Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

Jean-Claude Deville's contributions to survey theory and official statistics

by Pascal Ardilly, David Haziza, Pierre Lavallée and Yves Tillé

Release date: January 3, 2024

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at infostats@statcan.gc.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-514-283-9350

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Published by authority of the Minister responsible for Statistics Canada

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, December 2023
Vol. 49, No. 2, pp. 257-298
Statistics Canada, Catalogue No. 12-001-X

Jean-Claude Deville's contributions to survey theory and official statistics

Pascal Ardilly, David Haziza, Pierre Lavallée and Yves Tillé[1]

Abstract

Jean-Claude Deville, who passed away in October 2021, was one of the most influential researchers in the field of survey statistics over the past 40 years. This article traces some of his contributions that have had a profound impact on both survey theory and practice. This article will cover the topics of balanced sampling using the cube method, calibration, the weight-sharing method, the development of variance expressions of complex estimators using influence function and quota sampling.

Key Words: Calibration; Balanced sampling; Quota sampling; Variance estimation; Cube method; Weight-share method.

1. Introduction

Jean-Claude Deville, who passed away in October 2021, will undoubtedly leave an important legacy in survey statistics. For more than 40 years, as part of the National Institute of Statistics and Economic Studies (INSEE) and then *École nationale de la statistique et de l'analyse de l'information* (ENSAI) in France, he developed major innovations including calibration techniques; balanced sampling; indirect sampling and weight sharre methods; variance calculation, particularly for complex estimators; processing of non-ignorable non-response; and quota surveys. That being said, he has worked in all survey fields and beyond. His exceptional productivity is mainly attributable to a very fruitful imagination combined with a remarkable mastery of mathematical tools. It was also fed by the concrete cases encountered at INSEE, which, like all national statistical institutes, was constantly confronted with various constraints and obstacles that had to be overcome, generally quickly and at a low cost. As head of the statistical methodology unit, he had to meet the technical challenges presented to him as they arose.

The following is an overview of Jean-Claude Deville's developments, all of which have moved on to posterity and can be found in depth in the many articles he published throughout his career, some shared with colleagues with whom he had privileged relationships. Clearly, some of his developments have found considerable international applications since their publication. There has even been an "industrial" implementation for calibration, the development of which was designed with another prestigious statistician, Carl-Erik Särndal.

2. Unequal and balanced probability sampling

2.1 Innovations in sampling algorithms

A sample is said to be balanced on a variable if the Horvitz-Thompson estimators of the totals calculated from a sample are equal to or nearly equal to the population total $U = \{1, \ldots, k \ldots, N\}$. Formally, suppose

1. Pascal Ardilly, L'Institut national de la statistique et des études économiques (France); David Haziza, University of Ottawa (Canada). E-mail: dhaziza@uottawa.ca; Pierre Lavallée, Statistics Canada (retired); Yves Tillé, Université de Neuchâtel (Suisse).

that a vector of auxiliary variables $\mathbf{z}_k = (z_{k1}, \ldots, z_{kQ})^\top$ is known for all population units. Sample S is balanced on $\mathbf{z}_k$ if

$$\sum_{k \in S} \frac{\mathbf{z}_k}{\pi_k} = \sum_{k \in U} \mathbf{z}_k,$$

where π_k is the inclusion probability, i.e., the probability that unit k is selected in the random sample S.

The idea of selecting a balanced sample dates back to the very beginning of survey theory. Kiær (1896, 1899, 1903, 1905) was the first to propose what he called "representative counts". It is actually a selection of samples by quota. However, it was Gini and Galvani (1929) who first selected a balanced sample in official statistics. They selected 29 Italian districts (*circondari*) out of 214 to match several population averages as well as possible (Langel and Tillé, 2011; Tillé, 2016; Brewer, 2013). This method was harshly criticized by Jerzy Neyman because the sample was not randomly selected (see Bellhouse, 1988). Yates (1949) and Thionet (1953) proposed methods for which a sample is selected and then improved by successively replacing units to approach a balancing setting. Hájek (1964, 1981) proposed using rejective sampling, which consists of selecting a series of samples until a sufficiently balanced sample is obtained. However, this method has the drawback of changing the inclusion probabilities of units without being able to calculate them accurately afterwards (Choudhry and Singh, 1979; Dupačová, 1979; Fuller, 2009; Legg and Yu, 2010; Boistard, Lopuhaä and Ruiz-Gazen, 2012; Fuller, Legg and Li, 2017).

Jean-Claude Deville quickly became interested in sampling methods. In 1987, he published a book chapter with Jean-Marie Grosbras in which sampling methods were described and compared (Deville and Grosbras, 1987). The following year, along with Nicole Roth, they proposed a first balanced sampling method (Deville, Grosbras and Roth, 1988). The method applies only to equal probability of selection. The idea is to divide the variable space into quadrants and select one unit in the quadrant at each step that will contribute the most to achieving balancing. In the proceedings of the Örebro Conference held at Statistics Sweden in 1992, Jean-Claude Deville expressed his views on the three facets of the use of auxiliary information, namely constrained samples (i.e., balanced), conditional inference and weighting (Deville, 1992).

In parallel with this work, Jean-Claude Deville conducted very specific research on sampling issues. He proposed a formalization of sampling in a continuous population (Deville, 1989) long before the publication by Cordy (1993), who is often cited as the first reference in this field. He also proposed a selection method with unequal probabilities (Deville, 1998c), which is a variant of systematic sampling.

Then, with Yves Tillé, he proposed the splitting method (Deville and Tillé, 1998) to select samples with unequal probabilities of selection. This class of methods consists of starting with the vector of inclusion probabilities, $\boldsymbol{\pi}(0) = \boldsymbol{\pi}$, the sum of which is equal to an integer n. Then, at each step $t, 0, 1, 2, \ldots,$ this vector $\boldsymbol{\pi}(t)$ is randomly modified until a vector containing only values equal to 0 or 1 is obtained, which corresponds to the selection of a sample. For the method to be correct, three conditions must be met:

1. All components of the $\boldsymbol{\pi}(t)$ remain in the interval [0, 1].

2. The sum of the components of $\boldsymbol{\pi}(t)$ remains equal to n.
3. The martingale property must be satisfied:

$$\mathrm{E}_p\{\boldsymbol{\pi}(t)\,|\,\boldsymbol{\pi}(t-1)\} \;=\; \boldsymbol{\pi}(t-1), \quad \text{for all} \quad t, \tag{2.1}$$

where $\mathrm{E}_p(\cdot)$ is the expectation with respect to the sampling design that takes the sampling randomization into account.

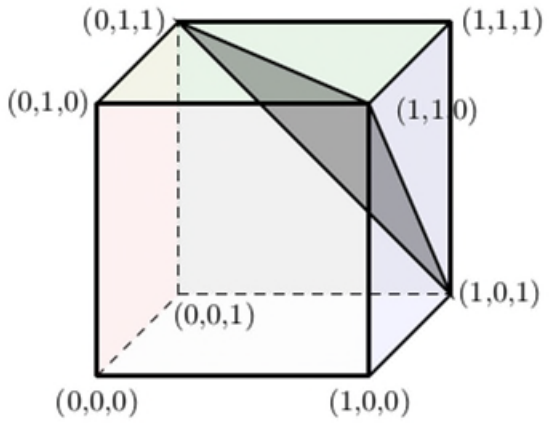
The martingale property is sufficient to show that the inclusion probabilities are respected at each stage. In fact, through the law of total expectation, we readily obtain $\mathrm{E}_p\{\boldsymbol{\pi}(t)\} = \boldsymbol{\pi}(0)$.

The splitting method is a very general way of representing a sampling method. Almost all selection algorithms can be viewed as a splitting procedure. This allows one to focus on basic steps. Checking the three conditions allows one to quickly check whether or not the method is correct.

One of the methods proposed as a special case of the splitting method is the pivot method for which only two components of the vectors $\boldsymbol{\pi}(t)$ are changed at each step. This method was generalized to select multiple non-overlapping samples in the same population with equal or unequal probabilities (Deville and Tillé, 2000b).

The transition from the splitting method to balanced sampling was relatively simple when Jean-Claude Deville and Yves Tillé realized that samples $(I_1, \ldots, I_N)^\top$ coded as vectors containing only 0 and 1 are the vertices of a N-cube of $\mathbb{R}^N$. In addition, conditions 1 and 2 of the splitting method can be interpreted geometrically. The vectors $\boldsymbol{\pi}(t)$ must remain in the simplex $\mathcal{P} = \left\{ c_k \in [0,1] \,\middle|\, \sum_{k=1}^{N} c_k = n \right\}$. Figure 2.1 shows a representation of this simplex for a sample of size $n = 2$ in a population of size $N = 3$. The splitting method is therefore a random walk in a simplex that must statisfy the martingale property.

Figure 2.1 Simplex bringing together samples of size $n = 2$ in a population of size $N = 3$ within a cube where the samples are the vertices. Here, the simplex is an equilateral triangle.



2.2 Balanced sampling using the cube method

The shift to balanced sampling then became self-evident. It was simply a matter of replacing condition 2 in the splitting method to obtain the general principles of a balanced sampling method. The following three conditions are therefore obtained:

1. All components of the $\boldsymbol{\pi}(t)$ remain in the interval [0.1].
2. At each step $t = 0, 1, 2, \ldots$ the vectors $\boldsymbol{\pi}(t) = (\pi_1(t), \ldots, \pi_N(t))^\top$ must meet the balancing equations:

$$\sum_{k \in U} \frac{\mathbf{z}_k}{\pi_k} \pi_k(t) \;=\; \sum_{k \in U} \mathbf{z}_k .$$

3. The martingale property must be satisfied

$$\mathrm{E}_p\{\boldsymbol{\pi}(t) \,|\, \boldsymbol{\pi}(t-1)\} \;=\; \boldsymbol{\pi}(t-1), \quad \text{for all } \; t. \tag{2.2}$$

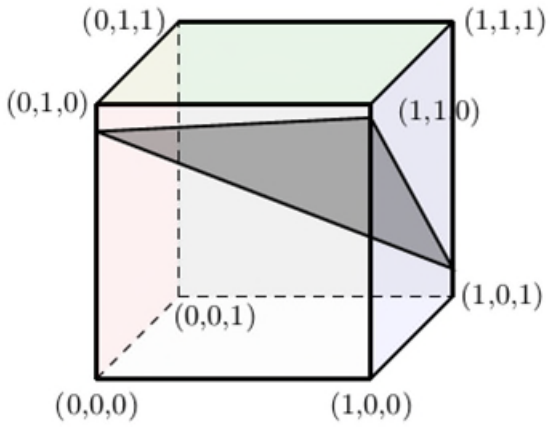
Conditions 1 and 2 now define a polytope

$$\mathcal{P} \;=\; \left\{ c_k \in [0,1] \,\middle|\, \sum_{k \in U} \frac{\mathbf{z}_k}{\pi_k} c_k = \sum_{k \in U} \mathbf{z}_k \right\},$$

in which the vectors $\boldsymbol{\pi}(t)$ must remain to statisfy the balancing constraints at each step. An example of a polytope is shown in Figure 2.2. However, when the constraints are complex, the vertices of the polytope $\mathcal{P}$ are not necessarily the vertices of the cube, meaning that there may not be exactly balanced samples. This will result in a roughly balanced sample. That is why the cube method consists of two phases: the flight phase and the landing phase.

The flight phase is a random walk in the polytope $\mathcal{P}$ that ends on one of the polytope vertices. The landing phase consists of selecting an approximately balanced sample close to the vertex of the polytope obtained at the end of the flight phase while satisfying the inclusion probabilities.

Figure 2.2 Polytope $\mathcal{P}$ in cases where the polytope vertices are not the vertices of the cube.



The cube method is a family of methods that allows such a random walk to be generated. For the flight phase, to go from $\boldsymbol{\pi}(t)$ to $\boldsymbol{\pi}(t+1)$, the cube method is conducted in the following manner.

1. A vector $\mathbf{u}(t) = (u_1(t), \ldots, u_N(t))^\top$ is generated so that

$$\sum_{k \in U} \frac{\mathbf{z}_k}{\pi_k} u_1(t) \;=\; \mathbf{0},$$

and $u_k(t) = 0$, if $\pi_k(t)$ is an integer (0 or 1). If such a vector does not exist, the flight phase stops.

2. We look for the largest positive values, λ_1 and λ_2, that satisfy

$$0 \le \pi_k(t) + \lambda_1 u_k(t) \le 1 \quad \text{and} \quad 0 \le \pi_k(t) - \lambda_2 u_k(t) \le 1, \quad \text{for all} \quad k \in U.$$

3. We update

$$\boldsymbol{\pi}(t+1) \;=\; \begin{cases} \boldsymbol{\pi}(t) + \lambda_1 \mathbf{u}(t) & \text{with probability } q \\ \boldsymbol{\pi}(t) - \lambda_2 \mathbf{u}(t) & \text{with probability } 1-q, \end{cases}$$

where $q = \lambda_2 / (\lambda_1 + \lambda_2)$.

There are several ways to generate the vector $\mathbf{u}(t)$, which allows you to define several variants of the method. After a maximum of N steps, the flight phase ends on the vertex of the polytope $\mathcal{P}$. This vertex is a vector containing at most Q values that are neither 0 nor 1, where Q is the number of auxiliary variables. To obtain a sample, one must apply the landing phase. Two variations are proposed in Deville and Tillé (2004).

The cube method was first published in the proceedings of the *Journées de méthodologie statistique* (Deville and Tillé, 2000a) and then as a chapter of a book (Deville and Tillé, 2001). The English publication was much more difficult but was eventually accepted in *Biometrika* (Deville and Tillé, 2004, 2005). A referee could not accept that the samples could be balanced and random at the same time. Another could not admit that the method worked without listing all possible samples. Another criticism was that the method did not provide exactly balanced samples. However, the existence of an exact solution does not depend on the method but on the geometry of the problem.

2.3 Implementation, method applications and research extensions

A first prototype of a SAS-IML function was written by three students from the ENSAI (Bousabaa, Lieber and Sirolli, 1999) under the supervision of Frédéric Tardieu and Yves Tillé. The first version was very slow, to the point its applicability was doubtful, but progress was quickly made. Chauvet and Tillé (2006a) proposed an implementation that considers only a small portion of the population at each stage, significantly reducing the computational time. An SAS function was written using this procedure (Chauvet and Tillé, 2006b). Several R packages also allow the selection of a balanced sample (Tillé and Matei, 2021; Grafström and Lisic, 2019; Jauslin, Eustache, Panahbehagh and Tillé, 2021). Their method is especially

simple because the functions depend on only two arguments: the matrix of balancing variables and the vector of inclusion probabilities.

