In memory of Jean-Claude Deville

Alina Matei

Institute of Statistics, University of Neuchâtel, Switzerland

IASS Webinar - 25 May 2022

Jean-Claude Deville (1944-2021)

It was with great sadness that we learned of the death of Jean-Claude, who passed away in November 2021 at the age of 77.

Life and work

- Jean-Claude studied at Ecole Polytechnique, and next to ENSAE (National School of Statistics and Economic Administration), both very prestigious and selective institutions in France.
- He joined INSEE (National Institute of Statistics and Economic Studies) in 1968 until 1998, when he became head of Survey Statistics Laboratory at ENSAI (National School of Statistics and Information Analysis), from where he retired in 2010.
- Jean-Claude was an elected member of the International Statistical Institute (1979), and member of the Council of the International Association of Survey Statisticians from 1993 to 1997.

Life and work

- He was a remarkable survey statistician, but he also did important contributions in the beginning of his career in demography, functional analysis and data analysis.
- He started to work in survey statistics in 1981 and published his first article about it at the age of 42.
- Jean-Claude was one of the world's leading experts in survey statistics, with huge contributions.
- In 2018, Jean-Claude received the Waksberg price, accorded by the journal 'Survey Methodology' each year to a prominent survey statistician.
  See https://community.amstat.org/surveyresearchmethodssection/programs/awards/waksberg

Main contributions to survey statistics

- His papers/contributions reflect the mixture of theory and practice, by focusing on statistical issues motivated by real-world problems.
- We can 'track' the history of his contributions on the web page of Statistical Methodology Days of INSEE; see http://jms-insee.fr/jcdeville_prixwaksberg2018/
- Main topics:
  - calibration/generalized calibration;
  - balanced sampling;
  - weight share method;
  - maximum entropy sampling designs;
  - non-response treatment;
  - variance estimation, etc.

Two important contributions

Calibration Estimators in Survey Sampling

JEAN-CLAUDE DEVILLE and CARL-ERIK SÄRNDAL*

This article investigates estimation of finite population totals in the presence of univariate or multivariate auxiliary information. Estimation is equivalent to attaching weights to the survey data. We focus attention on the several weighting systems that can be associated with a given amount of auxiliary information and derive a weighting system with the aid of a distance measure and a set of calibration equations. We briefly mention an application to the case in which the information consists of known marginal counts in a two- or multi-way table, known as *generalized raking*. The general regression estimator (GREG) was conceived with multivariate auxiliary information in mind. Ordinarily, this estimator is justified by a regression relationship between the study variable y and the auxiliary vector $\mathbf{x}$. But we note that the GREG can be derived by a different route by focusing instead on the weights. The ordinary sampling weights of the kth observation is $1/\pi_k$, where π_k is the inclusion probability of k. We show that the weights implied by the GREG are as close as possible, according to a given distance measure, to the $1/\pi_k$ while respecting side conditions called *calibration equations*. These state that the sample sum of the weighted auxiliary variable values must equal the known population total for that auxiliary variable. That is, the calibrated weights must give perfect estimates when applied to each auxiliary variable. That is a consistency check that appeals to many practitioners, because a strong correlation between the auxiliary variables and the study variable means that the weights that perform well for the auxiliary variable also should perform well for the study variable. The GREG uses the auxiliary information efficiently, so the estimates are precise; however, the individual weights are not always without reproach. For example, negative weights can occur, and in some applications this does not make sense. It is natural to seek the root of the dissatisfaction in the underlying distance measure. Consequently, we allow alternative distance measures that satisfy only a set of minimal requirements. Each distance measure leads, via the calibration equations, to a specific weighting system and thereby to a new estimator. These estimators form a family of *calibration estimators*. We show that the GREG is a first approximation to all other members of the family; all are asymptotically equivalent to the GREG, and the variance estimator already known for the GREG is recommended for use in any other member of the family. Numerical features of the weights and ease of computation become more than anything else the bases for choosing between the estimators. The reasoning is applied to calibration on known marginals of a two-way frequency table. Our family of distance measures leads in this case to a family of *generalized raking procedures*, of which classical raking ratio is one.

