{x}(k)$, for each $k \in U$.

Assumptions:

- The population total of $\mathbf{x}$
$$T_{\mathbf{x}} \;:=\; \sum_{k\in U} \mathbf{x}_k \in \mathbb{R}^{1\times J}$$
is known.
- For each $s \in \mathcal{S}$, the value $(y_k, \mathbf{x}_k)$ can be observed for each $k \in s$ via the sampling procedure.
- For each $k \in U$, $G_k(w\,;\,d)$ is an $\mathbb{R}$-valued function which satisfies:
    1. For each $d > 0$, $G_k(w\,;\,d)$ is non-negative, differentiable with respect to w, strictly convex in w, defined on an open interval $D_k(d)$ containing d, and such that $G_k(k\,;\,k) = 0$.
    2. $g_k(w\,;\,d) := \dfrac{\partial G_k(w\,;\,d)}{\partial w}$ is continuous in w and maps $D_k(d)$ bijectively onto its image.

Constrained Minimization Problem: For each $k \in U$, let $q_k > 0$ be chosen. For each $s \in \mathcal{S}$, the calibrated weights $w_k(s)$, for $k \in s$, are obtained by minimizing the following objective function:

$$f_s(w_k(s)\,;\,d_k) \;:=\; \sum_{k\in s} G_k(w_k(s)\,;\,d_k)$$

subject to the (vectorial) constraint on $w_k(s)$:

$$\mathbf{h}(w_k(s)\,;\,\mathbf{x}_k, T_{\mathbf{x}}) \;:=\; -T_{\mathbf{x}} + \sum_{k\in s} w_k(s)\,\mathbf{x}_k \;=\; 0.$$

Solution of the Constrained Minimization Problem for weights of calibration estimators:

Let $s \in \mathcal{S}$ be fixed. We write the objective function as

$$f(\{w_k(s)\,:\,k\in s\}) \;=\; \sum_{k\in s} G_k(w_k(s)\,;\,d_k),$$

and we write the constraints on $w_k(s)$ as:

$$h_j(\{w_k(s)\,:\,k\in s\}) \;=\; \sum_{k\in s} w_k(s)\,x_{kj} - T_{x_j} = 0, \quad j = 1,2,\ldots,J.$$

By the Method of Lagrange Multipliers, if $\mathbf{w}_0 = \{w_k(s)\,:\,k\in s\}$ is a solution to the constrained minimization problem, then $\mathbf{w}_0$ satisfies:

$$\nabla_w f(\mathbf{w}_0) \;\in\; \mathrm{span}\{\,\nabla_w h_j(\mathbf{w}_0)\,:\,j = 1,2,\ldots,J\,\}.$$

Now,

$$\frac{\partial f}{\partial w_k(s)} \;=\; g_k(w_k(s)\,;\,d_k) \quad\text{and}\quad \frac{\partial h_j}{\partial w_k(s)} \;=\; x_{kj}.$$

Thus, we seek $\lambda_1, \lambda_2, \ldots, \lambda_J$ such that

$$g_k(w_k(s)\,;\,d_k) \;=\; \frac{\partial f}{\partial w_k(s)} \;=\; \sum_{j=1}^{J} \lambda_j \frac{\partial h_j}{\partial w_k(s)} \;=\; \sum_{j=1}^{J} \lambda_j\, x_{kj} \;=\; \mathbf{x}_k^T \cdot \lambda.$$

By hypothesis, $g_k(\,\cdot\,;\,d_k)$ is bijective on $D_k(d_k)$. We denote its inverse by $g_k^{-1}(\,\cdot\,;\,d_k)$. Thus, we have

$$w_k(s) \;=\; g_k^{-1}\big(\,\mathbf{x}_k^T \cdot \lambda\,;\,d_k\,\big) \;=\; d_k \cdot F_k\big(\,\mathbf{x}_k^T \cdot \lambda\,\big)\,,$$

where $F_k(\,\cdot\,) := \dfrac{1}{d_k}\, g_k^{-1}(\,\cdot\,;\,d_k\,)$. The constraint equation can thus be rewritten as follows: For each $j = 1, 2, \ldots, J$,

$$\begin{aligned}
\sum_{k\in s} w_k(s)\, x_{kj} &= T_{x_j} \\
\sum_{k\in s} d_k F_k(\mathbf{x}_k^T \cdot \lambda)\, x_{kj} &= T_{x_j} \\
\sum_{k\in s} d_k \left[F_k(\mathbf{x}_k^T \cdot \lambda) - 1\right] x_{kj} &= T_{x_j} - \sum_{k\in s} d_k x_{kj} \;=\; T_{x_j} - \widehat{T}_{x_j}^{\mathrm{HT}}(s).
\end{aligned}$$

In vectorial form, we have:

$$\sum_{k\in s} d_k \left[F_k(\mathbf{x}_k^T \cdot \lambda) - 1\right] \mathbf{x}_k \;=\; T_{\mathbf{x}} - \widehat{T}_{\mathbf{x}}^{\mathrm{HT}}(s).$$

Note that, for each obtained sample $s \in \mathcal{S}$, the Lagrange multiplier vector $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_J) \in \mathbb{R}^J$ is the only unknown quantity in the above constraint equation. We now assume the above vectorial constraint equation is solvable for λ, and denote its solution by λ^*. Then, the **calibration estimator** of T_y is given by:

$$\widehat{T}_y^{\mathrm{Cal}}(s) \;=\; \sum_{k\in s} d_k \cdot F_k\big(\,\mathbf{x}_k^T \cdot \lambda^*\,\big) \cdot y_k$$

□

Observation:

If y is a deterministic linear function of $\mathbf{x}$, i.e. there exists some $\beta \in \mathbb{R}^J$ such that for each $k \in U$, we have $y_k = \beta^T \cdot \mathbf{x}_k$, then $\widehat{T}_y^{\mathrm{Cal}}(s) = T_y$, for each $s \in \mathcal{S}$.

Proof

$$\begin{aligned}
\widehat{T}_y^{\mathrm{Cal}}(s) &= \sum_{k\in s} d_k \cdot F_k\big(\mathbf{x}_k^T \cdot \lambda^*\big) \cdot y_k \;=\; \sum_{k\in s} d_k \cdot F_k\big(\mathbf{x}_k^T \cdot \lambda^*\big) \cdot \big(\beta^T \cdot \mathbf{x}_k\big) \;=\; \beta^T \cdot \left[\sum_{k\in s} d_k \cdot F_k\big(\mathbf{x}_k^T \cdot \lambda^*\big) \cdot \mathbf{x}_k\right] \\
&= \beta^T \cdot T_{\mathbf{x}} \;=\; \beta^T \cdot \left(\sum_{k\in U} \mathbf{x}_k\right) \;=\; \sum_{k\in U} \beta^T \cdot \mathbf{x}_k \;=\; \sum_{k\in U} y_k \\
&=: T_y,
\end{aligned}$$

where the third equality holds because d_k and $F_k(\,\cdot\,)$ are scalars, and the fourth equality holds since λ^* is a solution of the vectorial constraint equation. □

References

[1] Cochran, W. G. *Sampling Techniques*, third ed. Wiley Series in Probability and Mathematical Statistics. John-Wiley & Sons, 1977.

[2] Deville, J.-C., and Särndal, C.-E. Calibration estimators in survey sampling. *Journal of the American Statistical Association 87*, 418 (1992), 376–382.

[3] Särndal, C.-E., Swensson, B., and Wretman, J. *Model Assisted Survey Sampling*, third ed. Springer Series in Statistics. Springer, 1992.