Jean-Claude Deville was instrumental in changing the census procedure in France to a continuous system (Deville and Jacod, 1996). The cube method has been a valuable tool for constructing rotation groups (Durr and Dumais, 2002). The primary sampling units of the master sample were also selected using the cube method. The method was very quickly used in many applications (Tillé, 2011). As with calibration, balancing has become a standard procedure in survey statistics.

The cube method has also generated a lot of academic work. The accuracy of balancing is discussed in Chauvet, Haziza and Lesage (2015). Leuenberger, Eustache, Jauslin and Tillé (2022) suggest sorting observations in ascending depth order in the scatter plot, reducing the rounding problem. The issue of optimal inclusion probabilities is addressed in Nedyalkova and Tillé (2008, 2012) and Chauvet, Bonnéry and Deville (2011). These results generalized the optimal stratification of Neyman (1934). Several articles deal with balancing for stratified populations (Chauvet, 2009; Hasler and Tillé, 2014; Jauslin, Eustache and Tillé, 2021).

Several studies have been dedicated to spatial sampling. Grafström, Lundström and Schelin (2012) use the repulsive aspect of the pivot method to obtain samples properly spread out in space, increasing accuracy when data are autocorrelated. Grafström and Tillé (2013) then propose a variation of the cube method to obtain samples that are properly spread out and balanced on totals. Lastly, Jauslin and Tillé (2020a, b) balance on micro-strata containing the neighborhood of each unit to obtain particularly well-spread samples.

Jean-Claude Deville did a lot of work on maximum entropy plans, which he left several handwritten notes on (Deville, 2000b; Deville, nda, ndb, ndc). These results finally enabled a relatively quick implementation of this plan. Deville and Qualité (2005) then proposed an extension to the multidimensional case. As a result of a referee's remark during the submission of the article on the cube method, Jean-Claude Deville focused on determining a necessary and sufficient condition for balancing to have no rounding problems (Deville, 2015, 2014). Unfortunately, the condition obtained is very restrictive. In cases where the condition is met, it develops maximum entropy balanced designs (Deville, 2014).

Jean-Claude Deville quickly understood the value of the cube method for applications other than sampling. Several balanced imputation methods were proposed by Chauvet, Deville and Haziza (2011); Hasler and Tillé (2016); Eustache, Vallée and Tillé (2022). These methods have the advantage of properly restoring the distribution of the imputed variable while reducing the variance caused by random imputation. The cube method is also used in fields of application far from sampling such as in the MCMC (Monte Carlo Markov Chain) methods (Chopin and Ducrocq, 2021).

3. Calibration

The papers by Deville and Särndal (1992) and Deville, Särndal and Sautory (1993) on calibration methods (also called recovery methods) and published in the prestigious *Journal of the American Statistical*

Association are considered to be two of the most important and influential articles in the past 30 years in the field of sampling and official statistics. These two articles propose a unified theory of estimation in the presence of auxiliary information, the premises of which are discussed in Lemel (1976) and Huang and Fuller (1978). The two papers co-authored by Jean-Claude Deville have generated numerous research articles over the past three decades. The reader is referred to Särndal (2007), Haziza and Beaumont (2017), Devaud and Tillé (2019), and Zhang, Han and Wu (2022) for reviews on calibration methods. Post-stratification (e.g., Holt and Smith 1979), raking methods (Deming and Stephan, 1940; Stephan, 1942), generalized regression estimation (see, for example, Särndal, Swensson and Wretman, 1992) can be obtained as special cases of calibration methods.

Calibration methods use auxiliary information available at the estimation stage to ensure consistency between the survey estimates produced and known or estimated external totals. In practice, calibration methods are also used to reduce non-response and coverage errors.

3.1 Calibration in the absence of non-sampling errors

In this section, we consider an ideal theoretical framework for which non-response and coverage errors are assumed to be negligible. Calibration is based on the availability of a vector of auxiliary variables, $\mathbf{x}_k = (x_{1k},\ldots,x_{Jk})^\top$, and the corresponding vector of population totals, $\mathbf{t}_\mathbf{x} = (t_{x_1},\ldots,t_{x_J})^\top$, where $t_{x_j} = \sum_{k\in U} x_{jk}$, $j = 1,\ldots,J$. The vector $\mathbf{t}_\mathbf{x}$ is obtained from an external source such as the census, an administrative file or another survey.

When selecting a sample S of a population U, it is almost certain that the sample will suffer from a random distortion in terms of the vector of auxiliary variables $\mathbf{x}$, in the sense $\hat{\mathbf{t}}_{\mathbf{x},\pi} \neq \mathbf{t}_\mathbf{x}$, with $\hat{\mathbf{t}}_{\mathbf{x},\pi} = \sum_{k\in S} \mathbf{x}_k/\pi_k$. Unlike a systematic distortion (as is usually encountered in a non-response context), we face a random distortion since $\mathrm{E}_p(\hat{\mathbf{t}}_{\mathbf{x},\pi}) - \mathbf{t}_\mathbf{x} = \mathbf{0}$. The purpose of calibration is therefore to correct this distortion.

More formally, we are seeking a set of calibration weights $\{w_k; k\in S\}$ such that

$$\sum_{k\in S} \frac{d_k G(w_k/d_k)}{q_k} \tag{3.1}$$

is minimized subject to the J calibration constraints

$$\sum_{k\in S} w_k \mathbf{x}_k = \mathbf{t}_\mathbf{x}, \tag{3.2}$$

where $d_k = 1/\pi_k$ and q_k is a scaling factor selected by the user (see Deville and Särndal, 1992; Deville et al., 1993). In the majority of cases encountered in practice, we set $q_k = 1$ for all k. The function $G(\cdot)$ is a pseudo-distance function to measure the closeness between the weights before calibration d_k and the weights after calibration w_k.

Calibration weights w_k are given by

$$w_k = d_k F(q_k \hat{\boldsymbol{\lambda}}^\top \mathbf{x}_k), \tag{3.3}$$

where $\hat{\boldsymbol{\lambda}}$ is a vector of size J of estimated coefficients ensuring that the constraints (3.2) are satisfied, and $F(\cdot) = g^{-1}(\cdot)$ is the calibration function, defined as the inverse function of $g(.) \equiv \partial G(t)/\partial t$. The calibrated weight (3.3) can be viewed as the product of the weight before calibration, d_k, and an adjustment factor, $F(q_k \hat{\boldsymbol{\lambda}}^\top \mathbf{x}_k)$. In addition to the vector $\hat{\boldsymbol{\lambda}}$, the latter is dependent on the calibration function $F(\cdot)$ (and therefore pseudo-distance function $G(\cdot)$) as well as the characteristics of the unit k, q_k and $\mathbf{x}_k$. In some situations, the term $q_k \hat{\boldsymbol{\lambda}}^\top \mathbf{x}_k$ does not depend on k, in which case all calibration functions will lead to the same set of weights w_k. This occurs in the cases of post-stratification or ratio estimation (Haziza and Beaumont, 2017).

The calibration estimator of t_y is given by

$$\hat{t}_{y,C} = \sum_{k \in S} w_k y_k. \tag{3.4}$$

Deville and Särndal (1992) considered a range of functions $G(\cdot)$ some of which are presented in Table 3.1. For the generalized chi-squared distance, Deville and Särndal (1992) showed that calibration weights are given by

$$w_k = d_k (1 + q_k \hat{\boldsymbol{\lambda}}^\top \mathbf{x}_k),$$

where

$$\hat{\boldsymbol{\lambda}} = \left(\sum_{k \in S} d_k \mathbf{x}_k q_k \mathbf{x}_k^\top \right)^{-1} (\mathbf{t}_\mathbf{x} - \hat{\mathbf{t}}_{\mathbf{x},\pi}).$$

It follows that the calibration estimator coincides with the well-known generalized linear regression estimator (see, for example, Särndal et al., 1992)

$$\hat{t}_{y,C} = \hat{t}_{y,\pi} + (\mathbf{t}_\mathbf{x} - \hat{\mathbf{t}}_{\mathbf{x},\pi})^\top \hat{\mathbf{B}}, \tag{3.5}$$

where

$$\hat{\mathbf{B}} = \left(\sum_{k \in S} d_k \mathbf{x}_k q_k \mathbf{x}_k^\top \right)^{-1} \sum_{k \in S} d_k \mathbf{x}_k q_k y_k.$$

This result is one of the important breakthrough in the field of estimation in the presence of auxiliary information: it is possible to construct the generalized regression estimator using calibration. Deville et al. (1993) have established that the use of Kullback-Leibler information (see Table 3.1) leads to the standard raking estimator, which is another major contribution. The truncated generalized chi-squared distance and the logit distance (see Table 3.1) allow for bounds to be placed on the calibration adjustment factors, $F(q_k \hat{\boldsymbol{\lambda}}^\top \mathbf{x}_k)$, to limit the dispersion of the weights.

Although the calibration estimators are biased with respect to the sampling design, they are consistent, which is a desirable property (Deville and Särndal, 1992). When the sample size n is large enough, the

square bias of calibration estimators becomes negligible in front of their variance. Therefore, the mean square error of calibration estimators is approximately equal to their variance, provided that n is large enough.

Deville and Särndal (1992) showed that the variance of a calibration estimator can be approximated by

$$V_p(\hat{t}_{y,C}) \approx \sum_{k\in U}\sum_{\ell\in U}\Delta_{k\ell}\frac{E_k}{\pi_k}\frac{E_\ell}{\pi_\ell}, \tag{3.6}$$

where $E_k = y_k - \mathbf{x}_k^\top \mathbf{B}$ is the "census residual" associated with unit k with

$$\mathbf{B} = \left(\sum_{k\in U}\mathbf{x}_k q_k \mathbf{x}_k^\top\right)^{-1}\sum_{k\in U}\mathbf{x}_k q_k y_k.$$

This is a remarkable property: all calibration estimators have the same asymptotic variance regardless of the calibration function $F(\cdot)$. Expression (3.6) suggests that calibration estimators are efficient when residuals E_k are small, which will occur when the relationship between the variable of interest y and the calibration variables $\mathbf{x}$ is linear and strong. What if the relationship is not linear? In this case, the model may not fit the data well, leading to large residuals and a large variance. This has led Wu and Sitter (2001) to propose a model calibration procedure that allows for non-linear relationships through, for example, generalized linear models. However, unlike the classic calibration of Deville and Särndal (1992), model calibration requires the availability of the vector $\mathbf{x}$ for all population units. This requirement is generally not met in practice, especially in household surveys.

Table 3.1
A few distance functions introduced in Deville and Särndal (1992).

	Distance function $G(w_k/d_k)$	**Calibration adjustment factor $F(q_k\hat{\boldsymbol{\lambda}}^\top\mathbf{x}_k)$**
Generalized chi-square distance	$\frac{1}{2}\left(\frac{w_k}{d_k}-1\right)^2$	$1+q_k\hat{\boldsymbol{\lambda}}^\top\mathbf{x}_k$
Kullback-Leibler Information	$\frac{w_k}{d_k}\log\frac{w_k}{d_k}-\frac{w_k}{d_k}+1$	$\exp(q_k\hat{\boldsymbol{\lambda}}^\top\mathbf{x}_k)$
Inverse Kullback-Leibler information	$\log\frac{d_k}{w_k}+\frac{w_k}{d_k}-1$	$\frac{1}{1-q_k\hat{\boldsymbol{\lambda}}^\top\mathbf{x}_k}$
Hellinger distance	$2\left\{\sqrt{\frac{w_k}{d_k}}-1\right\}^2$	$\frac{1}{\sqrt{1-2q_k\hat{\boldsymbol{\lambda}}^\top\mathbf{x}_k}}$
Truncated generalized chi-square distance	$\begin{cases}\frac{1}{2}\left(\frac{w_k}{d_k}-1\right)^2 & L<\frac{w_k}{d_k}<M\\ \infty & \text{otherwise}\end{cases}$	$\begin{cases}1+q_k\boldsymbol{\lambda}^\top\mathbf{x}_k & (L-1)\le q_k\boldsymbol{\lambda}^\top\mathbf{x}_k\le(M-1)\\ M & q_k\boldsymbol{\lambda}^\top\mathbf{x}_k>(M-1)\\ L & q_k\boldsymbol{\lambda}^\top\mathbf{x}_k<(L-1)\end{cases}$
Logit distance	$\begin{cases}\left(a_k\log\frac{a_k}{1-L}+b_k\log\frac{b_k}{M-1}\right)\frac{d_k}{A} & L<\frac{w_k}{d_k}<M\\ \infty & \text{otherwise,}\end{cases}$	$\frac{L(M-1)+M(1-L)\exp(Aq_k\boldsymbol{\lambda}^\top\mathbf{x}_k)}{M-1+(1-L)\exp(Aq_k\boldsymbol{\lambda}^\top\mathbf{x}_k)}$

In multi-stage or multi-phase surveys, we face several layers of auxiliary information. For example, in two-stage sampling, we may have auxiliary information at the household level (number of individuals in the household, number of individuals in each age group, owner or tenant status, etc.) and information at the individual level (gender, age group, etc.). The reader is referred to Sautory and Le Guennec (2003) and Estevao and Särndal (2002, 2006) for a discussion of calibration methods in a multi-stage or multi-phase sampling.

3.2 Adjustment of non-response by calibration

Post-stratification and raking methods have long been used to treat unit non-response; see, for example, Thomsen (1978), Bethlehem and Keller (1987), and Bethlehem (1988). The first work on a unified non-response calibration approach is presented in Deville and Dupont (1993) and Dupont (1993). The approach was further investigated by Lundström and Särndal (1999) and Särndal and Lundström (2005). The idea is to obtain final weights w_k from the initial weights in order d_k to achieve the following two objectives: (i) reduce non-response bias and (ii) ensure consistency between survey estimates and known population totals.