KEY WORDS: Calibration; Multivariate auxiliary information; Raking; Regression estimators.

(a) 1992

Efficient balanced sampling: The cube method

By JEAN-CLAUDE DEVILLE

Laboratoire de Statistique d'Enquête, CREST–ENSAI,
École Nationale de la Statistique et de l'Analyse de l'Information, rue Blaise Pascal,
Campus de Ker Lann, 35170 Bruz, France
deville@ensai.fr

AND YVES TILLÉ

Groupe de Statistique, Université de Neuchâtel, Espace de l'Europe 4, Case postale 805,
2002 Neuchâtel, Switzerland
yves.tille@unine.ch

SUMMARY

A balanced sampling design is defined by the property that the Horvitz–Thompson estimators of the population totals of a set of auxiliary variables equal the known totals of these variables. Therefore the variances of estimators of totals of all the variables of interest are reduced, depending on the correlations of these variables with the controlled variables. In this paper, we develop a general method, called the cube method, for selecting approximately balanced samples with equal or unequal inclusion probabilities and any number of auxiliary variables.

Some key words: Calibration; Poststratification; Quota sampling; Sampling algorithm; Stratification; Sunter's method; Unequal selection probabilities.

(b) 2004

Framework

- Finite population with labels $U = \{1, 2, \ldots, k, \ldots, N\}$.
- A random sample s selected from U.
- $\pi_k = P(k \in s)$ first-order inclusion probability of unit k in the sample;
- The variable of interest is y (y_k represents the value of y for the survey unit k).
- A vector $\mathbf{x}_k = (x_{k1}, x_{k2}, \ldots, x_{kq})'$ of auxiliary variables which is assumed to be available for each $k \in U$ in the population.
- We want to estimate the unknown population mean or total of y.

Calibration - see Deville and Särndal (1992)

- A set of weights $\widetilde{d}_k$ is constructed by modifying the initial weights $d_k = 1/\pi_k, k \in s$, such that a distance between d_k and $\widetilde{d}_k$ is minimized, while satisfying the **calibration equations**

$$\sum_{k \in s} \widetilde{d}_k \mathbf{x}_k = \sum_{k \in U} \mathbf{x}_k = \text{known}. \tag{1}$$

- The final weights are

$$\widetilde{d}_k = d_k F(\boldsymbol{\lambda}^\top \mathbf{x}_k) = F(\boldsymbol{\lambda}^\top \mathbf{x}_k)/\pi_k,$$

where $F(.)$ is a function with suitable properties, like $F(\boldsymbol{\lambda}^\top \mathbf{x}_i) = 1 + \boldsymbol{\lambda}^\top \mathbf{x}_i$ or $F(\boldsymbol{\lambda}^\top \mathbf{x}_i) = exp(\boldsymbol{\lambda}^\top \mathbf{x}_i)$, and $\boldsymbol{\lambda}$ is a vector to determine by solving the calibration equations (1).

- A calibration estimator of $Y = \sum_{k \in U} y_k$ is

$$\widehat{Y}_{\text{cal}} = (\sum_{k \in s} \widetilde{d}_k y_k)/N.$$

Balanced sampling - see Deville and Tillé (2004)

According to Deville and Tillé (2004), a sampling design is said to be balanced on auxiliary variables $\mathbf{x}_k$ if

$$\sum_{k \in s} \frac{\mathbf{x}_k}{\pi_k} = \sum_{k \in U} \mathbf{x}_k = \text{known.} \tag{2}$$

The cube method allows to select a random sample s such that Equations (2) are fulfilled.

A similar problem?

