Published in International Statistical Review 66, issue 3, 303-322, 1998
which should be used for any reference to this work

Estimation in Surveys Using Conditional Inclusion Probabilities: Simple Random Sampling

Yves Tillé

CR ST- SAI, École ationale de la Statistique et de l'Analyse de l'Information, rue Blaise Pascal, Campus de Ker Lann, 35170 Bruz, France. e-mail: tille@ensai.fr

Summary

In survey sampling, auxiliary information on the population is often available. The aim of this paper is to develop a method which allows one to take into account such auxiliary information at the estimation stage by means of conditional bias adjustment. The basic idea is to attempt to construct a conditionally unbiased estimator. Four estimators that have a small conditional bias with respect to a statistic are proposed. It is shown that many of the estimators used in the literature in the case of simple random sampling can be obtained by using this estimation principle. The problem of simple random sampling with replacement, poststratification, and adjustment of a 2 × 2 dimensional contingency table to marginal totals are discussed in the conditional framework. Finally it is shown that the regression estimator can be viewed as an approximation of an application of the conditional principle.

Key words: Simple random sampling; Conditional estimation; Weighted observation.

1 Introduction

In survey sampling theory, conditional inference has been discussed especially in the context of post-stratification. Holt & Smith (1979) advocate in this case that "inferences should be made on the achieved sample configuration of the sample post-stratum frequencies". They argue for conditional inference after selecting the sample and point out that the post-stratified estimator "offers protection against extreme sample configurations". Other authors have considered how to obtain estimators which have small or no conditional bias. The idea of suppressing the conditional bias has been applied to the ratio estimator by Robinson (1987) who computes the conditional bias of the ratio estimator under assumptions of normality and then corrects it by using an estimator of the conditional bias. The computation and estimation of the conditional bias is presented as an estimation principle by Rao (1985) who applies a conditional bias adjustment to the mean estimator, the ratio estimator in simple random sampling and stratified sampling design and for domain estimation. More recently Rao (1994) has given a general set-up for estimation using auxiliary information and compares the calibration methods to the conditional approach. The aim of this paper is to propose a general estimator based on conditional inclusion probabilities that allows us to construct directly an estimator with a small conditional bias with respect to a statistic. It is also shown that many classical estimators used in survey sampling can be derived from this general estimator.

Before defining this general estimator, we give an example to show that the basic idea of conditional estimation is very intuitive. Suppose that a sample of size $n = 200$ is drawn by means of simple random sampling in a population of size $N = 2000$ composed of $N_1 = 1000$ males and $N_2 = 1000$

females. The aim is to estimate the average height $\bar{z}$ in the population. Denote by $\bar{z}_1$ and $\bar{z}_2$ respectively the male and female average heights and suppose that the sample obtained is composed of $n_1 = 90$ males of average height $\widehat{\bar{z}}_1 = 172$ and $n_2 = 110$ females of average height $\widehat{\bar{z}}_2 = 164$. Two estimators can be used to estimate the average height.

$$\widehat{\bar{z}}_A = \frac{n_1 \times \widehat{\bar{z}}_1 + n_2 \times \widehat{\bar{z}}_2}{n} = \frac{172 \times 90 + 164 \times 110}{200} = 167.6, \tag{1}$$

or

$$\widehat{\bar{z}}_B = \frac{N_1 \times \widehat{\bar{z}}_1 + N_2 \times \widehat{\bar{z}}_2}{N} = \frac{172 \times 1000 + 164 \times 1000}{2000} = 168. \tag{2}$$

Estimator $\widehat{\bar{z}}_B$ looks more appropriate than $\widehat{\bar{z}}_A$ because it includes a correction with respect to the real ratio males/females. Estimator $\widehat{\bar{z}}_A$ is the usual sample mean which is unconditionally unbiased for $\widehat{\bar{z}}_A$, but which is biased conditional on the number of males and females in the sample. Estimator $\widehat{\bar{z}}_B$ is well-known to survey statisticians and is called the *poststratified estimator*, is unconditionally unbiased for $\bar{z}$ and has a very slight conditional bias. We show below that the poststratified estimator can be obtained from the sample mean by a conditional bias adjustment. The aim of the paper is to formalise this idea and to show that, in survey sampling, many classical estimators can be viewed also as applications of a similar conditional bias adjustment.

The opportunity to apply a conditional bias adjustment ensues practically from the existence of auxiliary information that allows us to estimate a conditional expectation with respect to a statistic. We shall refer to this statistic as an *auxiliary statistic* and denote it by η. The objective is to decrease the variance of an estimator in the following way. Let $\hat{\theta}$ be an unbiased estimator of a parameter θ. If $B(\hat{\theta} \mid \eta) = E(\hat{\theta} \mid \eta) - \theta$, the conditional bias of $\hat{\theta}$ given η is known, it is possible to construct the adjusted estimator $\hat{\theta}^* = \hat{\theta} - B(\hat{\theta} \mid \eta)$. We get

$$Var(\hat{\theta}^*) = Var(\hat{\theta}) + Var\left\{B(\hat{\theta} \mid \eta)\right\} - 2Cov\left\{\hat{\theta}, B(\hat{\theta} \mid \eta)\right\}.$$

Since

$$\begin{aligned} Cov(\hat{\theta}, B(\hat{\theta} \mid \eta)) &= EE\left\{(\hat{\theta} - \theta)(E(\hat{\theta} \mid \eta) - \theta) \mid \eta\right\} \\ &= Var\left\{E(\hat{\theta} \mid \eta)\right\}, \end{aligned}$$

we obtain

$$Var(\hat{\theta}^*) = Var(\hat{\theta}) - Var\left\{E(\hat{\theta} \mid \eta)\right\}. \tag{3}$$

The variance of the adjusted estimator $\hat{\theta}^*$ will thus be no greater than that of the original estimator $\hat{\theta}$. The problem of this approach is that the conditional bias must in general be estimated, and that the benefits gained in attempting to decrease the conditional bias could be thwarted by the instability of the conditional bias estimator used.

This approach can be applied to the estimator $\widehat{\bar{z}}_A$ defined in (1). If the statistic used to compute the conditional bias is $\eta = (n_1, n_2)$, and if it is supposed that the probability of obtaining an empty poststratum is negligible, we obtain

$$E\left(\widehat{\bar{z}}_A \middle| n_1, n_2\right) = \frac{1}{n}\left(n_1\bar{z}_1 + n_2\bar{z}_2\right),$$

and thus

$$B\left(\widehat{\bar{z}}_A \middle| n_1, n_2\right) = \frac{1}{n}\left(n_1\bar{z}_1 + n_2\bar{z}_2\right) - \frac{1}{N}\left(N_1\bar{z}_1 + N_2\bar{z}_2\right),$$

which can be estimated by

$$\widehat{B}\left(\widehat{\bar{z}}_A \middle| n_1, n_2\right) = \frac{1}{n}\left(n_1\widehat{\bar{z}}_1 + n_2\widehat{\bar{z}}_2\right) - \frac{1}{N}\left(N_1\widehat{\bar{z}}_1 + N_2\widehat{\bar{z}}_2\right).$$

Finally we obtain the adjusted estimator

$$\widehat{\bar{z}}_A - \widehat{B}\left(\widehat{\bar{z}}_A \middle| n_1, n_2\right) = \widehat{\bar{z}}_B.$$

Estimator $\widehat{\bar{z}}_B$ is thus obtained from $\widehat{\bar{z}}_A$ by suppressing the conditional bias.

The approach proposed in this paper is a formalization of an approach that appears in numerous publications about survey sampling. The construction of a conditionally unbiased estimator is treated as a technique to develop a more precise estimator. The original feature of this paper is that, instead of obtaining $\hat{\theta}^*$ from $\hat{\theta}$ by a conditional bias adjustment, we propose to construct conditionally unbiased estimators directly using conditional inclusion probabilities. The existence of a conditionally unbiased estimator with respect to a statistic is discussed in Section 2. Four conditionally unbiased estimators are proposed. These estimators generalise the poststratified estimators and can be used to build a method of adjustment of a 2×2 table to known marginal totals (Section 3). Next it is shown that the regression estimator is an approximation of one of these estimators (Section 4). Finally the issues of conditional bias adjustment is discussed in Section 5.