We consider a population U in which a sample S is selected and only a subset S_r of units have responded. We therefore have $S_r \subset S \subset U$. We have two levels of auxiliary information: (1) The vector $\mathbf{x}_{Uk}$ that is observed for $k \in S_r$ and for which the vector of population totals, $\sum_{k\in U} \mathbf{x}_{Uk}$, is known. (2) The vector $\mathbf{x}_{Sk}$ that is observed for $k \in S$ and for which the vector of Horvitz-Thompson estimates, $\sum_{k\in S} d_k \mathbf{x}_{Sk}$, is available. Variables $\mathbf{x}_{Uk}$ are those that will ensure consistency between survey estimates and known population totals. Ideally, the variables $\mathbf{x}_{Sk}$ are those that explain both the response status R_k and the variables of interest. For each $k \in S_r$, we create the stacked vector $\mathbf{x}_k = \begin{pmatrix} \mathbf{x}_{Uk} \\ \mathbf{x}_{Sk} \end{pmatrix}$. We are seeking a final weighting system, $\{w_k; k \in S_r\}$, such that

$$\sum_{k\in S_r} \frac{d_k G(w_k/d_k)}{q_k}$$

is minimized subject to the calibration constraints

$$\sum_{k\in S_r} w_k \mathbf{x}_k = \begin{pmatrix} \sum_{k\in U} \mathbf{x}_{Uk} \\ \sum_{k\in S} d_k \mathbf{x}_{Sk} \end{pmatrix}.$$

The final weights are given by

$$w_k = d_k \times F(q_k \hat{\boldsymbol{\lambda}}_r^\top \mathbf{x}_k).$$

The estimator for t_y is given by

$$\hat{t}_{y,C} = \sum_{k\in S_r} \left\{ d_k \times F(q_k \hat{\boldsymbol{\lambda}}_r^\top \mathbf{x}_k) \right\} y_k. \tag{3.7}$$

Although any pseudo-distance function $G(\cdot)$ can be used, caution should be exercised. Indeed, selecting a pseudo-distance function in the context of non-response comes down to imposing a parametric model describing the relationship between the inverse of the response probabilities and the vector $\mathbf{x}$ (Haziza and Lesage, 2016). In general, an erroneous choice of the function $G(\cdot)$ will generally lead to a biased calibration estimator. An exception to this rule will occur when the variable of interest y is linearly related to the vector $\mathbf{x}$ and non-response is of the Missing At Random type (Rubin, 1976).

Another important contribution by Jean-Claude Deville is calibration on instrumental variables, also known as generalized calibration (Deville, 1998a, 2000a, 2002). This approach has also been studied and discussed by, among others, Sautory and Le Guennec (2003), Kott (2006), Chang and Kott (2008), Kott and Chang (2010), Haziza and Beaumont (2017), and Lesage, Haziza and D'Haultfoeuille (2019). This approach is especially useful in the context of nonignorable non-response (Rubin, 1976). In this case, the response probability depends on fully observed variables but also variables available for the respondents only. As a result, estimating the response probabilities is not easy. Generalized calibration leads to a consistent estimator of a total if the exclusion restriction conditions are met.

4. The weight-sharing method

Indirect sampling involves selecting a sample from a target population using a different sampling frame, but somewhat related to that target population. Many developments related to the indirect survey can be found in Lavallée's books (2002, 2007) to which we add more recent contributions such as Deville and Maumy-Bertrand (2006); Falorsi, Piersante and Bako (2016); Kiesl (2010); Medous, Goga, Ruiz-Gazen, Beaumont, Dessertaine and Puech (2023). We will see that Jean-Claude Deville played a leading role in the development of indirect sampling.

4.1 The very beginning: Longitudinal surveys

The genesis of the indirect sampling relates to a weighting problem in the context of longitudinal surveys. This involved weighting individuals interviewed in a social longitudinal survey that tracks individuals belonging to an household over time.

After a selection of households (and thus individuals) in the first wave, changes in household composition throughout the waves, partly due to marriages and deaths, made the weighting process difficult. The solution is achieved by using the weight-share method (see Lavallée, 1995).

The problem of weighting longitudinal household surveys has attracted interest from several authors, including Huang (1984); Judkins, Hubble, Dorsch, McMillen and Ernst (1984); Ernst, Hubble and Judkins (1984); Ernst (1989); and Kalton and Brick (1995). The article by Ernst (1989) clearly described the basis of the problem and proposed a solution related to the weight share method.

Consider a longitudinal survey of individuals drawn from households. Two waves of data are available: the wave A (or first wave) and the wave B (a subsequent wave). A sample S^A containing m^A individuals was drawn from the population in the wave A containing M^A individuals. Let $\pi_k^A > 0$, the probability of selection of individual k. In wave B, the population then contains M^B individuals distributed among N^B households U_i^B, where the household i contains M_i^B individuals.

The longitudinal survey process is as follows. For each individual k from S^A, a list is established of M_i^B individuals from the household i in wave B containing this individual, i.e., S^B, all of the n^B households identified by the individuals $k \in S^A$. Once households from S^B have been identified, all individuals k in households $i \in S^B$ are surveyed to measure the variable of interest y. The weight-share method assigns an estimation weight w_{ik} to each individual k in a surveyed household U_i^B. The method steps are as follows:

- **Step 1** For each individual k in the households i of S^B, we calculate the initial weight $w'_{ik} = \gamma_k / \pi_k^A$, where $\gamma_k = 1$ if $k \in S^A$, and 0 otherwise.
- **Step 2** For each household i of S^B, we obtain the total number of individuals M_i^{AB} in the household i present in wave A (but not necessarily contained in S^A).
- **Step 3** The final weight $w_i = \sum_{k \in U_i^B} w'_{ik} \Big/ M_i^{AB}$ is calculated.
- **Step 4** Lastly, we set $w_{ik} = w_i$ for all $k \in U_i^B$.

We could consider calculating the selection probability π_{ik}^B of the individual k in household i of S^B. This probability corresponds to the probability of selecting any of the M_i^B individuals in the household i, and therefore we must know each of the M_i^B probabilities π_k^A in the household i of S^B. Unfortunately, especially in the case of multistage surveys, the probabilities π_k^A are often unknown. In addition, apart from relatively simple cases (for example, when individuals k are independently selected in S^A), the calculation of the weights π_{ik}^A can be very complex. The weight-share method thus offers a simple solution to a weighting problem that is difficult, if not impossible, to carry out in practice.

4.2 A generalization of the problem

Imagine links (or correspondence) between individuals in both waves of the survey. Since it involves tracking individuals over time, these links can be seen as "one-to-one" (Figure 4.1). During discussions with Jean-Claude Deville, he came up with the following idea: "Why not generalize the links?" So instead of having "one-to-one" links, why not consider "many-to-many" links (Figure 4.2)? Figures 4.1 and 4.2 provide a graphical representation of the methods. The sample S^A is the yellow subset of the wave A. The green subsets of the wave B are clusters U_i^B (the households) encountered in the second wave.

Figure 4.1 Longitudinal links (“one-to-one”).

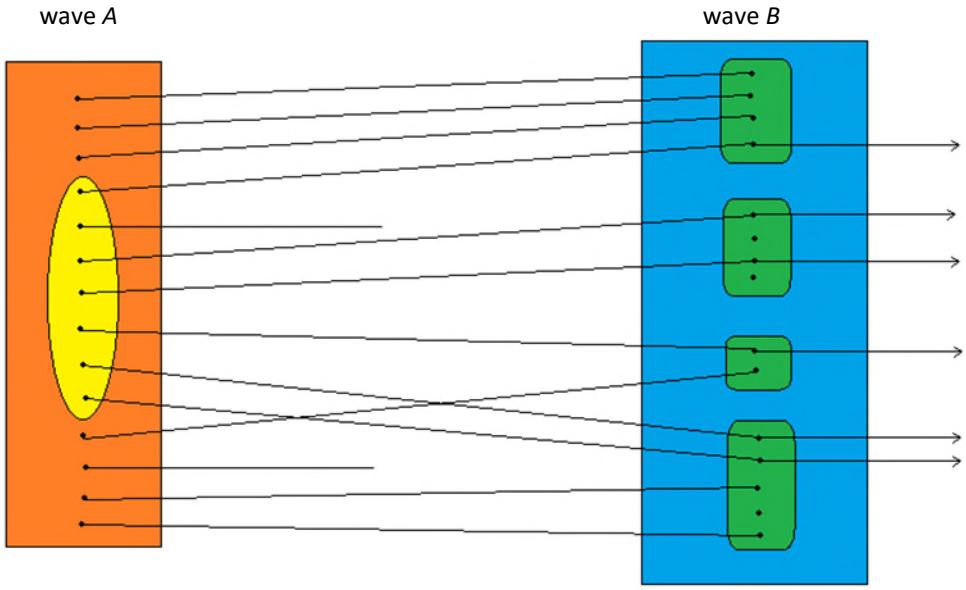
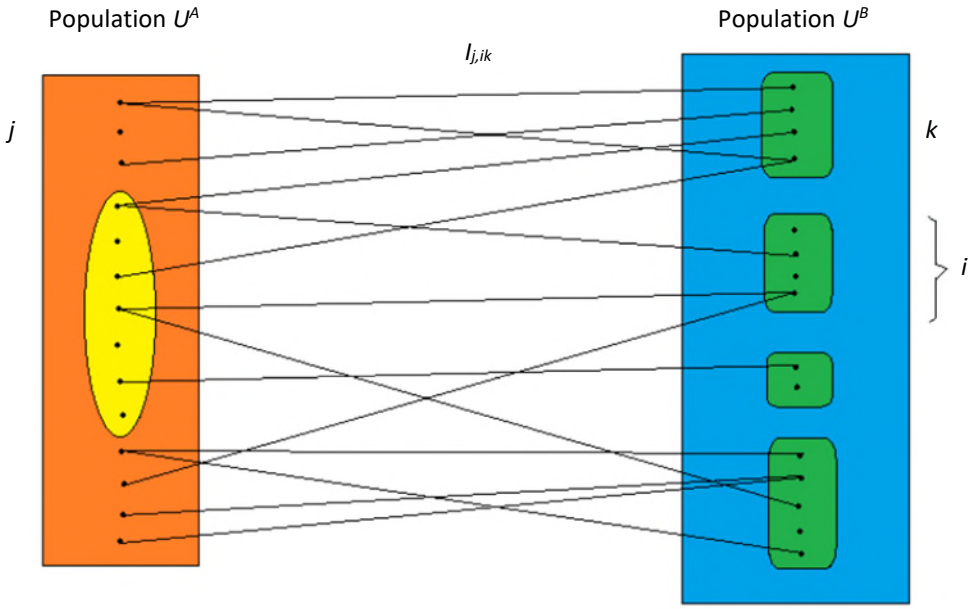


Figure 4.2 Arbitrary links (“many-to-many”).



With this new way of looking at links, the question then became how to associate a weight (unbiased) to the surveyed units of U^B (the target population) following the selection of units in U^A (survey frame). In fact, the problem was much broader than that of longitudinal surveys.

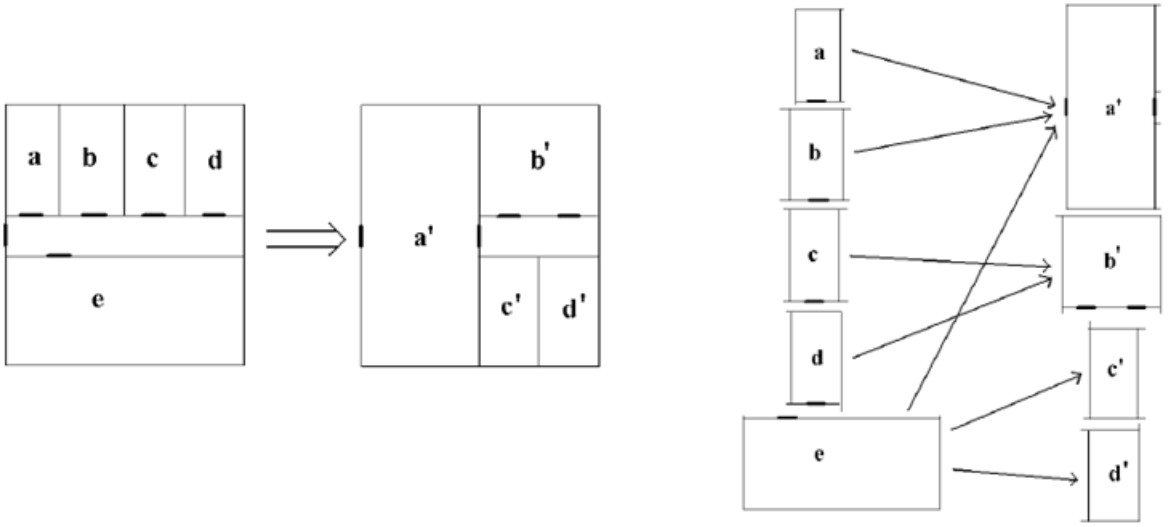
The new problem studied was the following. Let U^A and U^B be two populations related to each other. An estimate is required for U^B (target population), but a survey frame is available only for U^A. The

proposed solution is to then draw a sample from U^A to produce estimates for U^B using the existing correspondence (links) between the two populations. For this new approach, Jean-Claude Deville then coined the term indirect sampling.

Indirect sampling proceeds as follows. First, for each unit j of S^A, we identify the units k from the clusters i of U^B that have a link with j. Let U_i^B be the set of all the units k in cluster i. For each unit k identified, we list the M_i^B units from cluster i containing this unit. Lastly, we survey all the units k from the clusters $i \in S^B$ to measure the variable of interest y.

To illustrate indirect sampling, Jean-Claude Deville suggested an example where the goal is to survey people (units) living in dwellings (clusters). A sampling frame of dwellings is available, but unfortunately not up to date. This sampling frame does not include, among other things, the renovations impacting the divisions of dwelling in buildings. An example of this type of renovation is shown to the left of Figure 4.3. It can be seen that dwellings a, b, c, d and e were transformed to obtain dwellings a', b', c' and d'. Drawing a sample of dwellings from the sampling frame, we get to the new dwellings using the correspondence between the old and new dwellings. This correspondence is illustrated to the right of Figure 4.3.

Figure 4.3 Pre- and post-renovation dwellings (left) and indirect sampling of renovated dwellings (right).



Estimating the total t_y^B of the target population U^B can be done using S^A drawn from U^A. However, note that this can be a major challenge if the links between the units of U^A and U^B are not one-to-one. In fact, in this case, it is difficult, if not impossible, to associate a selection probability, or an estimation weight, to the units surveyed in U^B. The solution then is to use the Generalized Weight-Share Method (GWSM), which yields an estimation weight for each surveyed unit of the target population U^B.

As with longitudinal surveys, we consider a sample S^A containing m^A units drawn from U^A containing M^A units. The target population U^B contains M^B units and it is divided into N^B clusters, where the cluster i contains M_i^B units. The links (or correspondence) between the units j of U^A and the units k of the clusters i of U^B are identified by the variable $l_{i,jk}$, where $l_{j,ik}=1$ if there is a link between the unit $j \in U^A$ and the unit k of the cluster i of U^B, and 0 otherwise.

