Relationship between balanced sampling and calibrated estimator

Ieva Dirdaitė

Vilnius Gediminas technical university

2015-08-24

Content

The aim of the study

To compare the results of estimation of a finite population total in the case of two sampling strategies:

1. balanced sample of clusters with Horvitz-Thomson estimator of total,
2. simple random sample of clusters with the calibrated estimator of total.

Auxiliary information for both cases is the same.

Data of Lithuanian Labour force survey is used for simulation.

Notations

Let $U = \{1, 2, ..., N\}$ - finite population.
Study variable - y with values $y_1, ..., y_N$.

Parameter of interest - population total $t_y = \sum_{k=1}^{N} y_K$.

$\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_p$ - auxiliary variables known for all population elements.
Let $s \subset U$ - n size probability sample.
$\pi_k = P(s : k \in s)$ - first order inclusion probability of element $k,\ k = 1, 2, ..., N$.
$\mathbf{I} = (I_1, ..., I_N)'$ - sample vector, with

$$I_k = \begin{cases} 1, & if \ \ k \in s, \\ 0, & if \ \ k \notin s. \end{cases}$$

Horvitz–Thomson unbiased estimator of total $\hat{t}_y^{HT} = \sum_{k \in s} \frac{y_k}{\pi_k}$ will be used.
$d_k = \frac{1}{\pi_k}$ are called sampling weights.

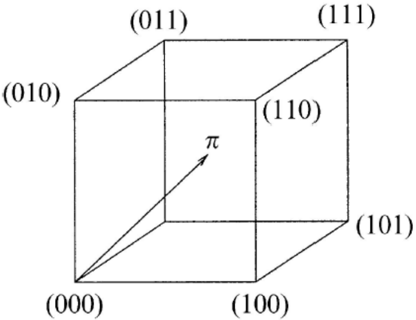
Balanced sampling

A sampling design is said to be balanced with respect to the auxiliary variables $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_p$ if and only if it satisfies the equations given by

$$\hat{t}_{\mathbf{x}_j}^{HT} = \sum_{k=1}^{N} \frac{x_{kj} I_k}{\pi_k} = \sum_{k \in s} \frac{x_{kj}}{\pi_k} = t_{x_j}. \quad (1)$$

Balanced sampling. Cube method

Cube method - algorithm to select balanced sample.



Geometric example of possible samples when population size is $N = 3$.

Calibrated estimator

Estimator of the total t_y $\hat{t}^w = \sum_{k \in s} w_k y_k$. is called calibrated if its weights w_k for any fixed sample s satisfy conditions:

1. they differ as little as possible from the design weights d_k:

$$L(w_k, d_k, k \in s) = \sum_{k \in s} \frac{(w_k - d_k)^2}{d_k q_k} \to min,$$

   q_k, $k \in U$ - freely choosen constants,

2. satisfy calibration equation

$$\hat{t}^w_{\mathbf{x}_j} = \sum_{k \in s} w_k x_{kj} = t_{\mathbf{x}_j},$$

   $j = 1, 2, ..., p.$

Study population

1. Labour Force survey data of statistics Lithuania.
2. Population of 21318 individuals, 11236 households (clusters).
3. Study variable for individual: unemployed (1), otherwise (0).
4. The aim - to estimate the number of unemployeed people (total).
5. Three auxiliary variables - sex, living place and age.
6. Sampling designs: balanced sampling of clusters with inclusion probabilities proportional to size and simple random sampling of clusters.
7. Three different sample sizes.
8. Six different sampling strategies.
9. Simulation is repeated for 10 samples.

Sampling strategies

1. Balanced sample of clusters and Horvitz and Thomson estimator.
2. Simple random sample of clusters and calibrated estimator.
3. Balanced sample of clusters, nonresponse and calibrated estimator.
4. Balanced sample of clusters, nonresponse and Horvitz and Thomson estimator.
5. Simple random sample of clusters, nonresponse and calibrated estimator.
6. Balanced sample of clusters and calibrated estimator.

Results

Estimators of total and accuracy measures when $B = 10$:

- total

$$\bar{\hat{t}}_y = \frac{1}{B} \sum_{k=1}^{B} \hat{t}_{yk},$$

- variance

$$\overline{\widehat{Var}}(\hat{t}_y) = \frac{1}{B} \sum_{k=1}^{B} \widehat{Var}(\hat{t}_{yk}),$$

- bias

$$\widehat{Bias}(\hat{t}_y) = \bar{\hat{t}}_y - t_y,$$

- relative mean squared error

$$\widehat{rMSE}(\hat{t}_y) = \frac{\sqrt{\widehat{Bias}^2(\hat{t}_y) + \overline{\widehat{Var}}(\hat{t}_y)}}{\bar{\hat{t}}_y}.$$

Results

	Balanced sampling	Simple random sampling and calibration	Balanced sampling, nonresponse and calibration	Balanced sampling and calibration	Simple random sampling, nonresponse and calibration	Balanced sampling and nonresponse
Total	1595	1730	1729	1740	1692	1992
Variance	169191	137509	208781	178696	208339	228248
Relative mean squared error	0.272	0.214	0.264	0.243	0.271	0.273
Bias	-137	-2	-3	8	-40	260

Results with sample size $n = 100$.

Results

	Balanced sampling	Simple random sampling and calibration	Balanced sampling, nonresponse and calibration	Balanced sampling and calibration	Simple random sampling, nonresponse and calibration	Balanced sampling and nonresponse
Total	1716	1768	1621	1672	1768	1899
Variance	17168	12863	17977	17166	18508	21374
Relative mean squared error	0.077	0.067	0.107	0.086	0.080	0.117
Bias	16	-36	111	60	-36	-167

Results with sample size $n = 1000$.

Results

	Balanced sampling	Simple random sampling and calibration	Balanced sampling, nonresponse and calibration	Balanced sampling and calibration	Simple random sampling, nonresponse and calibration	Balanced sampling and nonresponse
Total	1725	1744	1727	1727	1714	1924
Variance	2186	1530	2417	2157	2408	2914
Relative mean squared error	0.027	0.023	0.029	0.027	0.030	0.104
Bias	7	-12	5	5	18	-192

Results with sample size $n = 5000$.

Conclusions

1. The results of simple random sampling and calibrated estimator strategy were best for all three sample sizes.
2. Small and big samples with nonresponse: balanced sampling and calibrated estimator gave better results than other two strategies.
3. Medium samples with nonresponse: the results of simple random sampling and calibrated estimator were better than other two strategies' results.
4. The results of balanced sampling and nonresponse for all sample sizes were worse than the results of other 5 strategies.
5. For small samples the results can be improved by using auxiliary information in both - sample selection and estimation stages, however for big samples it is enough to use auxiliary information in just one of the stages.

References

1. Deville J.-C., Särndal C.-E., Calibrated Estimators in Survey Sampling, *Journal of the American Statistical Association*, 1992, 87, p. 376 - 382.
2. Särndal C.-E., Swensson B., Wretman J., *Model Assisted Survey Sampling*, New York: Springer-Verlag, 1992.
3. Tillé Y., *Sampling Algorithms*, New York: Springer, 2006.

Thank you for your attention!