2 Conditionally Unbiased Estimators

2.1 Existence of a Conditionally Unbiased Estimator

Suppose that a random sample S is drawn without replacement from a finite population $U = \{1, \ldots, k, \ldots, N\}$ following a sampling design $p(.)$. The probability of selecting the sample s is $Pr(S = s) = p(s)$ for all non-empty $s \subset U$. The indicator variables I_k take the value 1 if unit k is in the sample and 0 if not, for all $k \in U$. The probability that unit k is selected is $\pi_k = E(I_k)$ and is called the first-order inclusion probability. The probability of selecting both units k and ℓ is $\pi_{k\ell} = E(I_k I_\ell)$ and is called the second-order inclusion probability. Consider y_k and x_k, respectively the value of the variable of interest y and the value of the auxiliary variable x on the kth unit. The values x_k are assumed known for all $k \in U$.

The Horvitz–Thompson estimator (1952) of the mean

$$\bar{y} = \frac{1}{N}\sum_{k \in U} y_k \tag{4}$$

is

$$\hat{\bar{y}}_\pi = \frac{1}{N}\sum_{k \in S} \frac{y_k}{\pi_k}.$$

Let $\eta = \eta(x_k, k \in S)$ be a statistic. Since the population is finite, the statistic η takes a finite number of possible values denoted $\{\eta_1, \ldots, \eta_i, \ldots, \eta_I\}$. The objective is to estimate $\bar{y}$ with a conditional bias as small as possible with respect to η. Define the first-order conditional inclusion probabilities to be $\pi_{k|\eta} = E(I_k \mid \eta)$ for all $k \in U$ and the conditional joint inclusion probabilities to be $\pi_{k\ell|\eta} = E(I_k I_\ell \mid \eta)$ for all $k \in U, \ell \in U, k \neq \ell$.

An estimator constructed with conditional inclusion probabilities will be called a *conditionally weighted* (CW) *estimator*. First, we introduce the *simple* CW (SCW) estimator given by

$$\hat{\bar{y}}_{|\eta} = \frac{1}{N}\sum_{k \in S} \frac{y_k}{\pi_{k|\eta}}. \tag{5}$$

The conditional inclusion probabilities are assumed to be calculable for any possible value of η. The statistic η is called the *auxiliary statistic*. The conditional bias of the SCW-estimator is given by :

$$\begin{aligned} B(\hat{\bar{y}}_{|\eta} \mid \eta) &= E(\hat{\bar{y}}_{|\eta} \mid \eta) - \bar{y} \\ &= \frac{1}{N} \sum_{\substack{k \in U \\ \pi_{k|\eta}>0}} E\left(\frac{y_k I_k}{\pi_{k|\eta}} \middle| \eta \right) - \bar{y} \\ &= -\frac{1}{N} \sum_{k \in U} y_k I[\pi_{k|\eta} = 0], \end{aligned} \tag{6}$$

where $I[.]$ is the indicator function given by

$$I[\pi_{k|\eta} = 0] = \begin{cases} 1 \text{ if } \pi_{k|\eta} = 0 \\ 0 \text{ if } \pi_{k|\eta} > 0. \end{cases}$$

Remember that a well-known result of sampling theory is that a necessary condition for the existence of an unbiased estimator of $\bar{y}$ is that $\pi_k > 0$, for all $k \in U$. This result can be transposed conditionally to η and gives that a necessary condition for the existence of a conditionally unbiased estimator of $\bar{y}$ is that $\pi_{k|\eta} > 0$, for all $k \in U$, and for all possible values of η.

Note that in the general case, (see Example 2) the conditional inclusion probabilities can take the value 0 even if the unconditional inclusion probabilities are strictly positive and thus an exactly conditionally unbiased estimator generally does not exist. A weaker definition of conditional unbiasedness will also be considered.

DEFINITION 1: *An estimator of $\bar{y}$ denoted $\hat{\bar{y}}$ is said to be virtually conditionally unbiased (VCU) with respect to a statistic η if*

$$B(\hat{\bar{y}} \mid \eta) = \sum_{k \in U} y_k \alpha_k(\eta) I(\pi_{k|\eta} = 0),$$

for all $(y_1 \ \dots \ y_k \ \dots \ y_N) \in \boldsymbol{R}^N$, where the coefficients $\alpha_k(\eta)$ can depend on η.

When an estimator is VCU, this does not necessarily imply that the conditional bias is near zero but that the conditional bias only depends on the units having null conditional inclusion probabilities. Expression (6) directly shows that the SCW-estimator is VCU.

EXAMPLE 1: If a simple random sample is drawn without replacement from a population of size N with random size n_S where $n_S > 0$, then if $\eta = n_S$, we obtain $\pi_{k|\eta} = E(I_k|n_S) = n_S/N, k \in U$. Since $n_S/N > 0, k \in U$, an exactly conditionally unbiased estimator with respect to n_S always exists.

EXAMPLE 2: If a simple random sample of size $n = 2$ is drawn without replacement from a population $U = \{1, 2, 3, 4, 5\}$ of size $N = 5$ and if the statistic used is

$$\eta = k_1 + k_2$$

where k_1 and k_2 are the index numbers of the two units drawn in the sample, the possible values for the conditional inclusion probabilities are given in Table 1.

Table 1

Conditional inclusion probabilities: Example 2

η_i	$Pr(\eta = \eta_i)$	values of $\pi_{k\|\eta}$ k=1	2	3	4	5
3	1/10	1	1	0	0	0
4	1/10	1	0	1	0	0
5	1/5	1/2	1/2	1/2	1/2	0
6	1/5	1/2	1/2	0	1/2	1/2
7	1/5	0	1/2	1/2	1/2	1/2
8	1/10	0	0	1	0	1
9	1/10	0	0	0	1	1

So, if η takes the value 5, either the sample $\{1,4\}$ or $\{2,3\}$ was selected, thus $E[I_5|\eta = 5] = 0$, and $E[I_k|\eta = 5] = 1/2, k \neq 5$. In this case, many conditional inclusion probabilities equal zero and thus an exactly conditionally unbiased estimator does not exist in the class of linear estimators.

2.2 *Other Conditionally Weighted Estimators*

It is possible to derive other CW-estimators. First, it is always possible to construct an unconditionally unbiased estimator given by

$$\hat{\bar{y}}_{c|\eta} = \frac{1}{N} \sum_{k \in S} \frac{y_k}{h_k \pi_{k|\eta}} \tag{7}$$

where

$$h_k = EI[\pi_{k|\eta} > 0] = Pr\left(\pi_{k|\eta} > 0\right).$$

Estimator (7) will be called the corrected CW (CCW) estimator. Its conditional bias is

$$B(\hat{\bar{y}}_{c|\eta}|\eta) = \frac{1}{N} \sum_{k \in U} y_k \left(\frac{I[\pi_{k|\eta} > 0]}{h_k} - 1 \right). \tag{8}$$

The CCW-estimator is not VCU but it is unconditionally unbiased. Indeed, by (8), we get

$$B(\hat{\bar{y}}_{c|\eta}) = EB(\hat{\bar{y}}_{c|\eta} \mid \eta) = 0.$$

Both SCW and CCW-estimators can be criticised because they are not invariant by translation, i.e. these estimators do not increase by a value C when all the units y_k are increased by a value C. Indeed, for the SCW-estimator, we obtain

$$\frac{1}{N} \sum_{k \in S} \frac{y_k + C}{\pi_{k|\eta}} = \hat{\bar{y}}_{c|\eta} + \frac{C}{N} \sum_{k \in S} \frac{1}{\pi_{k|\eta}} \neq \hat{\bar{y}}_{c|\eta} + C.$$

To solve this problem, two 'ratio' versions of $\hat{\bar{y}}_{|\eta}$ and $\hat{\bar{y}}_{c|\eta}$ can be constructed.