To apply the GWSM, simply follow these steps (reminiscent of the weight-share method, but more general):

- **Step 1** For each unit k of the clusters i of S^B, we calculate the initial weight $w'_{ik} = \sum_{j \in U^A} l_{j,ik} \gamma_j \big/ \pi_j^A$, where $\gamma_j = 1$ if $j \in S^A$, and 0 otherwise.
- **Step 2** For each unit k of the clusters i of S^B, we obtain the total number of links $L_{ik}^B = \sum_{j \in U^A} l_{j,ik}$.
- **Step 3** The final weight $w_i = \sum_{k \in U_i^B} w'_{ik} \big/ \sum_{k \in U_i^B} L_{ik}^B$ is calculated.
- **Step 4** Lastly, we apply $w_{ik} = w_i$ for all $k \in U_i^B$.

Lavallée (2002, 2007) mentioned that indirect sampling and the GWSM are useful because they offer a simple solution to complex survey and weighting problems. In addition, the GWSM generally yields the same results as classical results in the context of simple problems. In fact, the GWSM is an interesting solution, although it is not always the most accurate (minimum variance) compared with another more complex estimation method.

4.3 Properties of the Generalized Weight-Share Method

The development of the properties of the GWSM took place during discussions with Jean-Claude Deville, which began in 1995. These led to the following theorem and its two corollaries:

Theorem 1 *Duality of the form of $\hat{t}_y^B$ with respect to U^A and U^B. The estimator $\hat{t}_y^B$ can be written in both forms:*

$$\hat{t}_y^B = \sum_{i \in S^B} \sum_{k \in U_i^B} w_{ik} y_{ik}$$

(with GWSM weights) and

$$\hat{t}_y^B = \sum_{j \in U^A} \gamma_j Z_j \Big/ \pi_j^A,$$

where $Z_j = \sum_{i \in U^B} \sum_{k \in U_i^B} l_{j,ik} Y_i \big/ L_i^B$.

Theorem 1 shows that we are, ultimately, in the presence of a simple Horvitz-Thompson estimator. From this finding, we obtain the following two corollaries:

Corollary 1 *Bias of* $\hat{t}_y^B$. *The estimator* $\hat{t}_y^B$ *is unbiased for the estimation of* Y^B, *with respect to the sampling design.*

Corollary 2 *Variance of* $\hat{t}_y^B$. *The variance formula for the estimator* $\hat{t}_y^B$, *with respect to the sampling design, is given by*

$$V_p(\hat{t}_y^B) = \sum_{j\in U^A}\sum_{j'\in U^A}\left(\pi_{jj'}^A - \pi_j^A\pi_{j'}^A\right)\frac{Z_j Z_{j'}}{\pi_j^A\pi_{j'}^A},$$

where $\pi_{jj'}^A$ *is the joint probability of selection of units* j *and* j'.

4.4 Calibration

Let's assume that we want to correct the GWSM weights so that the estimates produced correspond to known totals (auxiliary information). The most commonly used technique is the calibration developed by Deville and Särndal (1992).

In the context of indirect sampling, there are two possible sources of auxiliary information:

(i) From the survey frame U^A, we have a column vector $\mathbf{x}_j^A$ and its total $\mathbf{t}_x^A = \sum_{j\in U^A}\mathbf{x}_j^A$ (assumed to be known).

(ii) From the target population U^B, we have a column vector $\mathbf{x}_{ik}^B$ and its total $\mathbf{t}_x^B = \sum_{i\in U^B}\sum_{k\in U_i^B}\mathbf{x}_{ik}^B$ (assumed to be known).

The calibration constraints associated with the GWSM are:

(i) $\hat{\mathbf{t}}_x^{\mathrm{CAL},A} = \sum_{j\in S^A} w_j^{\mathrm{CAL},A}\mathbf{x}_j^A = \mathbf{t}_x^A$ and

(ii) $\hat{\mathbf{t}}_x^{\mathrm{CAL},B} = \sum_{i\in S^B}\sum_{k\in U_i^B} w_{ik}^{\mathrm{CAL},B}\mathbf{x}_{ik}^B = \mathbf{t}_x^B$, where $w_j^{\mathrm{CAL},A}$ is the calibration weight obtained from $d_j^A = 1/\pi_j^A$, and $w_{ik}^{\mathrm{CAL},B}$ is the calibration weight of the unit k from the surveyed cluster i where the GWSM was applied.

Based on Theorem 1, the latter constraint can be rewritten as: $\hat{\mathbf{t}}_x^{\mathrm{CAL},B} = \sum_{j\in S^A} w_j^{\mathrm{CAL},A}\boldsymbol{\Gamma}_j = t_x^B$, where $\boldsymbol{\Gamma}_j = \sum_{i\in U^B}\sum_{k\in U_i^B} l_{j,ik}X_i^B/L_i^B$. This constraint is now expressed in terms of units $j\in s^A$.

By defining the vectors

$$\mathbf{x}_j^{AB} = \begin{pmatrix}\mathbf{x}_j^A\\ \boldsymbol{\Gamma}_j\end{pmatrix} \quad\text{and}\quad \mathbf{t}_x^{AB} = \begin{pmatrix}\mathbf{t}_x^A\\ \mathbf{t}_x^B\end{pmatrix},$$

we get the single constraint encompassing U^A and U^B:

$$\hat{\mathbf{t}}_x^{\mathrm{CAL},AB} = \sum_{j\in S^A} w_j^{\mathrm{CAL},A}\mathbf{x}_j^{AB} = \mathbf{t}_x^{AB}.$$

The formulation of the problem for determining the estimator $\hat{t}_y^{\mathrm{CAL},B} = \sum_{j \in S^A} w_j^{\mathrm{CAL},A} Z_j$ associated with the GWSM is: Determine $w_j^{\mathrm{CAL},A}$, for $j \in S^A$, to minimize the total distance

$$\sum_{j \in S^A} G_j(w_j^{\mathrm{CAL},A}, d_j^A)$$

subject to the single constraint

$$\hat{\mathbf{t}}_x^{\mathrm{CAL},AB} = \sum_{j \in S^A} w_j^{\mathrm{CAL},A} \mathbf{x}_j^{AB} = \mathbf{t}_x^{AB}.$$

This formulation is consistent with that of Deville and Särndal (1992). Calibration can therefore be readily applied to indirect sampling and GWSM.

It is important to note that this calibration work was done without direct collaboration with Jean-Claude Deville. However, after all the work presented by Lavallée (2001) at the *Colloque francophone sur les sondages* in Brussels, he discovered the article of Deville (1998b), which provides the same solution to the calibration problem associated with GWSM.

4.5 Optimization of the links

The indicator variable $l_{j,ik}$ indicates whether or not there is a link between the units j of the sampling frame U^A and the units k of the clusters i of the target population U^B. However, it does not indicate the relative importance that some links might have over others. It is possible to replace $l_{j,ik}$ with a quantitative variable $\theta_{j,ik}$ representing the importance we want to give to the link $l_{j,ik}$. This variable $\theta_{j,ik}$ is defined on $[0,+\infty)$, where $\theta_{j,ik} = 0$ is equivalent to $l_{j,ik} = 0$. It should be noted that if the process for assigning values of $\theta_{j,ik}$ is independent of the selection of S^A, the GWSM remains unbiased.

By replacing the links $l_{j,ik}$ with $\theta_{j,ik}$, we obtain a new estimator (unbiased) $\hat{t}_{y\theta}^B$. The problem is then to determine optimal values $\theta_{j,ik}$ so as to minimize the variance of $\hat{t}_{y\theta}^B$. Indeed, since the estimator $\hat{t}_{y\theta}^B$ remains unbiased regardless of the values of $\theta_{j,ik}$, it must be possible to determine the values of the latter to maximize the precision of $\hat{t}_{y\theta}^B$. The problem is therefore to determine $\left[\theta_{j,ik}\right]_{M^A \times M^B}$ to minimize $V_p(\hat{t}_{y\theta}^B) = f(y_{ik};\ i = 1, \ldots, N^B;\ k = 1, \ldots, M_i^B)$.

Deville and Lavallée (2006) determined the values of $\theta_{j,ik}$ such that the variance of the estimator $\hat{t}_{y\theta}^B$ is (almost) minimal. The optimal solution is relatively complex, and often depends on the variable of interest y. However, Jean-Claude Deville came up with the idea of defining the concept of weak optimality as well as that of strong optimality independent of the y -variables.

Weak optimality consists of determining values of $\theta_{j,ik}$ to minimize the variance of $\hat{t}_{y\theta}^B$ for a very specific choice of a variable of interest: $y_{ik} = 1$ for a given unit k of a cluster i of U^B and $y_{i'k'} = 0$ for all other units of U^B. The optimization problem reduces to determining $[\theta_{j,ik}]_{M^A \times M^B}$ as to minimize $V_p(\hat{t}_{y\theta}^B) = f(y_{ik} = 1;\ y_{i'k'} = 0;\ \forall i \neq i' \text{ and } k \neq k')$. Deville and Lavallée (2006) mention that weak optimality consists of minimizing the variance $\mathrm{V}(w_{ik}^\theta \mid ik \in S^B)$ of the weight w_{ik}^θ obtained by the GWSM (with $\theta_{j,ik}$ instead of $l_{j,ik}$) for all possible values of $[\theta_{j,ik}]_{M^A \times M^B}$. We note that the resulting weakly optimal weighted links do

not involve the variable y itself (since the values of y have been replaced by 1 and 0). In addition, the weakly optimal values of $\theta_{j,ik}$ are generally relatively easy to calculate.

Strong optimality independent of y involves an additional step to weak optimality. It consists of checking that the weakly optimal values of $\theta_{j,ik}$ do not generally depend (i.e., for any value of y other than 1 and 0) on the variable of interest y. With this in mind, Deville and Lavallée (2006) proposed a criterion to check whether the weak optimality corresponds to the strong optimality (minimum variance of $\hat{t}_{y\theta}^{B}$). If this criterion is met, strong optimality does not depend on the variable of interest y.

5. The development of variance expression and its estimation for complex estimators

The development of a variance formula and its estimation for a sample estimator is an essential step in producing confidence intervals that will inform users of statistics on their reliability. Conventional theory uses either an analytical approach, which is, by nature, based on mathematical formulas, or a sample replication approach (bootstrap, jackknife, random groups). Roughly speaking, the analytical approach may be considered more applicable when sampling is complicated and the expression estimator is rather simple, whereas the replication approach is used more readily in the opposite configuration, that is, in the presence of simple sampling and a complex estimator. This was certainly a common strategy before the development of the linearization theory for complex statistics. Jean-Claude Deville has contributed a great deal to the theory of linearization. Of course, the technique of linearization of estimators defined as functions of linear components, typically differentiable functions of estimators of totals, such as a ratio or coefficient of regression or estimators that are the solution of estimating equations (Woodruff, 1971; Binder, 1983; Wolter, 1985; Binder, 1991; Francisco and Fuller, 1991; Binder and Kovačević, 1995; Binder, 1996), has been known for many years. In the late 1990s, Jean-Claude Deville proposed a formal framework based on the influence function in the journal *Survey Methodology* (Deville, 1999) to deal with highly non-linear statistics, such as fractiles or parameters defined as solutions to certain equations (implicit parameters), in an asymptotic and general setting. When the sample size is large, linearization eventually allows a very complex estimator to be approximated by a classical linear estimator of the Horvitz-Thompson type, and then the variance of the former is approximated by the variance of the latter, thus producing the desired result.

More specifically, the historical approach considers the parameter θ and estimator $\hat{\theta}$ as differentiable functions of the individual variables of interest. Linearization is then based on a Taylor expansion procedure of $\hat{\theta}$ around its expectation θ. Survey weights are present in the expression of $\hat{\theta}$ but are not treated as variables. Jean-Claude Deville reverses the approach in some ways by considering $\hat{\theta}$ as a function of survey weights and uses a derivative with respect to the weights; it is the influence function, introduced in the next part.

5.1 The theoretical framework

The proposed methodology is explained in three stages: first, an asymptotic framework, second, a formalization using the concept of measure on a probabilistic space, and third, the use of the concept of influence function, which is appropriately adapted to the context.

The asymptotic framework is the one defined in Isaki and Fuller (1982), and considers a series of nested populations, with respective sizes N going to infinity, within which samples s are selected, whose size n also goes to infinity. For any individual variable x_k, if t_x is the true total of x_k and

$$\hat{t}_x = \sum_{k \in s} w_k x_k$$

its estimator, we assume that $N^{-1}t_x$ has a limit, and that $N^{-1}(\hat{t}_x - t_x)$ converges towards 0 in probability and that $\sqrt{n}N^{-1}(\hat{t}_x - t_x)$ converges in distribution towards a Gauss distribution. Any complex statistic S constructed from true or estimated totals is based on similar assumptions; depending to its expression, it is consistent when it is multiplied by $N^{-\alpha}$, where α is a positive integer. The integer α is called the degree of homogeneity. A ratio is a homogeneous statistic of degree 0 and a variance is a homogeneous statistic of degree 2. The first axiom is therefore extended by assuming that $N^{-\alpha}S$ has a limit.

Then comes the formalization of the estimators by using the concept of measure. In the expression of any parameter, individuals in the finite population are "naturally" weighted by a weight equal to 1, interpreted as a mass associated with a measure M with finite support. In the same population, sampling results in weighting any individual k in the selected sample s by the survey weight w_k and any individual k outside s by 0. This leads to the measure $\hat{M}$. A parameter, however complex, can be expressed as a function of M, noted $T(M)$ and called "functional of M". As an example, consider a total

$$T(M) = \sum_{k=1}^{N} y_k = \sum_{k=1}^{N} y_k M(k).$$

Adopting a general notation familiar in the context of measure theory, we write $T(M) = \int y dM$, where we integrate over all the individuals in the population. Turning to estimators, again in the case of a total, we consider

$$T(\hat{M}) = \sum_{k \in s} w_k y_k = \sum_{k=1}^{N} y_k \hat{M}(k),$$

in which case $T(\hat{M}) = \int y d\hat{M}$. This parallelism applies to any complex parameter, initially expressed as $T(M)$ and estimated by $T(\hat{M})$, an estimator obtained by substituting M with $\hat{M}$.

