Published in Statistica Neerlandica 63, issue 3, 269-293, 2009
which should be used for any reference to this work

General framework for the rotation of units in repeated survey sampling

Desislava Nedyalkova*, Lionel Qualité† and Yves Tillé‡

Institute of statistics, University of Neuchâtel, Pierre à Mazel 7, 2000 Neuchâtel, Switzerland

Coordination of probabilistic samples is a challenging theoretical problem faced by statistical institutes. One of their aims is to obtain good estimates for each wave while spreading the response burden across the entire population. There is a collection of existing solutions that try to attend to these needs. These solutions, which were developed independently, are integrated in a general framework and their corresponding longitudinal designs are computed. The properties of these longitudinal designs are discussed. It is also noted that there is an antagonism between a good rotation and control over the cross-sectional sampling design. A compromise needs to be reached between the quality of the sample coordination, which appears to be optimal for a systematic longitudinal sampling design, and the freedom of choice of the cross-sectional design. In order to reach such a compromise, an algorithm that uses a new method of longitudinal sampling is proposed.

Keywords and Phrases: sampling algorithms, sample coordination, business surveys, systematic sampling.

1. Introduction

The negative coordination of samples is a challenging theoretical problem faced by statistical institutes. In business surveys, for example, several survey samplings are conducted each year on a relatively small population of large- or medium-sized companies. The paperwork burden asked of these companies can lead to reduced response rates and lesser quality. It is thus important for statistical institutes to have some control over the response burden of the units in the population while maintaining a probabilistic sampling system. In business surveys, the aims of such a system can be diametrically opposed: the institutes may want to ensure that a company will not be selected too often in order to limit its burden, or on the contrary, they may want to have a large overlap between the samples of two consecutive waves in order to have accurate estimations for the evolutions. These

*desislava.nedyalkova@unine.ch
†lionel.qualite@unine.ch
‡yves.tille@unine.ch

opposite properties are respectively called negative and positive coordination of samples.

There is a collection of existing solutions that try to attend to these needs. For instance, the use of random numbers for coordinating Poisson designs (Brewer, Early and Joyce, 1972), collocated sampling (Brewer, Early and Hanif, 1984) and the use of a measure of burden (De Ree, 1983; Van Huis, Koeÿers and De Ree, 1994a,b). These methods give partial but important solutions to real-life problems. However, one drawback of these methods is that they do not allow the important advances made in the domain of one-sample selection over the last decades to be integrated. For example, none of these allow using maximum fixed-size entropy sampling (see, e.g. Chen, Dempster and Liu, 1994) or balanced sampling (Deville and Tillé, 2004) as a cross-sectional sampling design.

Our aim is to provide the core of a general theory that includes the main existing sampling designs. However, in an attempt to keep this presentation simple, we will only describe negative coordination methods. The case of unit rotation (e.g. in partially renewed panels) is thus taken into account as it can be seen as a negative coordination problem. Indeed, sample rotation is usually achieved by splitting a sample into different parts and drawing for each new wave a non-overlapping sample that replaces one of these parts.

Another important issue where we made simplifications is the one of dynamic populations. In real-life problems, especially when we deal with business surveys, we need to allow for births and deaths of units in the population. This point is an important part of every rotation system and adds to its practical values. The algorithms present in this paper only require slight adaptations to work with a dynamic population.

The first part of our paper, in section 2, is devoted to a review of the useful concepts and notations for sampling on one and on several occasions. In section 3, we present some classical sampling designs in the context of longitudinal selection of a given unit over time. We also propose a new sampling algorithm that allows us to impose a minimum time between two selections of a unit.

After that, in section 4, we review the main existing methods for negative coordination of samples: the Brewer method that selects Poisson samples, the method of permutation of random numbers (see, e.g. Cotton and Hesse, 1992a,b), and the burden method. We show that it is possible to compute the cross-sectional and longitudinal sampling designs and in some cases even the joint sampling design. We also show that the longitudinal designs, which were never calculated before, are either systematic or Deville's (1998) systematic. These methods are not suitable if one wants to use a complex cross-sectional design (e.g. unequal inclusion probabilities and fixed size at the same time).