1. the SCW-ratio

$$\hat{\bar{y}}_{r|\eta} = \left(\sum_{k \in S} \frac{1}{\pi_{k|\eta}} \right)^{-1} \sum_{k \in S} \frac{y_k}{\pi_{k|\eta}}, \tag{9}$$

2. the CCW-ratio

$$\hat{\bar{y}}_{cr|\eta} = \left(\sum_{k \in S} \frac{1}{h_k \pi_{k|\eta}} \right)^{-1} \sum_{k \in S} \frac{y_k}{h_k \pi_{k|\eta}}. \tag{10}$$

They are invariant by translation. This property is appreciated by practitioners who consider it is more important to have an estimator where all the sums of percentages equal one hundred than a conditionally unbiased one. An approximation of their bias and conditional bias can be found by means of the linearization technique used for the research of a bias in the classical ratio estimator (see for instance Cochran, 1977, p. 161).

Among the four CW-estimators, which one should be chosen? Since a conditionally unbiased estimator rarely exists, it will usually be necessary to allow for a slight conditional bias. It is also interesting to note that it is always possible to correct the CW-estimator in order that it be unconditionally unbiased. Nevertheless, this correction enlarges the conditional bias and thus increases the MSE. For this reason, we prefer to use the estimator (5) and (9). We advocate the use of SCW-ratio when the sum of the inverses of the inclusion probabilities is not equal to N.

2.3 *Conditional Inclusion Probabilities*

To construct the CW-estimators, the conditional inclusion probabilities must be evaluated. Using Bayes's theorem, we obtain

$$\begin{aligned} E(I_k \mid \eta = \eta_i) &= Pr(k \in S \mid \eta = \eta_i) \\ &= \pi_k \frac{Pr(\eta = \eta_i \mid k \in S)}{Pr(\eta = \eta_i)}, i = 1, \dots, I. \end{aligned}$$

To compute the conditional inclusion probability, the probability distribution of η unconditionally and conditionally on the presence of each unit in the sample must be known. In the perspective of conditionally weighted estimation, the auxiliary information required is the knowledge of the probability distribution of the auxiliary statistic η. The probability distribution of η can theoretically be derived from the sampling design. Indeed,

$$Pr(\eta = \eta_i) = \sum_{s \mid \eta = \eta_i} p(s)$$

and

$$Pr(\eta = \eta_i \mid k \in S) = \frac{Pr(\eta = \eta_i \wedge k \in S)}{\pi_k} = \frac{1}{\pi_k} \sum_{\substack{s \mid \eta = \eta_i \\ s \ni k}} p(s).$$

In practice, it is far from self-evident exactly how the conditional inclusion probabilities are to be computed. We shall see that, in some cases, conditional inclusion probabilities can be calculated exactly; in other cases, approximations must be used.

3 Applications to Simple Random Sampling

3.1 *Sampling with Replacement*

One of the simplest applications of CW-estimators occurs in simple random sampling with replacement. Consider the following classical problem: m units are drawn with replacement with equal probabilities from a population of size N. The resulting sample is composed of the n_S distinct units. It is known (see Basu, 1958, Raj & Khamis, 1958, Pathak, 1961, Konijn, 1973, chap. IV, and Rao 1985) that

$$Pr(n_S = r) = \frac{1}{N^m} \binom{N}{r} \sum_{i=1}^{r} \binom{r}{i} i^m (-1)^{r-i}, r = 1, \dots, \min(N, m),$$

$$E(n_S) = N\left(1 - \frac{(N-1)^m}{N^m}\right) \tag{11}$$

and

$$Var(n_S) = \frac{(N-1)^m}{N^{m-1}} + (N-1)\frac{(N-2)^m}{N^{m-1}} - \frac{(N-1)^{2m}}{N^{2m-2}}. \tag{12}$$

Suppose now that n_S is used as the auxiliary statistic. Since, conditionally on n_S, the design is with equal probabilities and without replacement, we get $\pi_{k|n_S} = E(I_k \mid n_S) = n_S/N$ for all $k \in U$. The unconditional inclusion probabilities are $E(\pi_{k|n_S}) = N^{-1}E(n_S) = (1 - N^{-m}(N-1)^m)$. Moreover $Pr(\pi_{k|n_S} > 0) = 1$ and thus the four CW-estimators are all equal to the simple sample mean $\hat{\bar{y}}_{|\eta} = \bar{y}_S = n_S^{-1}\sum_{k\in S} y_k$ while the Horvitz–Thompson estimator is given by

$$\hat{\bar{y}}_\pi = \frac{n_S}{E(n_S)}\bar{y}_S = \left\{N\left(1 - \frac{(N-1)^m}{N^m}\right)\right\}^{-1}\sum_{k\in S} y_k.$$

Note that $\bar{y}_S$ is unbiased conditionally on n_S while $\hat{\bar{y}}_\pi$ is not. Moreover, $\bar{y}_S$ is invariant by translation while $\hat{\bar{y}}_\pi$ is not.

Konijn (1973, Chap IV), proved that the variance of the CW-estimator is given by

$$Var[\bar{y}_S] = E\left(\frac{\sigma_y^2}{n_S}\frac{N-n_S}{N-1}\right) = \frac{\sigma_y^2}{m}\frac{\sum_{j=1}^{N-1} j^{m-1}}{N^{m-1}} \tag{13}$$

where

$$\sigma_y^2 = \frac{1}{N}\sum_{k\in U}(y_k - \bar{y})^2.$$

The variance of the Horvitz-Thompson estimator is given by

$$\begin{aligned} Var[\hat{\bar{y}}_\pi] &= Var\left[\frac{n_S}{E(n_S)}\bar{y}_S\right] \\ &= EVar\left[\frac{n_S}{E(n_S)}\bar{y}_S \middle| n_S\right] + VarE\left[\frac{n_S}{E(n_S)}\bar{y}_S \middle| n_S\right] \\ &= \frac{\sigma_y^2}{N-1}\left(\frac{N}{E[n_S]} - \frac{E[n_S^2]}{E[n_S]^2}\right) + \frac{\bar{y}^2}{E[n_S]^2}Var[n_S] \end{aligned} \tag{14}$$

where n_S is defined as in (11). Examining (13) and (14), we see that the CW-estimator has a smaller variance if and only if

$$\bar{y}^2 > \frac{\sigma_y^2}{N-1}\left\{\frac{E[n_S]^2}{Var[n_S]}N\left(E\left[\frac{1}{n_S}\right] - \frac{1}{E[n_S]}\right) + 1\right\}.$$

This application is interesting. It shows that, even in case all the conditional inclusion probabilities are larger than zero, the CW-estimator has not necessarily a smaller variance than the Horvitz–Thompson estimator. Indeed, for this example, the Horvitz–Thompson estimator has a smaller variance if the population mean is close to zero. This simple result shows that it is impossible to determine the best estimation procedure without taking into account the relation between the auxiliary statistic and the interest variable.

This result could appear to be in contradiction to expression (3) that shows that if the conditional bias is corrected, the variance of the estimator will be smaller. Nevertheless, as we write in Section 1, the benefit gained in attempting to decrease the conditional bias could be thwarted by the instability of the conditional bias estimator used. For this reason, the CW-estimators are not better than the Horvitz–Thompson estimator for all possible values of $\bar{y}$.

3.2 *Poststratification*

Suppose that a simple random sample of size n is drawn without replacement from a population of size N. The population is assumed to be divided into H poststrata $U_h, h = 1, \dots, H$, of sizes $N_h, h = 1, \dots, H$. We denote $n_h, h = 1, \dots, H$, the sample poststratum sizes. Since the sampling design is simple without replacement, $\pi_k = n/N$ for all $k \in U$.