The third component of the theory uses the notion of influence function, which is used in the theory of robust statistics (Hampel, Ronchetti, Rousseeuw and Stahel, 1985). We consider a specific measure δ_k obtained by assigning a mass equal to 1 to the individual k, and the measure $M + t\delta_k$ leading to the mass

$1+t$ to the individual k and to the mass 1 for all other individuals. The influence function is defined – when the limit exists – by

$$IT(M,k) = \lim_{t\to 0} \frac{T(M+t\delta_k) - T(M)}{t}.$$

It can be shown, under certain technical conditions most often satisfied in practice, that when the measure space is equipped with a distance, if a measure M_2 converges to a measure M_1, then

$$T(M_2) = T(M_1) + \int IT(M_1,k)\, dM_2 - \int IT(M_1,k)\, dM_1 + R_\epsilon,$$

where R_ϵ is a random residual that converges to 0 in probability. This equality must be adapted to the postulated initial asymptotic conditions: to deal generally with the homogeneous statistics of degree α, it is the functionals $N^{-\alpha}T$ that have the required asymptotic properties and must therefore be considered here. Also, noting that the total mass associated with the measurement M is N, by setting a distance

$$d\left(\frac{M_1}{N}, \frac{M_2}{N}\right) = \left| \int y d\left(\frac{M_1}{N}\right) - \int y d\left(\frac{M_2}{N}\right) \right|$$

and by setting $M_1 = M$ and $M_2 = \hat{M}$, it follows that $\hat{M}/N$ converges toward M/N, which ultimately leads to:

$$N^{-\alpha}T(\hat{M}) = N^{-\alpha}T(M) + \int IT(M,k)\, d\left(\frac{\hat{M}}{N}\right) - \int IT(M,k)\, d\left(\frac{M}{N}\right) + R_\epsilon,$$

where the residual R_ϵ is negligeable in probability in front of $1/\sqrt{n}$, under these conditions, that is,

$$\text{for all } \epsilon > 0,\ P\left(\left| \sqrt{nR_\epsilon} > \epsilon \right| \right) \to 0.$$

By noting $IT(M,k) = z_k$, it follows that

$$N^{-\alpha}(T(\hat{M}) - T(M)) = \frac{1}{N}\sum_{k\in s} w_k z_k - \frac{1}{N}\sum_{k=1}^{N} z_k + R_\epsilon. \tag{5.1}$$

By noting $\hat{t}_z = \sum_{k\in s} w_k z_k$ the natural linear estimator of the total $t_z = \sum_{k=1}^{N} z_k$,

$$\sqrt{n}N^{-\alpha}(T(\hat{M}) - T(M)) = \sqrt{n}\frac{1}{N}(\hat{t}_z - t_z) + R_\epsilon.$$

According to the third asymptotic assumption, the term on the right has a Gaussian limit, and therefore the term on the left hand-side has an asymptotic variance, which is equal to that on the right hand-side. It is customary to use $N^{2(\alpha-1)}V(\hat{t}_z)$ as an approximate variance of $T(\hat{M})$ when n is considered to be "sufficiently large". The variable z_k is called a linearized variable associated with the functional T. To carry out the variance estimation in practice, when z_k depends on a finite number of parameters that can be

estimated by using the sample data, z_k will be replaced by its natural estimator $\hat{z}_k$. The estimated variance then differs from the true variance by a term whose order of magnitude is $n^{-1/2}$.

Deville's paper was once again innovative since it was followed by several other works on variance estimation. While Deville proposed to linearize the parameter of interest at the population level, Demnati and Rao (2004, 2010) derive the estimator directly from survey weights. This method is a simple way to calculate the influence function on the estimator. Graf (2011); Antal, Langel and Tillé (2011); Graf and Tillé (2014); and Vallée and Tillé (2019) derived the estimator by the sample selection indicators, allowing for both the non-linearity of the estimator and calibration. The results given by the different methods are not always identical because the weights may depend on the selection indicators, especially when the estimator is calibrated.

5.2 The tools

From the previous theory, we obtain rules for calculating linearized variables that allow the treatment of complex estimators in a simple fashion, by breaking them down. The "total" functional $T(M) = \sum_{i=1}^{N} y_i$ is the simplest. Since

$$T(M + t\delta_k) = \sum_{i=1}^{N} y_i M(i) + \sum_{i=1}^{N} y_i t\,\delta_k(i) = \sum_{i=1}^{N} y_i + t y_k,$$

it follows that $z_k = IT(M,k) = y_k$. The expression (5.1) is here tautological, with $R_\epsilon = 0$.

Various useful properties are cited in the founding article.

(i) Let $T(M)$ and $S(M)$ be two functionals:

$$I(T+S)(M,k) = IT(M,k) + IS(M,k)$$

and

$$I(T \cdot S)(M,k) = IT(M,k) \cdot S(M,k) + IS(M,k) \cdot T(M,k).$$

(ii) Let $T(M)$ be a vector of totals in $\mathbb{R}^p$ and f be a differentiable function of $\mathbb{R}^p$ in $\mathbb{R}$ with the (p,p) matrix of partial derivatives evaluated at $T(M)$ is noted $D_{f,T(M)}$. Then, $If(T)(M,k) = D_{f,T(M)} \cdot IT(M,k)$. This rule is useful for finding well known linearized, of smooth functions of totals, such as ratios or linear correlation coefficients.

Now consider functionals parametrized by a vector α of $\mathbb{R}^p$, noted $T(M,\alpha)$. For example, if $p=1$ and $\alpha \in [0,1]$, noting F the distribution function associated with the distribution of y_k, $T(M,\alpha) = F^{-1}(\alpha)$ is the quantile of order α of the distribution.

(iii) If $T(M,\lambda)$ as a function of λ has sufficient regularity, a reciprocal functional $\Lambda(M,\alpha)$ can therefore be defined satisfying $T(M,\Lambda(M,\alpha)) = \alpha$. For example, to treat the variance of a quantile, we will consider $T(M,\lambda) = F(\alpha)$ and we obtain $\Lambda(M,\alpha) = F^{-1}(\alpha)$. It can be shown that

$$I\Lambda(M,\alpha,k) = -\left\{\frac{\partial T}{\partial \alpha}(M,\Lambda(M,\alpha))\right\}^{-1} \cdot IT(M,\alpha,k). \tag{5.2}$$

This theorem is useful for obtaining, for example, the linearization of a quantile or implicit parameter (see 5.3).

(iv) Suppose that the parameter is written as a function of the value y, that is, let's consider a functional of the type $T(M,\phi(y))$ where ϕ is a function with the right technical properties, then the functional $S(M) = \int T(M,\phi(y))\,dM$. So,

$$IS(M,k) = T(M,\phi(y_k)) + \int IT(M,\phi(y),k)\,dM. \tag{5.3}$$

This theorem is used, for example, to determine the linearized Gini coefficient (see 5.3).

(v) If the parameter is itself a functional $S(M)$, we can obtain the influence function of $T(M,S(M))$ – that is, the linearization of a compound of functionals – as

$$IT(M,S(M),k) = IT(M,\alpha,k) + \frac{\partial T}{\partial \alpha}(M,\alpha) \cdot IS(M,k),$$

where α takes the value $S(M)$ in the final expression.

We also mention an interesting property: for any functional of degree α, we have

$$\sum_{k=1}^{N} IT(M,k) = \alpha \cdot T(M).$$

In particular if $\alpha = 0$ (a ratio for example), the population total of the linearized variables z_k is equal to zero.

5.3 A few applications

The above theory makes it possible to linearize virtually all the estimators that are encountered in survey statistics. It is therefore general in scope, and ultimately permits all analytical variance calculations, for (about) every conceivable parameter, under some asymptotic conditions, that is, when the sample size n is considered to be "sufficiently large".

The original article by Jean-Claude Deville presents several examples of applications for complex parameters. It contains, along with the technical developments justifying them, the cases of implicit parameters, quantiles, the Gini coefficient, the poverty line (defined as the proportion of individuals in a population whose income is less than half its median), and the Kendall rank coefficient of correlation. The variance estimation for a principal component and that of the projection of a point representing any subpopulation on a factorial axis as part of a multiple correspondence analysis are also discussed. Below, we provide some results regarding implicit parameters, quantiles, and the Gini inequality coefficient.

An implicit parameter is a vector of $\mathbb{R}^p$ solution of an equation of the form $H(M,\mathbf{B})=0,$ where $H(M,\mathbf{B})=\sum_{k\in U} l_k(\mathbf{B}),$ the functions l_k being regular functions of $\mathbb{R}^p$ in $\mathbb{R}^p$. Using (5.2), and noting $\mathbf{B}_0$ the solution of the equation, we get for all k of U,

$$I\mathbf{B}(M,k) = -\left\{\frac{\partial H}{\partial \mathbf{B}}(M,\mathbf{B}_0)\right\}^{-1} \cdot IH(M,\mathbf{B}_0,k),$$

which means that

$$I\mathbf{B}(M,k) = -\left\{\sum_{k\in U}\frac{\partial l_k}{\partial \mathbf{B}}(M,\mathbf{B}_0)\right\}^{-1} \cdot l_k(\mathbf{B}_0).$$

This situation is, for example, that of regression coefficients, the l_k stemming from normal equations. If the regression is linear, it obtain (classical notations)

$$l_k(\mathbf{B}) = \frac{1}{\sigma_k^2} \cdot \mathbf{x}_k(y_k - \mathbf{z}_k^\top \mathbf{B})$$

and finally the linearized vector of $\mathbb{R}^p$

$$\mathbf{z}_k = I\mathbf{B}(M,k) = -\left(\sum_{k\in U}\frac{\mathbf{x}_k\mathbf{z}_k^\top}{\sigma_k^2}\right)^{-1} \cdot \frac{1}{\sigma_k^2} \cdot \mathbf{x}_k(y_k - \mathbf{z}_k^\top \mathbf{B}_0)$$

which can also be found through the "traditional" approach. In the case of logistic regression, the conventional tools are no longer sufficient and then

$$\mathbf{z}_k = I\mathbf{B}(M,k) = \left\{\sum_{k\in U}\mathbf{x}_k\mathbf{z}_k^\top \cdot f(\mathbf{z}_k^\top \mathbf{B}_0)(1-f(\mathbf{z}_k^\top \mathbf{B}_0))\right\}^{-1} \mathbf{x}_k(y_k - f(\mathbf{z}_k^\top \mathbf{B}_0)),$$

where $f(u)=e^u(1+e^u).$

To obtain the linearized variable of a quantile, the functional "distribution function" must be considered:

$$F(x) = T(M,x) = \frac{1}{N} \cdot \text{Card}\{k\in U / x_k \leqslant x\} = \frac{1}{N} \cdot \int 1_{x_k\leqslant x} dM,$$

whose influence function is

$$IT(M,x,k) = \frac{1}{N} \cdot \{1_{x_k\leqslant x} - F(x)\}.$$

Assuming for simplicity that this increasing function is differentiable and invertible, we define the quantile q_α of order $\alpha,$ where $\alpha\in[0,1],$ using $F(q_\alpha)=\alpha.$ Application of the formula (5.2) would then lead to the linearized variable

$$z_k = Iq_\alpha(M,k) = -[N \cdot F'(q_\alpha)]^{-1} \cdot (1_{x_k\leqslant q_\alpha} - \alpha).$$

Jean-Claude Deville has proposed a clever idea that makes it simple to take into account the fact that $F(x)$ is a step function, which is neither derivable nor invertible.

Let's finish by specifying the linearization of a "Gini Index", which is a standard inequality index. The index considered in the article is defined as

$$\text{GINI} = \frac{1}{t_x}\int xF(x)\,dM,$$

where $t_x = \sum_{k\in U} x_k$. Using (5.3) we obtain

$$z_k = F(x_k)\cdot\frac{x_k - \overline{x}_{k,\text{inf}}}{t_x} - \text{GINI}\cdot\frac{x_k}{t_x},$$

by applying

$$\overline{x}_{k,\text{inf}} = \frac{\int x1_{x<x_k}\,dM}{\int 1_{x<x_k}\,dM}$$

which corresponds to the average of the x-values less than x_k.

6. Quota sampling

"Quota" sampling is not a survey method used frequently by national statistical institutes. The key argument is the resulting bias of the estimators, where smaller variances are not made available as opposed to with a well-chosen probabilistic method, given the large sample sizes commonly encountered in official statistics. Another more philosophical argument is the dependence of the quality of the estimate on a model, which is a set of simplifying assumptions of reality, and is sometimes overly simplifying. Jean-Claude Deville often made this second argument, which he considered to reflect a lack of neutrality that was not suitable for a national statistical institute, at least when it is possible to do otherwise (it may be argued that models are systematically used to deal with non-response, but this is unavoidable). It may be because this highly used empirical survey method (particularly in the private sector) is inherently risky and therefore controversial that Jean-Claude Deville felt the need to formalize the question. He appears to have been the first to do so in a comprehensive manner, writing an authoritative article in the journal *Survey Methodology* in 1991 (Deville, 1991). In this article, two types of models are distinguished: one model deals with sampling, another with the variable of interest. In each configuration, an estimator is given, its bias studied, and when possible, the author produces a theoretical expression of variance as well as an unbiased estimator of the variance.

6.1 The general framework

Recall the principle of quota sampling, by, for simplicity, restricting to the case of one or two so-called quota variables. Auxiliary information is made up of subpopulation sizes defined by the modalities of the

quota variables, which are therefore qualitative in nature. If there is a single qualitative variable with I modalities to define these subpopulations, we have the population count N_i for each modality i varying between 1 and I. If there are two qualitative variables with respectively I and J modalities, denoting by $N_{i,j}$ the population count of the cell (i, j), we know the marginal count $N_{i,.} = \sum_{j=1}^{J} N_{i,j}$ for any i varying between 1 and I as well as the marginal count $N_{.,j} = \sum_{i=1}^{I} N_{i,j}$ for all j varying between 1 and J ("marginal" quotas). It should be noted that availability of the cross counts $N_{i,j}$ ("cross" quotas) brings us back to the case of a single qualitative variable. The selection of the sample is done empirically, without a sampling frame, following a few collection guidelines to randomize the composition of the overall sample of size n as much as possible while imposing constraints on the n_i's (case of a quota variable) or assigning the $n_{i,j}$'s (case of two quota variables). These constraints, which will be referred to as "quota constraints", are set at the discretion of the survey statistician, but in practice, the most common method is the method of proportional quotas, where the idea is to ensure that the sample has the same structure as the population, i.e., $n_i = n\frac{N_i}{N}$ in the case of a single variable, or $n_{i,.} = n\frac{N_{i,.}}{N}$ for all i and $n_{.,j} = n\frac{N_{.,j}}{N}$ for all j in the context of cross quotas ("proportional quota" constraints). Although this is a reassuring standard, it does not correspond to the intuitive optimal situation, since it is always preferable to increase the sample sizes in cells with the greatest dispersion (refer to the Neyman optimum in stratified sampling – which differs from proportional allocation but is preferable).