In section 5, we give a general method that makes it possible to use any cross-sectional design along with a systematic longitudinal design. However, the cross-sectional design should be applied, at each step, on the conditional selection probabilities. That will result in a progressive loss of control over the cross-sectional designs. This difficulty sheds light on the antagonism between the requirements for the cross-sectional design and those for the longitudinal design. It seems that the quality of the coordination is contradictory with the control of the cross-sectional sampling design. Finally, in section 6, we develop new sampling strategies that allow us to have a good coordination while leaving a relatively free choice of cross-sectional designs.

2 Basic concepts and notation

2.1 Sampling on one occasion

A finite population is a set of N units. Each unit can be identified by a label. Let

$$U=\{1,\dots,k,\dots,N\}$$

be the set of these labels. The size N of the population is not necessarily known. A sample without replacement is a subset of the population and in vector notation is presented as:

$$\mathbf{s}=(s_1,\dots,s_k,\dots,s_N)'\in\{0,1\}^N,$$

where

$$s_k=\begin{cases}1 & \text{if unit } k \text{ is in the sample}\\ 0 & \text{if unit } k \text{ is not in the sample,}\end{cases}$$

for all $k\in U$. The sample size is:

$$n(\mathbf{s})=\sum_{k\in U}s_k.$$

A sampling design $p(\mathbf{s})$ is a probability distribution on the samples of U. Let $\mathbf{S}$ be the random sample, that is, the random vector of $\mathbb{R}^N$, whose distribution is given by

$$\Pr(\mathbf{S}=\mathbf{s})=p(\mathbf{s}).$$

The first-order inclusion probability π_k is the probability of selecting unit k in the sample, and $\boldsymbol{\pi}=(\pi_k)_{1\le k\le N}$ is the inclusion probability vector. It can be derived from the sampling design as follows:

$$\boldsymbol{\pi}=\sum_{\mathbf{s}\subset U}\mathbf{s}p(\mathbf{s}).$$

When the design has a fixed sample size n, then

$$\sum_{k\in U}\pi_k=n.$$

The joint inclusion probability $\pi_{k\ell}$ is the probability of selecting units k and ℓ together in the sample, and $\pi_{kk}=\pi_k$. The matrix of joint inclusion probabilities is given by

$$\mathbf{\Pi} = \sum_{\mathbf{s}\subset U} \mathbf{s}\mathbf{s}' p(\mathbf{s}).$$

A support $\mathcal{Q}$ is a set of samples. The support $\mathcal{Q}$ of a sampling design $p(\cdot)$ is defined by:

$$p(\mathbf{s}) > 0, \text{ for all } \mathbf{s}\in\mathcal{Q},$$

and

$$\sum_{\mathbf{s}\in\mathcal{Q}} p(\mathbf{s}) = 1.$$

The full support $\mathcal{S}$ is the set of all the possible samples, that is, $\mathcal{S} = \{0,1\}^N$ and $\text{card}(\mathcal{S}) = 2^N$. The support corresponding to the samples of fixed sample size n is defined by $\mathcal{S}_n = \{\mathbf{s} \in \mathcal{S} \mid \sum_{k\in U} s_k = n\}$. Note that $\text{card}(\mathcal{S}_n) = \binom{N}{n}$. Some sampling designs have very small supports. They are called minimum support designs. We refer to Definition 1.

Definition 1. *A sampling design* $p_0(\cdot)$ *with inclusion probabilities* $(\pi_k)_{1\le k\le N}$ *is said to be defined on a minimum support* $\mathcal{Q}_0$ *if, for every* $\mathcal{Q}\subset\mathcal{Q}_0$ *with* $\mathcal{Q}\neq\mathcal{Q}_0$*, there is no design* $p(\cdot)$ *with support* $\mathcal{Q}$ *and with* $\sum_{\mathbf{s}\in\mathcal{Q}} s_k p(\mathbf{s}) = \pi_k, k = 1,\dots,N$.

Péa, Qualité and Tillé (2007) showed that the systematic design is a minimum support design. They also presented new methods to construct minimum support designs.