Suppose now that $\eta = (n_1 \ \dots \ n_h \ \dots \ n_H)$. Note that

$$Pr(n_h = r_h, h = 1, \dots, H) = \binom{N}{n}^{-1} \prod_{h=1}^{H} \binom{N_h}{r_h},$$

where $r_h = 0, \dots, N_h, \sum_{h=1}^{H} r_h = n$, and, if $k \in U_\alpha$,

$$\begin{aligned} & Pr(n_h = r_h, h = 1, \dots, H \mid k \in S) \\ = & \binom{N-1}{n-1}^{-1} \binom{N_\alpha - 1}{r_\alpha - 1} \prod_{\substack{h=1 \\ h \neq \alpha}}^{H} \binom{N_h}{r_h}. \end{aligned}$$

It follows that $\pi_{k|\eta} = n_\alpha / N_\alpha$ and

$$Pr(\pi_{k|\eta} > 0) = 1 - Pr[n_\alpha = 0] = 1 - \frac{(N - N_\alpha)^{[n]}}{N^{[n]}}$$

where $N^{[n]} = N(N-1)\dots(N-n+1)$.

The four CW-estimators become :

1. from (5), the SCW-estimator

$$\hat{\bar{y}}_{|\eta} = \frac{1}{N} \sum_{\substack{h=1 \\ n_h > 0}}^{H} N_h \hat{\bar{y}}_h, \tag{15}$$

2. from (7), the CCW-estimator

$$\hat{\bar{y}}_{c|\eta} = \frac{1}{N} \sum_{\substack{h=1 \\ n_h > 0}}^{H} \frac{N_h \hat{\bar{y}}_h N^{[n]}}{N^{[n]} - (N - N_h)^{[n]}}, \tag{16}$$

3. from (9), the SCW-ratio

$$\hat{\bar{y}}_{r|\eta} = \left(\sum_{\substack{h=1 \\ n_h > 0}}^{H} N_h \right)^{-1} \sum_{\substack{h=1 \\ n_h > 0}}^{H} N_h \hat{\bar{y}}_h, \tag{17}$$

4. from (10), the CCW-ratio

$$\begin{aligned} \hat{\bar{y}}_{cr|\eta} = & \left(\sum_{\substack{h=1 \\ n_h > 0}}^{H} \frac{N_h}{N^{[n]} - (N - N_h)^{[n]}} \right)^{-1} \\ & \times \sum_{\substack{h=1 \\ n_h > 0}}^{H} \frac{N_h \hat{\bar{y}}_h}{N^{[n]} - (N - N_h)^{[n]}}, \end{aligned} \tag{18}$$

where $\hat{\bar{y}}_h$ denotes the simple sample mean within the hth poststratum.

These four estimators are four versions of the poststratified estimators. Estimator (15) was studied by Holt & Smith (1979). Doss, Hartley & Somayajulu (1979) discuss estimator (16) which is exactly unconditionally unbiased but not VCU. In the same paper, these authors discuss CCW-ratio (18). An interesting discussion of these estimators is given by Rao (1985). In practice, we advocate the use of SCW-ratio because it is invariant by translation and, if $n_h > 0, h = 1, \ldots, H$, it is calibrated on the N_h i.e. if the SCW-ratio is used to estimate N_h, one gets exactly N_h.

3.3 2×2 *Contingency Tables*

A classical application of the use of auxiliary information occurs in modifying cell estimates in a contingency table with the aid of known marginal counts. The application of CW-estimator to this case is rather complex. For this reason, we only consider a 2×2 contingency table. Suppose that the population is divided into four subpopulations $U_{hq}, h = 1, 2, q = 1, 2$, of size $N_{hq}, h = 1, 2, q = 1, 2$. We also denote $N_{h.} = N_{h1} + N_{h2}, h = 1, 2$, and $N_{.q} = N_{1q} + N_{2q}, q = 1, 2$. In the population, only the marginal totals of the table are assumed to be known.

Suppose that a sample of size n is drawn with a simple random sampling design without replacement and that the number of units selected in each subpopulation is denoted $n_{hq}, h = 1, 2, q = 1, 2$, with $n_{h.} = n_{h1} + n_{h2}, h = 1, 2$, and $n_{.q} = n_{1q} + n_{2q}, q = 1, 2$. The Horvitz–Thompson estimator of N_{11} is $\widehat{N}_{11\pi} = Nn_{11}/n$. However, the objective is to construct an estimator of N_{11} that takes into account the auxiliary information given by the marginal totals of the population.

Generally this problem is viewed as a calibration problem. The sample contingency table is adjusted to the marginal totals of the population table. (See, e.g., Stephan, 1942, Frielander, 1961, Ireland & Kullback, 1968, Fienberg, 1970, Thionet, 1959 and 1976, Froment & Lenclud, 1976, Durieux & Payen, 1976). The estimator obtained by applying such an adjustment is usually called a raking ratio estimator.

The interest function N_{11} is a total as in (4) and can thus be estimated by a SCW-estimator. The following result gives a VCU-estimator of N_{11} with respect to the auxiliary statistic $(n_{1.}, n_{.1})$ by using a SCW-estimator.

RESULT 1: *The SCW-estimator of N_{11} with respect to $\eta = (n_{1.}, n_{.1})$ is given by*

$$\widehat{N}_{11|\eta} = \frac{n_{11}}{\pi_{11|\eta}(N_{11})}, \tag{19}$$

where

$$\pi_{11|\eta}(N_{11}) = \frac{\sum_{z=b^-}^{b^+} z\psi(N_{11}, z)}{N_{11}\sum_{z=b^-}^{b^+} \psi(N_{11}, z)},$$

$$\psi(N_{11}, z) = \frac{N_{11}^{[z]}(N_{1.} - N_{11})^{[n_{1.}-z]}(N_{.1} - N_{11})^{[n_{.1}-z]}(N - N_{.1} - N_{1.} + N_{11})^{[n-n_{1.}-n_{.1}+z]}}{z!(n_{1.} - z)!(n_{.1} - z)!(n - n_{1.} - n_{.1} + z)!}, \tag{20}$$

$$b^- = \max(0, n_{1.} + n_{.1} - n, N_{11} - N_{1.} + n_{.1}, N_{11} - N_{.1} + n_{1.}) \tag{21}$$

and

$$b^+ = \min(n_{1.}, n_{.1}, N_{11}, N - N_{1.} - N_{.1} + N_{11} + n_{1.} + n_{.1} - n). \tag{22}$$

The proof is given in Appendix 1.

Note that this estimator (19) depends on N_{11}, which we want to estimate. It is thus impossible to construct a CW-estimator without knowing N_{11}. As the problem cannot be solved when only the marginal totals are known, we must use an approximation. The estimator that best respects the idea of suppression of conditional bias is denoted $\widehat{\widehat{N}}_{11|\eta}$ and satisfies the relation

$$\widehat{\widehat{N}}_{11|\eta} = \frac{n_{11}}{\pi_{11|\eta}(\widehat{\widehat{N}}_{11|\eta})}.$$

Knowing that

$$\widehat{\widehat{N}}_{11|\eta} \in [\max(n_{11}, N_{1.} + N_{.1} - N + n - n_{1.} - n_{.1} + n_{11}),\\ \min(N_{1.} - n_{1.} + n_{11}, N_{.1} - n_{.1} + n_{11})]$$

and that generally the hypergeometric distribution can be defined for non-integer values of N_{hq}, the problem can be easily solved numerically.

EXAMPLE 3: Suppose that a simple random sample without replacement of size $n = 22$ is drawn from a population of size $N = 40$ and that we obtain the results in Table 3.

Table 3

Sample data

$n_{11} = 3$	$n_{12} = 10$	$n_{1.} = 13$
$n_{21} = 6$	$n_{22} = 3$	$n_{2.} = 9$
$n_{.1} = 9$	$n_{.2} = 13$	$n = 22$

Suppose also that the following auxiliary information is known: $N_{1.} = 20$, $N_{2.} = 20$, $N_{.1} = 11$, $N_{.2} = 29$. If $\widehat{N}_{hqRR}$ denotes the raking ratio estimator for N_{hq}, the result is given in Table 4.