6.2 A sampling model

The case of quotas on a single variable can be treated by pretending that we are in the case of a stratified sampling design with simple random sampling in each stratum, each stratum being associated with a modality i of the quota variable, whereby the sample size n_i is imposed. It is difficult to imagine other alternatives, so there is nothing original to add in this simple context.

The interesting case is that of marginal quotas. The model has two phases. First, the selection can be viewed as a simple random draw under constraint, the constraints being those imposed on the $n_{i,.}$ and on the $n_{.,j}$, i.e., the constraints of quotas. That is a pretty daring assumption, that assumes that interviewers are completely neutral in selecting their respective samples. Technically, a simple random draw under constraint reduces to considering that any sample not meeting quotas has a zero probability of selection and that all samples meeting quotas have the same probability of selection. Practically, it could be implemented by conducting successive and independent simple random draws, while rejecting samples until quota constraints are satisfied (rejective selection). Secondly, the existence of these constraints is somewhat forgotten.

Let $P_{i,j}$ be the weight $\frac{N_{i,j}}{N}$ of the cell (i, j). The objective is to estimate these weights. Indeed, the average $\overline{Y}$ of any variable of interest Y defined in the population is written $\overline{Y} = \sum_{(i,j)} \frac{N_{i,j}}{N} \cdot \overline{Y}_{i,j}$ and, under the postulated model, the true average $\overline{Y}_{i,j}$ in the cell (i, j) is estimated without bias by the simple average $\overline{y}_{i,j}$ in the sample intersecting this cell, so that the natural estimator by quotas of $\overline{Y}$ will be $\hat{\overline{Y}}_{\text{quota}} = \sum_{(i,j)} \hat{P}_{i,j} \cdot \overline{y}_{i,j}$ where $\hat{P}_{i,j}$ is an estimate of $P_{i,j}$ – as much as possible without bias. With a standard simple random draw

selection, the sample sizes per cell (i, j), i.e., $n_{i,j}$, follow a multinomial distribution. In the second stage of modeling, we postulate that, subject to the quota constraints, the distribution of the $n_{i,j}$'s remains mutinomial. In this context, the estimates $P_{i,j}$ are obtained, for example, using maximum likelihood estimation, which leads to maximizing the objective function $\prod_{(i,j)} P_{i,j}^{n_{i,j}}$. Solutions must be compatible with known marginal information, which also requires

$$\sum_{j=1}^{J} P_{i,j} = \frac{N_{i,.}}{N} \quad \text{for all} \quad i \quad \text{and} \quad \sum_{i=1}^{I} P_{i,j} = \frac{N_{.,j}}{N}, \quad \text{for all} \quad j.$$

The solutions obtained are in the form $\hat{P}_{i,j} = \frac{n_{i,j}}{n}(a_i + b_j)^{-1}$ (the unknown coefficients a_i and b_j are the Lagrange coefficients associated with the constraints). Since there are $I + J - 1$ independent constraints, for example, by imposing the identifier constraint $b_J = 0$, we obtain the a_i and the b_j, by solving the system (S)

$$\sum_{j=1}^{J} n_{i,j}(a_i + b_j)^{-1} = n\frac{N_{i,.}}{N}, \quad \text{for all} \quad i = 1, \ldots, I,$$

$$\sum_{i=1}^{I} n_{i,j}(a_i + b_j)^{-1} = n\frac{N_{.,j}}{N}, \quad \text{for all} \quad j = 1, \ldots, J-1.$$

At this stage, it is clear that if quotas are proportional (it should be remembered that this is the scenario almost systematically chosen in practice), the system is resolved by systematically choosing $a_i = 1$ and $b_j = 0$, in which case $\hat{P}_{i,j} = \frac{n_{i,j}}{n}$ and $\hat{\bar{Y}}_{\text{quota}} = \bar{y}$, the customary sample mean. This provides a well-known result. The case of non-proportional quotas provides a complex estimator, whose analytical form is not explicit but asymptotically unbiased, since the specification of the model – not discussed in the founding article – is not in question. This point should prompt caution because, if proportional quota constraints are actually respected on average with a simple random selection, it seems to us that the sampling model becomes *a priori* more fragile when quotas significantly move away from proportional quotas, because of the fact that the second phase of this model is more difficult to accept. If the first phase of the model is probably the only way to build a theoretical basis for treating quotas, we could on the other hand – this would be another exercise – try to maximize the density of the $n_{i,j}$ under the quota constraints to obtain the $\hat{P}_{i,j}$.

Jean-Claude Deville proposed a variance expression for $\hat{\bar{Y}}_{\text{quota}}$, the sampling design being the only randomization mechanism that generates variability here. It begins by breaking down the true averages per cell $\bar{Y}_{i,j}$ according to $\bar{Y}_{i,j} = A_i + B_j + E_{i,j}$ by imposing the constraints $B_J = 0$,

$$\sum_{j=1}^{J} N_{i,j} E_{i,j} = 0, \quad \text{for all} \quad i = 1, \ldots, I$$

and

$$\sum_{i=1}^{I} N_{i,j} E_{i,j} = 0, \quad \text{for all} \quad j = 1, \ldots, J-1.$$

Then, he introduces the coefficients a_i^0 and b_j^0 ensuring the equalities

$$E_p\left\{\frac{n_{i,j}}{n}(a_i+b_j)^{-1}\right\}=\frac{N_{i,j}}{N}(a_i^0+b_j^0)^{-1}.$$

Indeed, $\frac{n_{i,j}}{n}(a_i+b_j)^{-1}$ is an estimator of $P_{i,j}=\frac{N_{i,j}}{N}$ but it is not unbiased of $P_{i,j}$ in general, and the coefficients a_i^0 and b_j^0 satisfy the equations of the system (S)

$$\sum_{j=1}^{J}N_{i,j}(a_i^0+b_j^0)^{-1}=N_{i,.},\quad \text{for all}\quad i=1,\ldots,I,$$

$$\sum_{i=1}^{I}N_{i,j}(a_i^0+b_j^0)^{-1}=N_{.,j},\quad \text{for all}\quad j=1,\ldots,J-1.$$

If quotas are proportional, then $a_i=1$ and $b_j=0$ and given the multinomial model, we also have $a_i^0=1$ and $b_j^0=0$ (these values are solutions of the previous non-linear system in all cases, but they are not the appropriate values for non-proportional quotas). By applying

$$S_{i,j}^2=\frac{1}{N_{i,j}}\sum_{k\in(i,j)}(y_k-\bar{Y}_{i,j})^2,$$

a variance expression is obtained:

$$V(\hat{\bar{Y}}_{\text{quota}})=\frac{1}{n}\sum_{(i,j)}\frac{N_{i,j}}{N}\{E_{i,j}^2+(a_i^0+b_j^0)^{-1}S_{i,j}^2\}.$$

The optimal strategy is not one of proportional quotas, but one that "inflates" the quotas $n_{i.}$ and $n_{.j}$ for the modalities i and j corresponding to the high dispersions $S_{i,j}^2$. This rule, which is very intuitive, does not appear clearly in its principle if we stick to this variance calculation. On the other hand, it becomes evident when we condition on the sizes $n_{i,j}$.

When the quotas are proportional, we therefore have

$$V(\hat{\bar{Y}}_{\text{quota}})=\frac{1}{n}\sum_{(i,j)}\frac{N_{i,j}}{N}(E_{i,j}^2+S_{i,j}^2).$$

It is clearly in our interest to have $E_{i,j}=0$ for all (i,j), therefore an additive model. The article also proposes a variance estimator with low bias that is not complicated when quotas are proportional because it is obtained in this case from a classic residuals calculation in a certain standard linear regression.

6.3 Models for the variable of interest

This whole part is about taking a radically different view from the previous one, because this time the model pertains to the variable of interest Y. This variable is considered to be random using the model-based approach developed by Royall (1970, 1976a, b, 1988) (see also Valliant, Dorfman and Royall, 2000;

Chambers and Clark, 2012). All the model versions considered in the article fall within the general framework of the linear model, of the type $\mathbf{Y} = \mathbf{XB} + \mathbf{u}$, where $\mathbf{B}$ is a vector of unknown real coefficients, $\mathbf{X}$ a matrix whose columns are explanatory variables, and $\mathbf{u}$ is a vector of errors with a mean equal to 0 and an unknown model variance $\mathbf{V}$.

In this context, a vector of values y_k, $k = 1, \ldots, N$, is composed of two vectors: $\mathbf{Y}_s$ of size n that includes the observed values y_k – those for which $k \in s$ – and $\mathbf{Y}_r$ of size $N - n$ that includes the unobserved values y_k. To define an optimal estimator of the true total $t_y = \sum_{k \in U} y_k$, it is natural to try to minimize the mean square error $\mathrm{E}_m(\hat{t} - t_y)^2$, where $\mathrm{E}_m(\cdot)$ denotes the expectation under the randomization with respect to the values y_k (which therefore has nothing to do with the sampling randomization). We also impose an unbiasedness condition of the form $\mathrm{E}_m(\hat{t} - t_y) = 0$. Finally, we are seeking a simple estimator, therefore linear, i.e., of the form of $\hat{t} = \mathbf{g}_s^\top \mathbf{Y}_s$, where $\mathbf{g}_s$ is a vector (to be found) of size n. The solution to this problem leads to

$$\hat{t}_{y,\mathrm{opti}} = \mathbf{1}_s^\top \mathbf{Y}_s + \mathbf{1}_r^\top \left\{ \mathbf{X}_r \hat{\mathbf{B}} + \mathbf{V}_{rs} \mathbf{V}_{ss}^{-1} (\mathbf{Y}_s - \mathbf{X}_s \hat{\mathbf{B}}) \right\},$$

where $\mathbf{1}_s$ (respectively $\mathbf{1}_r$) is a vector of size n (respectively $N - n$) of 1's, the terms $\mathbf{X}_s$, $\mathbf{X}_r$, $\mathbf{V}_{ss}$ and $\mathbf{V}_{rs}$ representing the sub-matrices of $\mathbf{X}$ and $\mathbf{V}$ formed by the rows and columns associated with the sets of indices s and r. The estimated coefficient $\hat{\mathbf{B}}$ is the general least squares estimator, i.e., $\hat{\mathbf{B}} = (\mathbf{X}_s^\top \mathbf{V}_{ss}^{-1} \mathbf{X}_s)^{-1} (\mathbf{X}_s^\top \mathbf{V}_{ss}^{-1} \mathbf{Y}_s)$. We will place ourselves *a priori* in the case where $\mathbf{V}_{rs} = 0$, a simplifying assumption that is easy to accept unless the selection involves clusters. As a result, the optimal estimator simplifies itself greatly since

$$\hat{t}_{y,\mathrm{opti}} = \mathbf{1}_s^\top \mathbf{Y}_s + \mathbf{1}_r^\top (\mathbf{X}_r \hat{\mathbf{B}}) = \sum_{k \in s} y_k + \sum_{k \in r} \hat{y}_k,$$

where $\hat{t}_{yk}$ is the k^{th} coordinate of the vector $\mathbf{X}_r \hat{\mathbf{B}}$, that is, the optimal predictor of the unknown value y_k. It is this expression that will subsequently be used to express the estimators arising from the quota method. It should be noted that if 1 is one of the regressors (which is extremely common), we have $\sum_{k \in s} y_k = \sum_{k \in s} \hat{y}_k$, so that $\hat{t}_{y,\mathrm{opti}} = \sum_{k \in U} \hat{y}_k$.

The estimator $\hat{t}_{y,\mathrm{opti}}$ is unbiased by construction in the sense that $\mathrm{E}_m(\hat{t}_{y,\mathrm{opti}} - t_y) = 0$. The assessment of its accuracy also relies on its variance, defined by $\mathrm{E}_p \mathrm{E}_m(\hat{t}_{y,\mathrm{opti}} - t_y)^2$, where $\mathrm{E}_p(\cdot)$ the expectation with respect to the sampling design. It is also equal to $\mathrm{E}_m \mathrm{E}_p(\hat{t}_{y,\mathrm{opti}} - t_y)^2$.

The combination of a sampling randomization and a model randomization is not easy to treat if the model on y_k changes when we have information on the membership – or non-membership – of the individual k in the selected sample. Otherwise, any calculation would become unmanageable, and moreover, we would not even be able to credibly formalize the model's dependence on the sample, which we recall, is random. Since it is not desirable (and not possible practically speaking) to seek to refine the assumptions beyond a certain degree, it is considered that the selection law of s and the one that generates Y are independent.

This is a non-informative model. Under these conditions, the model applied on y_k applies blindly to any individual k without there being any need to know if $k \in s$ or if $k \notin s$. This bias is essential to make variance calculations, but unfortunately, it is questionable. Typically, in the case of empirical surveys, one can have doubts, at least when the operation is not conducted under very tight control of the interviewers' practice, that there is no relationship between the values of Y and whether they belong to the sample. This is what creates the main risk of bias in the estimators and what feeds the traditional criticism of empirical sampling. But let's move on from this risk, accept it, and summarize, under these conditions, the theory set out in Jean-Claude Deville's article (1991).

The case of quotas based on a single qualitative variable and the case of cross-quotas are based on the same linear model, which is very simple: noting i the modality of the quota variable and k the individual identifier, we have $y_{i,k} = m_i + u_{i,k}$ with $\mathrm{E}_m(u_{i,k}) = 0$ and $\mathrm{E}_m(u_{i,k}^2) = \sigma_i^2$. This reflects the expected natural situation, where the qualitative variable (or the cross-classification of the two variables) explains the variable of interest well. The sampling technique makes it so that the sample sizes n_i in each modality are independent of the selected sample s. We have $\hat{y}_{i,k} = \hat{m}_i = \overline{y}_i$, the simple average of $y_{i,k}$ calculated on individuals in the sample belonging to the modality i. In this case, it is easy to check that the optimal unbiased estimator is

$$\hat{t}_{y,\text{opti}} = \sum_{i=1}^{I} N_i \overline{y}_i.$$

This is exactly the expression of the classical post-stratified estimator. It can be shown that

$$\mathrm{E}_m \mathrm{E}_p (\hat{t}_{y,\text{opti}} - t_y)^2 = \sum_{i=1}^{I} N_i^2 \left(1 - \frac{n_i}{N_i}\right) \frac{\sigma_i^2}{n_i}$$

by using the fact that the n_i's are independent of s, and then obtain an unbiased estimate of this variance.