Calibration:

$$\sum_{k\in s} \widetilde{d}_k \mathbf{x}_k = \sum_{k\in U} \mathbf{x}_k = \text{known}.$$

Balanced sampling:

$$\sum_{k\in s} \frac{\mathbf{x}_k}{\pi_k} = \sum_{k\in U} \mathbf{x}_k = \text{known}.$$

Representativeness

In 2011, Jean-Claude was keenyote speaker at the Third Baltic-Nordic Conference on Survey Statistics - BaNoCoSS-2011
Lecture : 'Generalized calibration, balanced sampling and application to nonresponse' (page 12; see http://www.mathstat.helsinki.fi/msm/banocoss/2011/ISpeakers.html)

12

5-Representativeness à la Hajek (Sampling from a finite population, chap 16).

A strategy (plan + weights) is representative with respect of the vector of variables x if for any sample s the weights w verify $\Sigma_s\, w_k\, x_k = X = \Sigma_U\, x_k$.
The total X has to be known and an **auxiliary information.**

When this information comes from a source disconnected from the sampling frame, a way is to modify slightly the unbiased HT weights such that the equality is verified. This is **calibration**.

When the information x_k is available in the frame, a way is to keep the unbiased HT weights and to find some sampling algorithm such that the equality is verified . This means find a method to get a **random balanced sample** with given inclusion probabilities.

Properties of those strategies are evaluated using the **mean square error** criterion and, in some sense, **conditional inference**.

His personality

- A very good mathematician.
- Hobbies: chess, jazz, movies, basket-ball.
- During a panel discussion on 'Can we teach survey methodology?' in 1995,

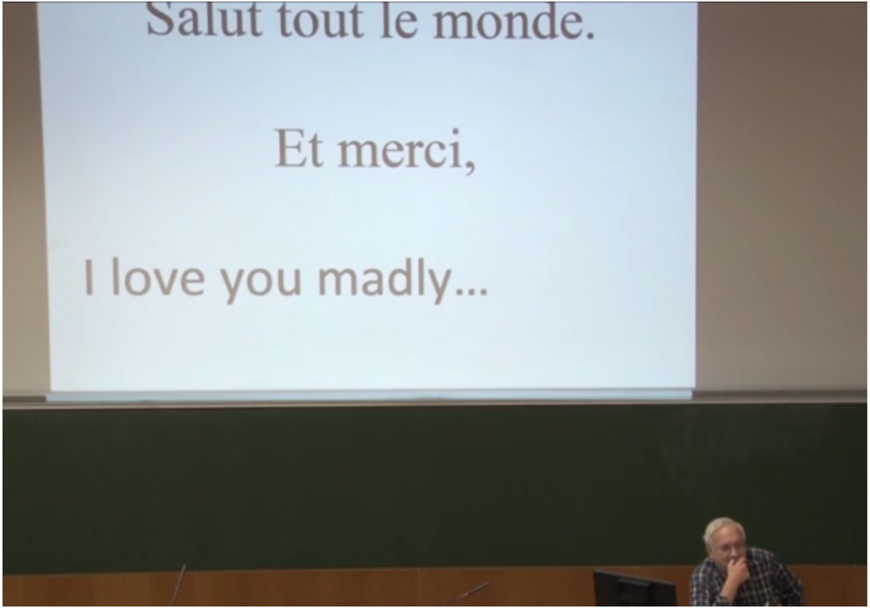
Jean-Claude declared:

'As far as the usefulness of mathematics is concerned, it intervenes as other useful techniques in statistical methodology. Without mathematics, some areas of methodology may be unattainable (balanced sampling, robust statistics for example) and one must be satisfied with 'rustic' solutions.
Jankélévitch said: "Yes, one can live without music, love and philosophy, but ...much less well!"
In the same way, one can do statistics without mathematics, but much less well.'

Good bye, Jean-Claude...

A prominent statistician, a passionate speaker, a very cultured person... Those who had the privilege of meeting and talking with him know how brilliant he was !

Thank you for all and good bye, Jean-Claude !



2012: Francophone Conference on Surveys, Rennes, France

References

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.