Table 4

Adjusted data by the raking ratio-estimators

$\widehat{N}_{11RR} = 2.13$	$\widehat{N}_{12RR} = 17.87$	$N_{1.} = 20$
$\widehat{N}_{21RR} = 8.87$	$\widehat{N}_{22RR} = 11.13$	$N_{2.} = 20$
$N_{.1} = 11$	$N_{.2} = 29$	$N = 40$

The result is not very coherent. Indeed, $\widehat{N}_{11RR} = 2.13$ is smaller than $n_{11} = 3$ while N_{11} is necessarily larger or equal to n_{11}. The raking ratio estimator does not take into account that the population is finite. In cases with small sample sizes as in this example, this can lead to absurd results. The approximation of the SCW-estimator of the N_{hq} denoted $\widehat{\widehat{N}}_{hq|\eta}$ is given in Table 5.

Table 5

The CW-estimator

$\widehat{\widehat{N}}_{11|\eta} = 3.16$	$\widehat{\widehat{N}}_{12|\eta} = 16.84$	$N_{1.} = 20$
$\widehat{\widehat{N}}_{21|\eta} = 7.84$	$\widehat{\widehat{N}}_{22|\eta} = 12.16$	$N_{2.} = 20$
$N_{.1} = 11$	$N_{.2} = 29$	$N = 40$

The result is now very coherent. Indeed, $\widehat{N}_{|\eta 11} = 3.16$ is larger that n_{11}.

It thus appears that, in this example at least, the approximation of the SCW-estimator gives a more coherent result than the raking ratio estimator. As for the raking ratio estimator, it seems impossible to give an exact expression for the variance or MSE of the estimator. The following example is the comparison of the mean square error (MSE) for a simple example.

EXAMPLE 4: A simple random sample of size $n = 8$ was drawn without replacement from a population of size $N = 12$. We examined all the 2×2 population contingency tables with non-null cells. For each table, all the possible sample contingency tables with non-null cells were computed with their probabilities of being selected. Both estimators were computed: the SCW-estimator $\widehat{\widehat{N}}_{11|\eta}$ and the raking ratio estimator $\widehat{N}_{11RR}$. Next, the exact expectations and mean square errors of these estimators were computed in the following way

$$\begin{aligned}
E_1 &= E[\widehat{\widehat{N}}_{11|\eta} | n_{hq} > 0, h = 1, 2, q = 1, 2], \\
E_2 &= E[\widehat{N}_{11RR} | n_{hq} > 0, h = 1, 2, q = 1, 2], \\
MSE_1 &= E[(\widehat{\widehat{N}}_{11|\eta} - N_{11})^2 | n_{hq} > 0, h = 1, 2, q = 1, 2], \\
MSE_2 &= E[(\widehat{N}_{11RR} - N_{11})^2 | n_{hq} > 0, h = 1, 2, q = 1, 2].
\end{aligned}$$

The expectations and mean square errors were calculated conditional on $n_{hq} > 0$ because the raking ratio estimator does not exist when one of the $n_{hq} = 0$. The effect was defined as the ratio of the two MSE.

$$\text{Effect} = \frac{MSE_1}{MSE_2}.$$

Finally, we give the chi-square statistic for the population contingency table :

$$\chi^2 = \sum_{h=1}^{2} \sum_{q=1}^{2} \frac{\left(N_{hq} - \frac{N_{h.} N_{.q}}{N} \right)^2}{\frac{N_{h.} N_{.q}}{N}}.$$

Results are given in Table 6.

In one case only, the raking ratio estimator (see bold line in Table 6) gives a better result than SCW-estimator. A detailed examination of Table 6 shows that the larger the value of χ^2, the better the relative precision of the CW-estimator. For small populations and sample sizes, this example shows that it is generally possible to get an estimator which is much more precise than the raking ratio estimator.

4 Regression Estimation and the Use of Auxiliary Information on Means

4.1 Auxiliary Information on Means

Suppose that the known auxiliary information is the population mean

$$\bar{x} = \frac{1}{N} \sum_{k \in U} x_k.$$

Table 6

Expectation and MSE of CW (1) and raking-ratio (2) estimators

N_{11}	$N_{1.}$	$N_{.1}$	E_1	E_2	MSE_1	MSE_2	Effect	χ^2
1	2	2	1.00	0.81	0.00	0.04	0.00	1.92
1	2	3	1.00	0.95	0.00	0.03	0.00	0.80
1	3	3	1.17	1.15	0.09	0.10	0.89	0.15
2	3	3	1.71	1.58	0.14	0.19	0.74	3.70
1	2	4	1.00	1.00	0.00	0.05	0.00	0.30
1	3	4	1.22	1.22	0.13	0.17	0.76	0.00
2	3	4	1.77	1.73	0.12	0.14	0.85	2.00
1	4	4	1.28	1.31	0.19	0.26	0.74	0.19
2	**4**	**4**	**1.95**	**1.93**	**0.20**	**0.18**	**1.11**	**0.75**
3	4	4	2.62	2.48	0.18	0.28	0.64	4.69
1	2	5	1.00	1.01	0.00	0.05	0.00	0.07
1	3	5	1.22	1.23	0.13	0.18	0.70	0.11
2	3	5	1.78	1.78	0.12	0.16	0.76	1.03
1	4	5	1.29	1.32	0.19	0.28	0.69	0.69
2	4	5	2.00	2.00	0.23	0.24	0.95	0.17
3	4	5	2.69	2.64	0.18	0.24	0.75	2.74
1	5	5	1.31	1.34	0.19	0.30	0.64	1.66
2	5	5	2.06	2.07	0.29	0.33	0.88	0.01
3	5	5	2.89	2.84	0.26	0.27	0.96	1.19
4	5	5	3.60	3.46	0.18	0.31	0.57	5.18
1	2	6	1.00	1.00	0.00	0.05	0.00	0.00
1	3	6	1.22	1.22	0.12	0.18	0.71	0.44
2	3	6	1.78	1.78	0.12	0.18	0.71	0.44
1	4	6	1.30	1.32	0.19	0.27	0.70	1.50
2	4	6	2.00	2.00	0.23	0.26	0.89	0.00
3	4	6	2.70	2.68	0.19	0.27	0.70	1.50
1	5	6	1.32	1.39	0.18	0.27	0.67	3.09
2	5	6	2.07	2.10	0.29	0.34	0.87	0.34
3	5	6	2.93	2.90	0.29	0.34	0.87	0.34
4	5	6	3.68	3.61	0.18	0.27	0.67	3.09
1	6	6	1.40	1.55	0.17	0.31	0.55	5.33
2	6	6	2.12	2.19	0.27	0.31	0.88	1.33
3	6	6	3.00	3.00	0.35	0.40	0.87	0.00
4	6	6	3.88	3.81	0.27	0.31	0.88	1.33
5	6	6	4.60	4.45	0.17	0.31	0.55	5.33

Consider the π-estimator of $\bar{x}$ given by

$$\hat{\bar{x}} = \frac{1}{N}\sum_{k\in S}\frac{x_k}{\pi_k}.$$

The aim is to use $\hat{\bar{x}}$ as an auxiliary statistic to estimate $\bar{y}$. Note that the SCW-estimator is given by

$$\hat{\bar{y}}_{|\hat{\bar{x}}} = \frac{1}{N}\sum_{k\in S}\frac{y_k}{\pi_{k|\hat{\bar{x}}}}$$

where $\pi_{k|\hat{\bar{x}}} = E\left(I_k \mid \hat{\bar{x}}\right)$. In order to construct the SCW-estimator, of $\bar{y}$, we thus need an approximation of $\pi_{k|\hat{\bar{x}}}$. If the random vector $\hat{\bar{x}}$ takes the value z, we have by Bayes's theorem that

$$E\left(I_k \mid \hat{\bar{x}} = z\right) = Pr\left(k \in S \mid \hat{\bar{x}} = z\right) = \pi_k \frac{Pr\left(\hat{\bar{x}} = z \mid k \in S\right)}{Pr\left(\hat{\bar{x}} = z\right)}.$$

In order to compute the conditional inclusion probabilities, it is thus necessary to know the probability distribution of $\hat{\bar{x}}$ unconditionally and conditionally on the presence of each unit in the sample. Except for some particular cases, this probability distribution is very complex; for this reason we shall construct approximations of the conditional inclusion probabilities. Next, these approximations will be used to construct an approximation of the SCW-estimator.