Now let's look at marginal quotas in the context of an additive model, that is, for an individual k belonging to the modalities i and j of the two quota variables respectively: $y_{i,j,k} = \alpha_i + \beta_j + u_{i,j,k}$ with $\mathrm{E}_m(u_{i,j,k}) = 0$ and $\mathrm{E}_m(u_{i,j,k}^2) = \sigma_i^2 + \gamma_j^2$. The residuals are also considered to be mutually independent. The unbiased optimal estimator becomes

$$\hat{t}_{y,\text{opti}} = \sum_{i,j} n_{i,j} \overline{y}_{i,j} + \sum_{i,j} (N_{i,j} - n_{i,j})(\hat{\alpha}_i + \hat{\beta}_j).$$

Since the estimators of α_i and β_j are complicated to obtain in the general framework presented here, Jean-Claude Deville suggests using the least squares estimators associated with an ordinary model. In this context, after writing out the normal equations, and thus easily obtaining the $\hat{\alpha}_i$ and $\hat{\beta}_j$, a bit of algebra leads to

$$\hat{t}_{y,\text{opti}} = \sum_{i} N_{i.} \hat{\alpha}_i + \sum_{j} N_{.j} \hat{\beta}_j.$$

In the case of proportional quotas, we check that $\hat{t}_{y,\text{opti}} = N\bar{y}$. This is the estimator that is used classically in quota sampling, which is often used due to its simplicity. In the case of non-proportional quotas, we can always express analytically $\hat{t}_{y,\text{opti}}$, particularly as a linear combination of simple estimators per cell $\bar{y}_{i,j}$, i.e., $\hat{t}_{y,\text{opti}} = \sum_{i,j} \hat{N}_{i,j}\, \bar{y}_{i,j}$, but the expressions of $\hat{N}_{i,j}$ are complex.

When quotas are proportional, the true variance $V(\hat{t}_{y,\text{opti}})$ can be expressed in terms of $N_{i.}$, $N_{.j}$, σ_i^2 and γ_j^2, and a variance estimator can also be obtained as

$$\hat{\mathrm{V}}(\hat{t}_{y,\text{opti}}) = N^2 \frac{1-\frac{n}{N}}{n} \sum_{i,j} \frac{n_{i,j}}{n} s_{i,j}^2,$$

where $s_{i,j}^2$ is the standard dispersion of the $y_{i,j,k}$'s collected in the cell (i,j). It is unbiased in the sense that $\mathrm{E}_p\mathrm{E}_m\hat{\mathrm{V}}(\hat{t}_{y,\text{opti}}) = \mathrm{V}(\hat{t}_{y,\text{opti}})$. The feasibility of a variance calculation and its unbiased estimate is closely linked to the proportionality aspect of quotas: without this property, the development of the variance expression is no longer possible because the estimator $\hat{t}_{y,\text{opti}}$ depends on the sample sizes per cell $n_{i,j}$ and these are themselves depend on the drawn sample… based on an unknown probability, by definition of an empirical selection.

Finally, the article discusses the more general case of marginal quotas in a context where the model loses its additive aspect and becomes $y_{i,j,k} = \alpha_i + \beta_j + \gamma_{i,j} + u_{i,j,k}$. The optimal estimator is $\hat{t}_{y,\text{opti}} = \sum_{i,j} N_{i,j}\, \bar{y}_{i,j}$ but past this stage, we find ourselves in a context where not only do we not know $N_{i,j}$ but it does not even seem possible to produce a "natural" estimator. If the quotas are proportional, the practice – and clearly the only possible outcome – is to revert to the simple, standard expression $\hat{t} = N\bar{y}$. The price to pay is that of bias, which is

$$\mathrm{E}_p\mathrm{E}_m(N\bar{y} - t_y) = \mathrm{E}_p\left\{\sum_{i,j}\left(N\frac{n_{i,j}}{n} - N_{i,j}\right)\gamma_{i,j}\right\}.$$

There is no reason for it to be zero, but when the instructions are well designed and the collection meets them, we can hope that the ratios $n_{i,j}/n$ and $N_{i,j}/N$ will be numerically close, and therefore the bias will be small.

6.4 Operational conclusion

The conclusion that can be drawn on the treatment of quota sampling is as follows. If one uses a single quota variable or if one uses cross-classified quotas (which is the best!) by using multiple quota variables, there is no difficulty in performing point estimation without bias and then calculating a variance. For marginal quotas, with a sampling model, an estimator is obtained in all circumstances. It is asymptotically unbiased and its variance can always be estimated. With a model on the variable of interest, we do not know how to manage non-additive models. If the model is additive, an unbiased estimator can always be produced, but if the quotas are not proportional, then we don not know how to estimate the variance.

7. Conclusion

This article does not cover all of the original developments made by Jean-Claude Deville in the field of survey statistics. He has written numerous articles and made numerous presentations at conferences on other less significant aspects. As a result, his major works have themselves known a variety of fates. Calibration was certainly the most significant, best known and most widely used breakthrough worldwide, to the point that, today, there is no survey that is not calibrated on some external data, at least in the world of official statistics. The weight-share method is dependent on fairly specific survey conditions, but is nonetheless widely used, in particular, when repeated surveys are implemented over time and more specifically in the form of rotational sampling. The analytical development of variance expression for complex estimators regularly finds applications but relies on a rather complicated theory and has some serious competitors, namely the resampling methods. Balanced sampling has also had some success, but its use in official statistics is developing more slowly than calibration methods, probably because changing a sampling design is a much more consequential decision than changing an estimation procedure. Tillé (2011) already listed a list of applications 10 years after the method was developed. Interest in balanced sampling has since continued to grow. In France, the National Institute of Statistics and Economic Studies now makes a wide use of balanced sampling designs. As for the theory of quotas, this is mainly a clarification operation that helped somewhat demystify this technique that official statistics still looks at with some suspicion. Jean-Claude Deville has undoubtedly touched upon it all, and with brilliance, something that the scientific community readily recognized by awarding him the prestigious Waksberg Prize in 2018, just three years before his death.

References

Antal, E., Langel, M. and Tillé, Y. (2011). Variance estimation of inequality indices in complex sampling designs. Invited Lecture at the 58th Congress of the International Statistical Institute.

Bellhouse, D.R. (1988). A brief history of random sampling methods. *Handbook of Statistics Volume 6: Sampling*, (Eds., P.R. Krishnaiah and C.R. Rao), New York, Amsterdam. Elsevier/North-Holland, 1-14.

Bethlehem, J.G. (1988). Reduction of nonresponse bias through regression estimation. *Journal of Official Statistics*, 4, 251-260.

Bethlehem, J.G., and Keller, W.J. (1987). Linear weighting of sample survey data. *Journal of Official Statistics*, 3, 141-153.

Binder, D.A. (1983). On the variances of asymptotically normal estimators from complex survey. *International Statistical Review*, 51, 279-292.

Binder, D.A. (1991). Use of estimating functions for interval estimation from complex surveys. *Proceedings of the Survey Research Methods Section*, American Statistical Association, 34-42.

Binder, D.A. (1996). Linearization methods for single phase and two-phase samples: A cookbook approach. *Survey Methodology*, 22, 1, 17-22. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/1996001/article/14389-eng.pdf.

Binder, D.A., and Kovačević, M.S. (1995). Estimating some measures of income inequality from survey data: An application of the estimating equations approach. *Survey Methodology*, 21, 2, 137-145. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/1995002/article/14396-eng.pdf.

Boistard, H., Lopuhaä, H.P. and Ruiz-Gazen, A. (2012). Approximation of rejective sampling inclusion probabilities and application to high order correlations. *Electronic Journal of Statistics*, 6, 1967-1983.

Bousabaa, A., Lieber, J. and Sirolli, R. (1999). *La Macro Cube*. Technical report, Rennes: ENSAI.

Brewer, K. (2013). Three controversies in the history of survey sampling. *Survey Methodology*, 39, 2, 249-262. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2013002/article/11883-eng.pdf.

Chambers, R.L., and Clark, R.G. (2012). *An Introduction to Model-Based Survey Sampling with Applications*. Oxford: Oxford University Press.

Chang, T., and Kott, P.S. (2008). Using calibration weighting to adjust for nonresponse under a plausible model. *Biometrika*, 95, 555-571.

Chauvet, G. (2009). Stratified balanced sampling. *Survey Methodology*, 35, 1, 115-119. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2009001/article/10888-eng.pdf.

Chauvet, G., Bonnéry, D. and Deville, J.-C. (2011). Optimal inclusion probabilities for balanced sampling. *Journal of Statistical Planning and Inference*, 141(2), 984-994.

Chauvet, G., Deville, J.-C. and Haziza, D. (2011). On balanced random imputation in surveys. *Biometrika*, 98, 459-471.

Chauvet, G., Haziza, D. and Lesage, É. (2015). Examining some aspects of balanced sampling in surveys. *Statistica Sinica*, 25, 313-334.

Chauvet, G., and Tillé, Y. (2006a). A fast algorithm of balanced sampling. *Journal of Computational Statistics*, 21, 9-31.

Chauvet, G., and Tillé, Y. (2006b). Fastcube SAS-IML Macro. Université de Neuchâtel.

Chopin, N., and Ducrocq, G. (2021). Fast compression of MCMC output. *Entropy*, 23(8).

Choudhry, G.H., and Singh, M.P. (1979). Sampling with unequal probabilities and without replacement – A rejective method. *Survey Methodology*, 5, 2, 162-177. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/1979002/article/54834-eng.pdf.

Cordy, C.B. (1993). An extension of the Horvitz-Thompson theorem to point sampling from a continuous universe. *Statistics and Probability Letters*, 18(5), 353-362.

Deming, W.E., and Stephan, F.F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are know. *Annals of Mathematical Statistics*, 11, 427-444.

Demnati, A., and Rao, J.N.K. (2004). Linearization variance estimators for survey data. *Survey Methodology*, 30, 1, 17-26. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2004001/article/6991-eng.pdf.

Demnati, A., and Rao, J.N.K. (2010). Linearization variance estimators for model parameters from complex survey data. *Survey Methodology*, 36, 2, 193-201. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2010002/article/11381-eng.pdf.

Devaud, D., and Tillé, Y. (2019). Deville and Särndal's calibration: Revisiting a 25 years old successful optimization problem. *TEST*, 4, 1033-1065.

Deville, J.-C. (nda). 15ème round. cette fois y'a vraiment qêqchose. Manuscript internal note, Paris: Insee.

Deville, J.-C. (ndb). Comparaison des plans à probabilités inégales avec and sans remise. Manuscript internal note, Paris: Insee.

Deville, J.-C. (ndc). Échantillonnage à entropie max (rédaction rapide). Manuscript internal note, Paris: Insee.

Deville, J.-C. (1989). Une théorie simplifiée des sondages. *Les Ménages : Mélanges en L'honneur de Jacques Desabie*, (Eds., P. L'Hardy and C. Thélot), Paris: Insee, 191-214.

Deville, J.-C. (1991). A theory of quota surveys. *Survey Methodology*, 17, 2, 163-181. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/1991002/article/14504-eng.pdf.

Deville, J.-C. (1992). Constrained samples, conditional inference, weighting: Three aspects of the utilisation of auxiliary information. *Proceedings of the Workshop on the Uses of Auxiliary Information in Surveys*, Örebro. Statistics Sweden, 1-18.

Deville, J.-C. (1998a). La correction de la nonréponse par calage ou par échantillonnage équilibré. *Actes du Colloque de la Société Statistique du Canada,* Sherbrooke, Canada.

Deville, J.-C. (1998b). Les enquêtes par panel : en quoi différent-elles des autres enquêtes ? suivi de : Comment attraper une population en se servant d'une autre. *Actes des Journées de Méthodologie Statistique,* INSEE Méthodes, Paris: Insee, No. 84-85-86, 63-82.

Deville, J.-C. (1998c). Une nouvelle (encore une!) méthode de tirage à probabilités inégales. Working paper no 9804, Statistical Methodology, Paris: Insee.

Deville, J.-C. (1999). Variance estimation for complex statistics and estimators: Linearization and residual techniques. *Survey Methodology*, 25, 2, 193-203. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/1999002/article/4882-eng.pdf.

Deville, J.-C. (2000a). Generalized calibration and application to weighting for non-response. *CompStat, Proceedings in Computational Statistics: 14th Symposium Held in Utrecht, The Netherlands*, New York: Springer, 65-76.

Deville, J.-C. (2000b). Note sur l'algorithme de Chen, Dempster et Liu. Manuscript internal note, Rennes: CREST-ENSAI.

Deville, J.-C. (2002). La correction de la nonréponse par calage généralisé. *Actes des Journées de Méthodologie Statistique*, Paris: Insee-Méthodes, 3-20.

Deville, J.-C. (2014). Échantillonnage équilibré exact poissonnien. *8ème Colloque Francophone sur les Sondages*, Université de Bourgogne, Dijon, 1-6.

Deville, J.-C. (2015). Quelques éléments de géométrie et d'algèbre pour comprendre la nature d'un échantillonnage équilibré. *Actes des Journées de Méthodologie Statistique*, Paris: Insee-Méthodes, 1-8.

Deville, J.-C., and Dupont, F. (1993). Non-réponse : principes et méthodes. *Actes des Journées de Méthodologie Statistique*, Paris: Insee-Méthodes, 53-69.

Deville, J.-C., and Grosbras, J.-M. (1987). Algorithmes de tirage. *Les Sondages*, (Eds., J.-J. Droesbeke, B. Fichet and P. Tassi), Paris: Economica, 209-233.

Deville, J.-C., Grosbras, J.-M. and Roth, N. (1988). Efficient sampling algorithms and balanced sample. *CompStat, Proceedings in Computational Statistics 8th Symposium held in Copenhagen*, Heidelberg: Physica Verlag, 255-266.

Deville, J.-C., and Jacod, M. (1996). Replacing the traditional French census by a large scale continuous population survey. *Annual Research Conference*. US Bureau of Census.

Deville, J.-C., and Lavallée, P. (2006). Indirect sampling: The foundations of the generalized weight share method. *Survey Methodology*, 32, 2, 165-176. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2006002/article/9551-eng.pdf.

Deville, J.-C., and Maumy-Bertrand, M. (2006). Extension of the indirect sampling method and its application to tourism. *Survey Methodology*, 32, 2, 177-185. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2006002/article/9552-eng.pdf.

