A Simple and Efficient Way of Rounding Calibration Weights

Yves Tillé
University of Neuchatel

July 10, 2019

Abstract

Sartore et al. (2019) have proposed a method to round calibration weights to integer values. Their method is based on a discrete coordinate descent algorithm. We propose a much simpler method based on balanced sampling that achieves the same aim. This method provides random, unbiased and balanced rounded weights.
Keywords: balanced sampling, cube method, calibration, weights.

1 The method

Suppose that a random sample S of size n is selected from a finite population U of size N with inclusion probabilities $\pi_1, \dots, \pi_N$. The basic Horvitz-Thompson (HT) weights are $d_k = 1/\pi_k$. For the population total $Y = \sum_{k \in U} y_k$, the HT-estimator

$$\widehat{Y} = \sum_{k \in S} d_k y_k$$

is unbiased (Horvitz and Thompson, 1952). Besides, suppose also that a vector $\mathbf{x}_k$ of p auxiliary variables is known for each sample unit. The vector of population totals of $\mathbf{x}_k$ is supposed to be available from a census or a register.

The first non-integr calibration weights w_k can be computed by means of the Deville-Särndal methodology (Deville and Särndal, 1992). The weights satisfy

$$\sum_{k \in S} w_k \mathbf{X}_k = \sum_{k \in U} \mathbf{x}_k.$$

Deville and Särndal have proposed several solutions to impose lower and upper bounds to the weights by instance a logistic calibration function can be used. It is also possible at this stage to use a soft calibration by using a penalized method like the ridge regression or the Lasso (Tibshirani, 1996; Rao and Singh, 1997; Théberge, 1999, 2000; Beaumont and Bocci, 2008; Guggemos and Tillé, 2010). It is further desirable that all weights be greater or equal to 1 because a sampled unit represents itself at least.

In order to round the weights, we propose to proceed as follows.

- Compute the integer values $\phi_k = w_k - \lfloor w_k \rfloor$ for all $k \in S$.
- Select a subsample from S with probabilities ϕ_k. Let $c_k = 1$ if unit k is selected in this subsample and $c_k = 0$ otherwise. Besides, the subsample must satisfy at best the following balancing equations:

$$\sum_{k \in S} (\lfloor w_k \rfloor + c_k) \mathbf{x}_k \approx \sum_{k \in S} w_k \mathbf{x}_k = \sum_{k \in U} \mathbf{x}_k. \quad (1)$$

- Compute the rounded weights $\tilde{w}_k = w_k + c_k$.

Equation (1) is equivalent to

$$\sum_{k\in S} c_k \mathbf{x}_k \approx \sum_{k\in S} \phi_k \mathbf{x}_k$$

or also to

$$\sum_{k\in S} \frac{c_k \tilde{\mathbf{x}}_k}{\phi_k} \approx \sum_{k\in S} \tilde{\mathbf{x}}, \qquad (2)$$

where $\tilde{\mathbf{x}} = \phi_k \mathbf{x}_k$. The selection of the subsample satisfying at best Equation (2) can be realized by using the cube method described in Deville and Tillé (2004).

The method has several advantage. The rounding is done randomly. The rounding does not increase the bias because $\mathrm{E}(\tilde{w}_k|S) = w_k$. So,

$$\mathrm{E}\left(\sum_{k\in S} \widetilde{w}_k y_k\right) = \mathrm{E}\left(\sum_{k\in S} w_k y_k\right).$$

Simulations can simply evaluate the additional variance due to the rounding. A large number of random rounding can be generated and the variance due to the rounding can be approximated.