It is possible to derive the means and variances of $\hat{\bar{x}}$ unconditionally and conditionally on the presence of each unit in the sample:

$$\bar{x} = E\left(\hat{\bar{x}}\right) = \frac{1}{N}\sum_{\ell\in U} x_\ell, \tag{23}$$

$$\bar{x}_{|k} = E\left(\hat{\bar{x}} \mid k \in S\right) = \frac{1}{N}\sum_{\substack{\ell\in U\\ \ell\neq k}}\frac{x_\ell \pi_{k\ell}}{\pi_k \pi_\ell} + \frac{x_k}{\pi_k N}, \tag{24}$$

$$V_x = Var\left(\hat{\bar{x}}\right) = \frac{1}{N^2}\sum_{\ell\in U}\frac{x_\ell x_\ell}{\pi_\ell}(1-\pi_\ell) + \frac{1}{N^2}\sum_{\ell\in U}\sum_{\substack{m\in U\\ m\neq \ell}}\frac{x_\ell x_m}{\pi_\ell \pi_m}(\pi_{\ell m} - \pi_\ell \pi_m) \tag{25}$$

and

$$\begin{aligned} V_{x|k} &= Var\left(\hat{\bar{x}} \mid k \in S\right) \\ &= \frac{1}{N^2}\sum_{\substack{\ell\in U\\ \ell\neq k}}\frac{x_\ell x_\ell \pi_{k\ell}}{\pi_k \pi_\ell^2}\left(1 - \frac{\pi_{k\ell}}{\pi_k}\right) \\ &\quad + \frac{1}{N^2}\sum_{\substack{\ell\in U\\ \ell\neq k}}\sum_{\substack{m\in U\\ m\neq \ell\\ m\neq k}}\frac{x_\ell x_m}{\pi_k \pi_\ell \pi_m}\left(\pi_{k\ell m} - \frac{\pi_{k\ell}\pi_{km}}{\pi_k}\right) \end{aligned} \tag{26}$$

where $\pi_{k\ell m}$ is the third-order inclusion probability. Note that V_x can also be written

$$V_x = \frac{1}{N}\sum_{k\in U}(\bar{x}_{|k} - \bar{x})x_k. \tag{27}$$

For simple random sampling without replacement, we get

$$\pi_k = \frac{n}{N}, \pi_{k\ell} = \frac{n}{N}\frac{n-1}{N-1} \text{ and } \pi_{k\ell m} = \frac{n}{N}\frac{n-1}{N-1}\frac{n-2}{N-2}.$$

By (24), (25), (26), we get

$$\bar{x}_{|k} = \bar{x} + \frac{N-n}{N-1}\frac{x_k - \bar{x}}{n}, \tag{28}$$

$$V_x = \frac{N-n}{N-1}\frac{\sigma_x^2}{n}, \tag{29}$$

$$V_{x|k} = \frac{N(N-n)(n-1)}{(N-2)(N-1)n^2}\left\{\sigma_x^2 - \frac{(x_k - \bar{x})^2}{N-1}\right\} \tag{30}$$

where

$$\sigma_x^2 = \frac{1}{N}\sum_{k\in U}(x_k - \bar{x})^2.$$

4.2 SCW-Estimation and Regression Estimation

For simple random sampling without replacement, the normality of the mean estimator was proved by Madow (1948) under some conditions and for a large sample size. If we suppose that $\hat{\bar{x}}$ has a normal distribution conditionally and unconditionally in the presence on $k \in S$, then

$$a_k(\hat{\bar{x}}) = \frac{n}{N\pi_{k|\hat{\bar{x}}}} = \frac{Pr(\hat{\bar{x}})}{Pr(\hat{\bar{x}} \mid k \in S)} = \frac{f(\hat{\bar{x}})}{f_k(\hat{\bar{x}})}$$

where $f(.)$ (resp. $f_k(.)$) is the density function of a normal variable with mean $\bar{x}$ (resp. $\bar{x}_{|k}$) and variance V_x (resp. $V_{x|k}$). Thus,

$$a_k(\hat{\bar{x}}) = \frac{V_x^{-1/2}\exp - \dfrac{(\hat{\bar{x}} - \bar{x})^2}{2V_x}}{V_{x|k}^{-1/2}\exp - \dfrac{(\hat{\bar{x}} - \bar{x}_{|k})^2}{2V_{x|k}}}. \tag{31}$$

The SCW-estimator is thus defined by

$$\hat{\bar{y}}_{|\eta} = \frac{1}{n}\sum_{k\in S} a_k(\hat{\bar{x}}) y_k.$$

The following result gives an approximation of the SCW-estimator:

RESULT 2: *In simple random sampling, if $\hat{\bar{x}}$ has a normal distribution unconditionally and conditionally on the presence of each unit in the sample, then an approximation of the SCW-estimator of $\bar{y}$ conditionally to $\hat{\bar{x}}$ is given by*

$$\hat{\bar{y}}_{|\hat{\bar{x}}} = \hat{\bar{y}}_\pi + (\bar{x} - \hat{\bar{x}})\frac{1}{n\sigma_x^2}\sum_{k\in S}(x_k - \bar{x})y_k + O_p(n^{-1}), \tag{32}$$

where $n \times O_p(n^{-1})$ is a quantity that remains bounded in probability when n tends to infinity.

Proof of Result 2 is given in Appendix 2.

The approximation given in expression (32) is very similar to the regression estimator given by

$$\hat{\bar{y}}_R = \hat{\bar{y}}_\pi + (\bar{x} - \hat{\bar{x}})\frac{1}{n\widehat{\sigma}_x^2}\sum_{k\in S}(x_k - \hat{\bar{x}})y_k. \tag{33}$$

The only difference between the regression estimator and the approximation of the SCW-estimator is the way we estimate σ_x^2 and

$$\sigma_{xy} = \frac{1}{N}\sum_{k\in U}(x_k - \bar{x})(y_k - \bar{y}).$$

Indeed, in (32), the covariance σ_{xy} is estimated by

$$\frac{1}{n}\sum_{k\in S}(x_k - \bar{x})y_k$$

while, in (33), it is estimated by

$$\frac{1}{n}\sum_{k\in S}(x_k - \hat{\bar{x}})y_k = \frac{1}{n}\sum_{k\in S}(x_k - \hat{\bar{x}})(y_k - \hat{\bar{y}}).$$

Moreover, estimator (32) uses the population variance σ_x^2 while estimator (33) uses an estimator $\widehat{\sigma}_x^2$ of σ_x^2. The SCW-estimator approximation thus needs more complete auxiliary information.

Result 2 is interesting because it shows that the regression estimator can be introduced without using a superpopulation model. Indeed, it is a natural approximation of the SCW-estimator for large sample size. Moreover, Result 2 also shows that, for simple random sampling, the regression estimator leads to valid conditional inference for large sample size. Indeed, from Result 2, it is obvious that it is asymptotically conditionally unbiased.

5 Discussion

As a conditionally unbiased estimator rarely exists, the conditional unbiasedness concept must thus be enlarged to be really operational. A general method to construct a conditionally VCU-estimator with respect to an auxiliary statistic is given. Nevertheless, this method brings up a new problem: how to choose the statistic η?

We advocate the choice of the estimator having the smallest unconditional MSE. Note that a CW-estimator does not necessarily have a smaller MSE than the Horvitz–Thompson estimator. If we consider the example developed in Section 3.1, we see that even if all the conditional inclusion probabilities are strictly positive, the MSE of the CW-estimator can be larger than the MSE of the Horvitz–Thompson estimator. In poststratification also, it is known that when the interest variable is independent of the poststratification variable, it is preferable to use the Horvitz–Thompson estimator.