Deville, J.-C., and Qualité, L. (2005). Échantillonnage multidimensionnel (de plusieurs échantillons à la fois) à entropie maximum : définition, propriétés, algorithmes et programmes. *Actes des Journées de Méthodologie Statistique*, Paris: Insee-Méthodes, 1-8.

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

Deville, J.-C., Särndal, C.-E. and Sautory, O. (1993). Generalized raking procedure in survey sampling. *Journal of the American Statistical Association*, 88, 1013-1020.

Deville, J.-C., and Tillé, Y. (1998). Unequal probability sampling without replacement through a splitting method. *Biometrika*, 85, 89-101.

Deville, J.-C., and Tillé, Y. (2000a). Échantillonnage équilibré par la méthode du cube, variance et estimation de variance. *Actes des Journées de Méthodologie Statistique*, Paris: Insee-Méthodes, 15-35.

Deville, J.-C., and Tillé, Y. (2000b). Selection of several unequal probability samples from the same population. *Journal of Statistical Planning and Inference*, 86, 215-227.

Deville, J.-C., and Tillé, Y. (2001). Échantillonnage équilibré par la méthode du cube, variance et estimation de variance. *Enquêtes, Modèles et Applications*, (Eds., J.-J. Droesbeke and L. Lebart), Paris: Dunod, 444-362.

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

Deville, J.-C., and Tillé, Y. (2005). Variance approximation under balanced sampling. *Journal of Statistical Planning and Inference*, 128, 569-591.

Dupačová, J. (1979). A note on rejective sampling. *Contribution to Statistics (Jaroslav Hájek Memorial Volume)*, Academia Prague, 71-78.

Dupont, F. (1993). Calage et redressement de la non-réponse totale : validité de la pratique courante de redressement et comparaison des méthodes alternatives pour l'enquête sur la consommation alimentaire de 1989. *Actes des Journées de Méthodologie Statistique,* 15-16 December, 1993, Insee-Méthodes No 56-57-58, Complement, 9-42.

Durr, J.-M., and Dumais, J. (2002). Redesign of the french census of population. *Survey Methodology*, 28, 1, 43-49. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2002001/article/6414-eng.pdf.

Ernst, L.R. (1989). Weighting issues for longitudinal household and family estimates. *Panel Surveys*, (Eds., D. Kasprzyk, G. Duncan, G. Kalton and M.P. Singh), New York: John Wiley & Sons, Inc., 135-159.

Ernst, L.R., Hubble, D.L. and Judkins, D.R. (1984). Longitudinal family and household estimation in sipp. *Proceedings of the Survey Research Methods Section,* American Statistical Association, 682-687.

Estevao, V.M., and Särndal, C.-E. (2002). The ten cases of auxiliary information for calibration in twophase sampling. *Journal of Official Statistics*, 18, 233-255.

Estevao, V.M., and Särndal, C.-E. (2006). Survey estimates by calibration on complex auxiliary information. *International Statistical Review*, 74, 127-147.

Eustache, E., Vallée, A.-A. and Tillé, Y. (2022). Balanced donor imputation handling Swiss cheese nonresponse. *Statistica Sinica*, accepted.

Falorsi, P., Piersante, A. and Bako, B. (2016). Indirect sampling, a way to overcome the weakness of the lists in agricultural survey. *Proceedings ICAS VII: Seventh International Conference on Agricultural Statistics*, Rome, 24-26 October, 2016.

Francisco, C.A., and Fuller, W.A. (1991). Quantile estimation with a complex survey design. *The Annals of Statistics*, 454-469.

Fuller, W.A. (2009). Some design properties of a rejective sampling procedure. *Biometrika*, 96, 933-944.

Fuller, W.A., Legg, J.C. and Li, Y. (2017). Bootstrap variance estimation for rejective sampling. *Journal of the American Statistical Association*, 112(520), 1562-1570.

Gini, C., and Galvani, L. (1929). Di una applicazione del metodo rappresentativo al censimento italiano della popolazione (1. dicembre 1921). *Annali di Statistica*, Series 6, 4, 1-107.

Graf, É., and Tillé, Y. (2014). Variance estimation using linearization for poverty and social exclusion indicators. *Survey Methodology*, 40, 1, 61-79. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2014001/article/14000-eng.pdf.

Graf, M. (2011). Use of survey weights for the analysis of compositional data. *Compositional Data Analysis: Theory and Applications*, (Eds., V. Pawlowsky-Glahn and A. Buccianti), Chichester: Wiley, 114-127.

Grafström, A., and Lisic, J. (2019). *BalancedSampling: Balanced and Spatially Balanced Sampling*. R package version 1.5.5.

Grafström, A., Lundström, N.L.P. and Schelin, L. (2012). Spatially balanced sampling through the pivotal method. *Biometrics*, 68(2), 514-520.

Grafström, A., and Tillé, Y. (2013). Doubly balanced spatial sampling with spreading and restitution of auxiliary totals. *Environmetrics*, 14(2), 120-131.

Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *Annals of Mathematical Statistics*, 35, 1491-1523.

Hájek, J. (1981). *Sampling from a Finite Population*. New York: Marcel Dekker.

Hampel, F.R., Ronchetti, E., Rousseeuw, P.J. and Stahel, W.A. (1985). *Robust Statistics: The Approach Based on the Influence Function*. New York: John Wiley & Sons, Inc.

Hasler, C., and Tillé, Y. (2014). Fast balanced sampling for highly stratified population. *Computational Statistics and Data Analysis*, 74, 81-94.

Hasler, C., and Tillé, Y. (2016). Balanced k -nearest neighbor imputation. *Statistics*, 105, 11-23.

Haziza, D., and Beaumont, J.-F. (2017). Construction of weights in surveys: A review. *Statistical Science*, 32(2), 206-226.

Haziza, D,. and Lesage, É. (2016). A discussion of weighting procedures for unit nonresponse. *Journal of Official Statistics*, 32(1), 129-145.

Holt, D., and Smith, T.M.F. (1979). Post-stratification. *Journal of the Royal Statistical Society*, A142: Part 1, 33-46.

Huang, E.T., and Fuller, W.A. (1978). Non-negative regression estimation for sample survey data. *Proceedings of the Social Statistics Section,* American Statistical Association, 300-305.

Huang, H. (1984). Obtaining cross-sectional estimates from a longitudinal survey: Experiences of the income survey development program. *Proceedings of the Social Statistics Section,* American Statistical Association, 670-675.

Isaki, C.T., and Fuller, W.A. (1982). Survey design under a regression population model. *Journal of the American Statistical Association*, 77, 89-96.

Jauslin, R., Eustache, E., Panahbehagh, B. and Tillé, Y. (2021). *StratifiedSampling: Different Methods for Stratified Sampling*. R Foundation for Statistical Computing, Vienna, Autricha. R package version 0.3.0.

Jauslin, R., Eustache, E. and Tillé, Y. (2021). Enhanced cube implementation for highly stratified population. *Japanese Journal of Statistics and Data Science*, 4, 783-795.

Jauslin, R., and Tillé, Y. (2020a). Spatial spread sampling using weakly associated vectors. *Journal of Agricultural, Biological and Environmental Statistics*, 25(3), 431-451.

Jauslin, R., and Tillé, Y. (2020b). *WaveSampling: Weakly Associated Vectors Sampling*. R package version 0.1.1 http://CRAN.R-project.org/package=WaveSamplinghttp://CRAN.R-project.org/package=WaveSampling.

Judkins, D., Hubble, D., Dorsch, J., McMillen, D. and Ernst, L. (1984). Weighting of persons for sipp longitudinal tabulations. *Proceedings of the Social Statistics Section,* American Statistical Association, 676-687.

Kalton, G., and Brick, J. (1995). Weighting schemes for household panel surveys. *Survey Methodology*, 21, 1, 33-44. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/1995001/article/14412-eng.pdf.

Kiær, A.N. (1896). Observations et expériences concernant des dénombrements représentatifs. *Bulletin de l'Institut International de Statistique*, 9, 176-183.

Kiær, A.N. (1899). Sur les méthodes représentatives ou typologiques appliquées à la statistique. *Bulletin de l'Institut International de Statistique*, 11, 180-185.

Kiær, A.N. (1903). Sur les méthodes représentatives ou typologiques. *Bulletin de l'Institut International de Statistique*, 13, 66-78.

Kiær, A.N. (1905). Discours sans intitulé sur la méthode représentative. *Bulletin de l'Institut International de Statistique*, 14, 119-134.

Kiesl, H. (2010). Selecting kindergarten children by three stage indirect sampling. *Proceedings of the Survey Research Methods Section,* American Statistical Association, 2730-2738.

Kott, P.S. (2006). Using calibration weighting to adjust for nonresponse and coverage errors. *Survey Methodology*, 32, 2, 133-142. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2006002/article/9547-eng.pdf.

Kott, P.S., and Chang, T. (2010). Using calibration weighting to adjust for nonignorable unit nonresponse. *Journal of the American Statistical Association*, 105(491), 1265-1275.

Langel, M., and Tillé, Y. (2011). Corrado Gini, a pioneer in balanced sampling and inequality theory. *Metron*, 69, 45-65.

Lavallée, P. (1995). Cross-sectional weighting of longitudinal surveys of individuals and households using the weight share method. *Survey Methodology*, 21, 1, 25-32. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/1995001/article/14413-eng.pdf.

Lavallée, P. (2001). La méthode généralisée du partage des poids et le calage sur marges. *Enquêtes, Modèles et Applications*, (Eds., J.-J. Droesbeke and L. Lebart), Paris: Dunod, 396-403.

Lavallée, P. (2002). *Le Sondage Indirect ou la Méthode Généralisée du Partage des Poids*. (Ed., l'Université de Bruxelles, Paris: Ellipses.

Lavallée, P. (2007). *Indirect Sampling*. New York: Springer.

Legg, J.C., and Yu, C.L. (2010). A comparison of sample set restriction procedures. *Survey Methodology*, 36, 1, 69-79. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2010001/article/11249-eng.pdf.

Lemel, Y. (1976). Une généralisation de la méthode du quotient pour le redressement des enquêtes par sondages. *Annales de l'Insee*, 22-23, 273-281.

Lesage, É., Haziza, D. and D'Haultfoeuille, X. (2019). A cautionary tale on instrumental calibration for the treatment of nonignorable unit nonresponse in surveys. *Journal of the American Statistical Association*, 114, 906-915.

Leuenberger, M., Eustache, E., Jauslin, R. and Tillé, Y. (2022). Balancing a sample almost perfectly. *Statistics and Probablity Letters*, 180, 109229.

Lundström, S., and Särndal, C.-E. (1999). Calibration as a standard method for treatment of nonresponse. *Journal of Official Statistics*, 15, 305-327.

Medous, E., Goga, C., Ruiz-Gazen, A., Beaumont, J.-F., Dessertaine, A. and Puech, P. (2023). Many-to-one indirect sampling with application to the french postal traffc estimation. *The Annals of Applied Statistics*, 17(1), 838-859.

Nedyalkova, D., and Tillé, Y. (2008). Optimal sampling and estimation strategies under linear model. *Biometrika*, 95, 521-537.

Nedyalkova, D., and Tillé, Y. (2012). Bias robustness and efficiency in model-based inference. *Statistica Sinica*, 22, 777-794.

Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97, 558-606.

Royall, R.M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika*, 57, 377-387.

Royall, R.M. (1976a). Likelihood functions in finite population sampling theory. *Biometrika*, 63, 605-614.

Royall, R.M. (1976b). The linear least squares prediction approach to two-stage sampling. *Journal of the American Statistical Association*, 71, 657-664.

Royall, R.M. (1988). The prediction approach to sampling theory. *Handbook of Statistics, Sampling*, (Eds., P.R. Krishnaiah and C.R. Rao), Amsterdam: Elsevier, Volume 6, 399-413.

Rubin, D.B. (1976). Inference and missing data. *Biometrika*, 63, 581-592.

Särndal, C.-E. (2007). The calibration approach in survey theory and practice. *Survey Methodology*, 33, 2, 99-119. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2007002/article/10488-eng.pdf.

Särndal, C.-E., and Lundström, S. (2005). *Estimation in Surveys with Nonresponse*. New York: John Wiley & Sons, Inc.

Särndal, C.-E., Swensson, B. and Wretman, J.H. (1992). *Model Assisted Survey Sampling*. New York: Springer.

Sautory, O., and Le Guennec, J. (2003). La macro CALMAR2 : redressement d'un échantillon par calage sur marges – documentation de l'utilisateur. Technical report, Paris: Insee.

Stephan, F.F. (1942). An iterative method of adjusting sample frequency data tables when expected marginal totals are known. *Annals of Mathematical Statistics*, 13, 166-178.

Thionet, P. (1953). *La Théorie des Sondages*. Institut National de la Statistique et des Études Économiques, Études théoriques vol. 5, Paris: Imprimerie nationale.

Thomsen, I. (1978). A second note on the efficiency of weighting subclass means to reduce the effects of non-response when analyzing survey data. *Statistisk Tidskrift*, 16, 278-285.

Tillé, Y. (2011). Ten years of balanced sampling with the cube method: An appraisal. *Survey Methodology*, 37, 2, 215-226. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2011002/article/11609-eng.pdf.

Tillé, Y. (2016). The legacy of Corrado Gini in survey sampling and inequality theory. *Metron*, 74(2), 167-174.

Tillé, Y., and Matei, A. (2021). *Sampling: Survey Sampling*. R package version 2.9.

Vallée, A.-A., and Tillé, Y. (2019). Linearization for variance estimation by means of sampling indicators: Application to nonresponse. *International Statistical Review*, 87(2), 347-367.

Valliant, R., Dorfman, A.H. and Royall, R.M. (2000). *Finite Population Sampling and Inference: A Prediction Approach*. New York: John Wiley & Sons, Inc.

Wolter, K.M. (1985). *Introduction to Variance Estimation*. New York: Springer.

Woodruff, R.S. (1971). A simple method for approximating the variance of a complicated estimate. *Journal of the American Statistical Association*, 66, 411-414.

Wu, C., and Sitter, R.R. (2001). A model calibration approach to using complete auxiliary information from survey data. *Journal of the American Statistical Association*, 96, 185-193.

Yates, F. (1949). *Sampling Methods for Censuses and Surveys*. London: Charles Griffin.

Zhang, S., Han, P. and Wu, C. (2022). Calibration techniques encompassing survey sampling, missing data analysis and causal inference. *International Statistical Review*, accepted for publication.