2 An example

A sample of size $n = 14$ has been selected from a population of size $N = 1000$ with equal inclusion probabilities $\pi_k = n/N, k \in U$. There are three calibration variables given in Table 1. With the raking-

Table 1: Values of the three auxiliary variables in the sample

x_1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
x_2	1	2	3	4	5	6	7	8	9	10	11	12	13	14
x_3	1	4	9	16	25	36	49	64	81	100	121	144	169	196

ratio method, the vector of calibrated weights is $\mathbf{w} = (57.20, 59.35, 61.51, 63.70, 65.89, 68.10, 70.32, 72.54, 74.75, 76.96, 79.16, 81.35, 83.51, 85.65)^\top$. Ten realizations of for the rounded weights are given in Table 2. The estimated totals for the auxiliary variables

$$\widehat{\widetilde{X}}_j = \sum_{k\in S} \tilde{w}_k x_{k1}$$

are given in Table 3 for the ten rounded weighting systems. A very simple R code to round the calibrated weights are given in Section 3.

3 R code

```
# require sampling package
library(sampling)
# generate data
n=14;N=1000;pik=rep(n/N,n);d=1/pik
Xs=cbind(rep(1,n),1:n,(1:n)^2)
T=c(1000,8000,80000)
# calibrate
w=d*calib(Xs,d,T,method="raking")
# round
phi=w-floor(w);Xtilde=X*phi
wtilde=floor(w)+samplecube(Xtilde,phi,comment=FALSE)
```

Table 2: Exemple of 10 randomly rounded weights

$\tilde{\mathbf{w}}_1$	$\tilde{\mathbf{w}}_2$	$\tilde{\mathbf{w}}_3$	$\tilde{\mathbf{w}}_4$	$\tilde{\mathbf{w}}_5$	$\tilde{\mathbf{w}}_6$	$\tilde{\mathbf{w}}_7$	$\tilde{\mathbf{w}}_8$	$\tilde{\mathbf{w}}_9$	$\tilde{\mathbf{w}}_{10}$
57	58	57	58	58	57	57	57	57	57
59	59	60	60	59	59	60	59	59	59
62	61	62	61	62	62	61	62	62	62
64	64	64	63	63	64	63	64	64	64
66	66	65	66	66	65	66	66	66	66
68	68	68	68	68	68	68	68	68	68
70	71	70	70	70	71	70	71	70	70
73	72	73	73	72	73	73	72	72	72
74	74	75	75	75	74	75	75	75	75
77	77	77	77	77	77	77	77	77	77
79	79	79	79	79	79	79	79	79	80
82	82	81	82	82	81	81	81	82	81
84	83	83	83	84	84	84	84	83	83
85	86	86	85	85	86	86	85	86	86

Table 3: Values of the estimated auxiliary variables with the ten rounded weighting systems

$\widehat{\widetilde{X}}_1$	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
$\widehat{\widetilde{X}}_1$	8000	7998	7995	7992	7998	8004	8006	7996	8002	8001
$\widehat{\widetilde{X}}_1$	79992	79996	79935	79884	79994	80068	80104	79914	80036	80013

References

Beaumont, J.-F. and Bocci, C. (2008). Another look at ridge calibration. *Metron*, 66(1):5–20.

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

Deville, J.-C. and Tillé, Y. (2004). Efficient balanced sampling: The cube method. *Biometrika*, 91:893–912.

Guggemos, F. and Tillé, Y. (2010). Penalized calibration in survey sampling: Design-based estimation assisted by mixed models. *Journal of Statistical Planning and Inference*, 140(11):3199–3212.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47:663–685.

Rao, J. N. K. and Singh, A. C. (1997). A ridge-shrinkage method for range-restricted weight calibration in survey sampling. In *Section on Survey Research Methods, American Statistical Association, Washington DC, 57-65.*

Sartore, L., Toppin, K., Young, L., and Spiegelman, C. (2019). Developing integer calibration weights for census of agriculture. *Journal of Agricultural, Biological and Environmental Statistics*, 24(1):26–48.

Théberge, A. (1999). Extension of calibration estimators in survey sampling. *Journal of the American Statistical Association*, 94:635–644.

Théberge, A. (2000). Calibration and restricted weights. *Survey Methodology*, 26:99107.

Tibshirani, R. J. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, B73(3):267–288.