The following result is useful in order to choose the auxiliary statistic. The conditional bias of the Horvitz–Thompson estimator is given by:

$$B\left(\hat{\bar{y}}_\pi \mid \eta\right) = \frac{1}{N}\sum_{k\in U} y_k \frac{\pi_{k|\eta} - \pi_k}{\pi_k}.$$

By expression (5), the SCW-estimator of this bias is given by

$$\widehat{B}_{|\eta}\left(\hat{\bar{y}}_\pi \mid \eta\right) = \frac{1}{N}\sum_{k\in S} y_k \left(\frac{1}{\pi_k} - \frac{1}{\pi_{k|\eta}}\right).$$

After correcting the Horvitz–Thompson estimator by means of this estimator of the conditional bias, we get the SCW-estimator. Indeed,

$$\begin{aligned}\hat{\hat{y}}_\pi - \widehat{B}_{|\eta}\left(\hat{\hat{y}}_\pi \mid \eta\right) &= \frac{1}{N}\sum_{k\in S}\frac{y_k}{\pi_k} - \frac{1}{N}\sum_{k\in S} y_k\left(\frac{1}{\pi_k} - \frac{1}{\pi_{k|\eta}}\right) \\ &= \frac{1}{N}\sum_{k\in S}\frac{y_k}{\pi_{k|\eta}}.\end{aligned}$$

Thus, the SCW-estimator can be presented as a Horvitz–Thompson estimator to which a conditional unbiasedness correction is applied.

Note that if the conditional bias was known, we should get the following result as seen in Section 1:

$$Var\left(E[\hat{\hat{y}}_\pi - B(\hat{\hat{y}}_\pi \mid \eta)]\right) = Var[\hat{\hat{y}}_\pi] - Var[E(\hat{\hat{y}}_\pi | \eta)]. \tag{34}$$

Expression (34) gives a first indication on the choice of the auxiliary statistic. The auxiliary statistic η must be chosen in such a way that $Var[E(\hat{\hat{y}}_\pi|\eta)]$ is as large as possible. The statistics η and $\hat{\hat{y}}_\pi$ must thus be very dependent. Nevertheless, another element must be taken into account. In order that the conditional bias should remain small, the conditional inclusion probabilities must be strictly positive. The choice of the auxiliary statistic must thus be made according to two quite contradictory criteria: η must be highly dependent on $\hat{\hat{y}}_\pi$ but most of the $\pi_{k|\eta}$ must stay strictly positive.

Finally, the conditional inclusion probabilities must be computable. Auxiliary information is thus necessary to calculate the probability distribution of the auxiliary statistic unconditionally and conditionally on the presence of each unit in the sample. Generally, the knowledge of the probability distribution of η can be derived from the sampling design $p(s)$ and from the vector $(x_1 \ldots x_k \ldots x_N)$. In some cases, as in poststratification, the conditional inclusion probabilities are computable without difficulty. In some other cases, as in the table adjustment problem, the calculus becomes more difficult. In more complex sampling designs, it becomes practically impossible to find the exact value of the conditional inclusion probability.

The complexity of this problem is due to the fact that three variables (or three groups of variables) interact: the interest variable; the auxiliary variables used *a priori* (as the stratification variables or the variables used to sample with unequal probabilities); the auxiliary variables used *a posteriori* (variables used at the estimation stage). The use of an auxiliary variable without any reference to a model in a complex sampling design, must thus take into account this third-order interaction. For this reason, the computation of the conditional inclusion probabilities becomes very complex. It is possible, however, to give approximations of these inclusion probabilities even for the complex cases. These approximations will be given in a further paper.

Acknowledgements

The author is grateful to Carl Särndal for helpful comments and suggestions on earlier versions of this manuscript. I also wish to thank an Associate Editor and a Referee, both of whom provided useful comments that have improved this article.

References

Basu, D. (1958). On sampling with and without replacement. *Sankhyā*, **20**, 287–294.

Cochran, W.G. (1977). *Sampling Techniques*, 3rd edition, New York: John Wiley.

Doss, D.C., Hartley, H.O. & Somayajulu, G.R. (1979). An exact small sample theory for post stratification. *Journal of Statistical Planning and Inference*, **3**, 235–248.

Durieux, B. & Payen, J.-F. (1976). Ajustement d'une matrice sur ses marges: la méthode ASAM. *Annales de l'INSEE*, **22–23**, 311–337.

Fienberg, S.E. (1970). An iterative procedure for estimation in contingency tables. *Annals of Mathematical Statistics*, **41**, 907–917.

Frielander, D. (1961). A technique for estimating a contingency table, given the marginal totals and some supplementary data. *Journal of the Royal Statistical Society*, **A124**, 412–420.
Froment, R. & Lenclud, B. (1976). Ajustement de tableaux statistiques. *Annales de l'INSEE*, **22–23**, 29–53.
Holt, D. & Smith, T.M.F. (1979). Post stratification. *Journal of the Royal Statistical Society*, **A142**, 33–46.
Horvitz, D.G. & Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, **47**, 663–685.
Isaki, C.T. & Fuller, W.A. (1982). Survey design under a regression population model. *Journal of the American Statistical Association*, **77**, 89–96.
Ireland, C.T. & Kullback, S. (1968). Contingency tables with given marginals. *Biometrika*, **55**, 179–188.
Konijn, H.S. (1973). *Statistical Theory of sample survey design and analysis*, Amsterdam: North-Holland.
Madow, W.G. (1948). On the limiting distributions of estimated based on samples from finite universes. *Annals of Mathematical Statistics*, **19**, 535–545.
Pathak, P.K. (1961). On simple random sampling with replacement. *Sankhyā*, **A24**, 415–420.
Raj, D. & Khamis S.D. (1958). Some remarks on sampling with replacement. *Annals of Mathematical Statistics*, **29**, 550–557.
Rao, J.N.K. (1985). Conditional inference in survey sampling. *Survey Methodology*, **11**, 15–31.
Rao, J.N.K. (1994). Estimating totals and distribution functions using auxiliary information at the estimation stage. *Journal of Official Statistics*, **10**, 153–165.
Robinson, J. (1987). Conditioning ratio estimates under simple random sampling. *Journal of the American Statistical Association*, **82**, 826–831.
Stephan, 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. (1959). L'ajustement des résultats des sondages sur ceux des dénombrements. *Revue de l'Institut International de Statistique*, **27**, 8–25.
Thionet, P. (1976). Construction et reconstruction de tableaux statistiques. *Annales de l'INSEE*, **22–23**, 5–27.

Appendix 1 : Proof of Result 1

Note that the use of the auxiliary statistic $(n_{1.}, n_{.1})$ leads to the same result as $(n_{1.}, n_{2.}, n_{.1}, n_{.2})$ because of the relations $n_{2.} = n - n_{1.}$ and $n_{.2} = n - n_{.2}$. If $k \in U_{11}$, we get

$$Pr(n_{hq} = r_{hq}, h = 1,2, q = 1,2) = \binom{N}{n}^{-1} \prod_{h=1}^{2}\prod_{q=1}^{2} \binom{N_{hq}}{r_{hq}},$$

$$\begin{aligned} & Pr(n_{hq} = r_{hq}, h = 1,2, q = 1,2 \mid k \in S) \\ = \; & \binom{N-1}{n-1}^{-1} \binom{N_{11}-1}{r_{11}-1}\binom{N_{12}}{r_{12}}\binom{N_{21}}{r_{21}}\binom{N_{22}}{r_{22}} \\ = \; & \frac{Nr_{11}}{nN_{11}} Pr(n_{hq} = r_{hq}, h = 1,2, q = 1,2), \end{aligned}$$

$$\begin{aligned} & Pr(n_{1.} = r_{1.}, n_{.1} = r_{.1}) \\ = \; & \sum_z Pr(n_{11} = z, n_{12} = r_{1.} - z, n_{21} = r_{.1} - z, n_{22} = n - r_{1.} - r_{.1} + z) \end{aligned}$$

where $\sum_z$ extends from $\max(0, r_{1.} + r_{.1} - n, N_{11} - N_{1.} + r_{.1}, N_{11} - N_{.1} + r_{1.})$ to $\min(r_{1.}, r_{.1}, N_{11}, N - N_{1.} - N_{.1} + N_{11} + r_{1.} + r_{.1} - n)$. Moreover,

$$\begin{aligned} & Pr(n_{1.} = r_{1.}, n_{.1} = r_{.1} \mid k \in S) \\ = \; & \sum_z Pr(n_{11} = z, n_{12} = r_{1.} - z, n_{21} = r_{.1} - z, n_{22} = n - r_{1.} - r_{.1} + z \mid k \in S) \\ = \; & \frac{N}{nN_{11}} \sum_z z Pr(n_{11} = z, n_{12} = r_{1.} - z, n_{21} = r_{.1} - z, n_{22} = n - r_{1.} - r_{.1} + z). \end{aligned}$$

Since

$$E(I_k \mid n_{1.} = r_{1.}, n_{.1} = r_{.1}) = \frac{n}{N} \frac{Pr(n_{1.} = r_{1.}, n_{.1} = r_{.1} \mid k \in S)}{Pr(n_{1.} = r_{1.}, n_{.1} = r_{.1})},$$

the conditional inclusion probabilities for the units that belong to U_{11} can be written

$$\pi_{11|\eta}(N_{11}) = \frac{\sum_{z=b^-}^{b^+} z\psi(N_{11}, z)}{N_{11} \sum_{z=b^-}^{b^+} \psi(N_{11}, z)}$$

where b^-, b^+, and $\psi(N_{11}, z)$ are given in (21), (22), and (20).

Appendix 2 : Proof of Result 2

The following lemmas will be used in the proof of Result 2.

LEMMA 1: *If* $\gamma_k = V_x^{-1/2}(\bar{x}_{|k} - \bar{x})$, *t he n* $\gamma_k = O(n^{-1/2})$, *whe re* $n^\alpha O\left(n^{-\alpha}\right)$, $\alpha < 0$, *is a quant ity t hat re mains bou nded whe n n* $\rightarrow 0$.

Proof

By (28) and (29), we get

$$\begin{aligned}\gamma_k &= \left(\frac{N-n}{N-1}\frac{\sigma_x^2}{n}\right)^{-1/2} \times \frac{N-n}{N-1}\frac{x_k - \bar{x}}{n} \\ &= \left[\frac{N-n}{n(N-1)}\right]^{1/2} \frac{x_k - \bar{x}}{\sigma_x} \\ &= O(n^{-1/2}).\end{aligned}$$

LEMMA 2:

$$\frac{V_{x|k}}{V_x} = 1 + O\left(\frac{1}{n}\right).$$

Proof

By (29) and (30), we get

$$\begin{aligned}\frac{V_{x|k}}{V_x} &= \frac{N(n-1)}{(N-2)n}\left\{1 - \frac{(x_k - \bar{x})^2}{(N-1)\sigma_x^2}\right\} \\ &= 1 - \frac{1}{n}\left\{\frac{N-2n}{(N-2)} + \frac{N(n-1)}{(N-2)}\frac{(x_k - \bar{x})^2}{(N-1)\sigma_x^2}\right\} \\ &= 1 + O\left(\frac{1}{n}\right). \qquad (35)\end{aligned}$$

LEMMA 3:

If $V_{x|k}V_x^{-1} - 1 = O(n^{-1})$, *t he n* $V_x V_{x|k}^{-1} = 1 + O(n^{-1})$.

The proof is straightforward.

Proof of Result 2

If we define $\hat{\hat{x}}_c = V_x^{-1/2}(\hat{\hat{x}} - \bar{x})$, $\gamma_k = V_x^{-1/2}(\bar{x}_{|k} - \bar{x})$, and

$$c_k = \left(\frac{V_{x|k}}{V_x}\right)^{1/2} \exp\frac{\hat{\hat{x}}_c^2}{2}\left(\frac{V_x}{V_{x|k}} - 1\right), \tag{36}$$

then (31) can be written

$$\begin{aligned} a_k(\hat{\hat{x}}) &= \left(\frac{V_{x|k}}{V_x}\right)^{1/2} \exp\left\{-\frac{\hat{\hat{x}}_c^2}{2} + \frac{(\hat{\hat{x}}_c - \gamma_k)^2 V_x}{2V_{x|k}}\right\} \\ &= c_k \exp \gamma_k \frac{V_x}{2V_{x|k}}(\gamma_k - 2\hat{\hat{x}}_c). \end{aligned} \tag{37}$$

By using a Taylor development for the vector γ_k of (37), we get

$$a_k(\hat{\hat{x}}) = c_k\left(1 - \gamma_k \frac{V_x}{V_{x|k}}\hat{\hat{x}}_c\right) + \mathcal{R}(\gamma_k^{(0)}). \tag{38}$$

where

$$\begin{aligned} \mathcal{R}(\gamma_k^{(0)}) &= c_k\left\{\exp \gamma_k^{(0)} \frac{V_x}{2V_{x|k}}(\gamma_k^{(0)} - 2\hat{\hat{x}}_c)\right\} \\ &\times \gamma_k^2 \frac{V_x}{V_{x|k}}\left\{\frac{V_x}{V_{x|k}}\left(\gamma_k^{(0)} - \hat{\hat{x}}_c\right)^2 - 1\right\}, \end{aligned}$$

where $\gamma_k^{(0)}$ is a vector whose elements are included between the correspondent elements of γ_k and 0. By Lemma 1, we directly get $\gamma_k^2 = O\left(n^{-1}\right)$ and thus $\mathcal{R}(\gamma_k^{(0)}) = O_p(n^{-1})$. On the other hand, we have by (36), Lemma 2 and Lemma 3 that

$$c_k = \left\{1 + O(n^{-1})\right\}^{1/2} \exp\left\{\frac{\hat{\hat{x}}_c^2}{2} \times O(n^{-1})\right\} = 1 + O_p(n^{-1}). \tag{39}$$

By (38) and (39), we get

$$\begin{aligned} a(\hat{\hat{x}}) &= \left(1 + O_p^{-1}\right)\left\{1 - \gamma_k\left(1 + O(n^{-1})\right)\hat{\hat{x}}_c + O_p(n^{-1})\right\} \\ &= \left\{1 - \gamma_k\hat{\hat{x}}_c + O_p(n^{-1})\right\} \\ &= \left\{1 + (\bar{x} - \hat{\hat{x}})V_x^{-1}(\bar{x}_{|k} - \bar{x}) + O_p(n^{-1})\right\}. \end{aligned}$$

Finally, we get

$$\begin{aligned} \hat{\hat{y}}_{|\hat{\hat{x}}} &= \frac{1}{n}\sum_{k\in S} a_k(\hat{\hat{x}})y_k \\ &= \hat{\hat{y}}_\pi + (\bar{x} - \hat{\hat{x}})V_x^{-1}\frac{1}{n}\sum_{k\in S}(\bar{x}_{|k} - \bar{x})y_k + O_p(n^{-1}) \\ &= \hat{\hat{y}}_\pi + (\bar{x} - \hat{\hat{x}})\frac{1}{n\sigma_x^2}\sum_{k\in S}(x_k - \bar{x})y_k + O_p(n^{-1}). \end{aligned}$$

Résumé

Dans les enquêtes par sondage, une information auxiliaire sur l'ensemble de la population est souvent disponible. Le but de cet article est de développer une méthode qui permet de prendre en compte cette information auxiliaire à l'étape de l'estimation au moyen d'un ajustement du biais conditionnel. L'idée de base est de tenter de construire un estimateur sans biais conditionnel. Quatre estimateurs ayant un faible biais conditionnel sont proposés. On montre ensuite que beaucoup d'estimateurs présentés dans la littérature dans le cas du plan simple sans remise peuvent être obtenus en utilisant ce principe d'estimation. Les problèmes du sondage aléatoire simple avec remise, de la poststratification, de l'ajustement d'un tableau de contingence de dimension 2×2 sont discutés dans le contexte de l'estimation conditionnelle. Finalement on montre que l'estimateur par la régression peut être obtenu en cherchant une approximation de ce principe conditionnel.