2.2 Sampling on several occasions

In coordination problems, we are interested in drawing samples from a population at times $t = 1,2,\dots,T$. At time t, a sample without replacement is a subset of the population.

Definition 2. *The cross-sectional sample is denoted by a vector*

$$\mathbf{s}^t = (s_1^t,\dots,s_k^t,\dots,s_N^t)' \in \{0,1\}^N,$$

for all $t\in\{1,2,\dots,T\}$*, and the longitudinal sample by a vector*

$$\mathbf{s}_k = (s_k^1,\dots,s_k^t,\dots,s_k^T)' \in \{0,1\}^T,$$

where

$$s_k^t = \begin{cases} 1 & \textit{if, at time } t\textit{, unit } k \textit{ is in the sample } \mathbf{s}^t \\ 0 & \textit{if, at time } t\textit{, unit } k \textit{ is not in the sample } \mathbf{s}^t, \end{cases}$$

for all $k\in U$.

Definition 3. *A sampling design* $p(\mathbf{s}^t), t = 1,2,\dots,T$, *will be called a cross-sectional sampling design.*

Definition 4. *A sampling design $p(\mathbf{s}_k), k=1,2,\ldots,N$, will be called a longitudinal sampling design.*

The joint (or complete) sampling design $p(\mathbf{s})$ is given by

$$p(\mathbf{s})=p(\mathbf{s}^1,\ldots,\mathbf{s}^t,\ldots,\mathbf{s}^T).$$

From this joint sampling design, we can derive the marginal cross-sectional design for a time t

$$p(\mathbf{s}^t)=\sum_{\mathbf{s}^1,\ldots,\mathbf{s}^{t-1},\mathbf{s}^{t+1},\ldots,\mathbf{s}^T} p(\mathbf{s}^1,\ldots,\mathbf{s}^t,\ldots,\mathbf{s}^T),$$

and the marginal longitudinal design for a unit k,

$$p(\mathbf{s}_k)=\sum_{\mathbf{s}_1,\ldots,\mathbf{s}_{k-1},\mathbf{s}_{k+1},\ldots,\mathbf{s}_N} p(\mathbf{s}_1,\ldots,\mathbf{s}_k,\ldots,\mathbf{s}_N).$$

Let S_k^t be the random variable that takes the value 1 if unit k is selected at time t and 0 otherwise. The first-order inclusion probabilities and the joint inclusion probabilities of the cross-sectional design at time t are given, respectively, by:

$$\pi_k^t=\mathrm{E}(S_k^t) \text{ and } \pi_{k\ell}^t=\mathrm{E}(S_k^t S_\ell^t),$$

where $\mathrm{E}(\cdot)$ is the expectation under the probability distribution $p(\cdot), k, \ell \in U, t=1,\ldots,T$. The longitudinal joint inclusion probabilities for times t and u are given by:

$$\pi_k^{tu}=\mathrm{E}(S_k^t S_k^u), \quad k\in U, t,u=1,\ldots,T.$$

Finally, we can define:

$$\pi_{k\ell}^{tu}=\mathrm{E}(S_k^t S_\ell^u), \quad k,\ell\in U, t,u=1,\ldots,T.$$

We have that $\pi_{k\ell}^{tu}=\pi_{\ell k}^{ut}$, where $k,\ell\in U$, but $\pi_{k\ell}^{tu}$ is not necessarily equal to $\pi_{\ell k}^{tu}$. These definitions can easily be adapted to a dynamic population denoted $U^t, t=1,\ldots,T$, which holds N^t units at time t. Naturally, if a unit k does not belong to U^t, then $s_k^t=0$ and the inclusion probabilities π_k^t and $\pi_{k\ell}^{tu}$ are also null.

In a repeated sampling design, the objectives of the cross-sectional and longitudinal designs are completely different. The cross-sectional design must be organized so as to obtain a complete coverage of the population and optimize the accuracy of the estimators. The aim of the longitudinal design could be to organize an equitable rotation of the units in the samples in order to fairly share the response burden of the companies. In some studies, the aim could be to have the best possible longitudinal estimations. Fixed size of the longitudinal design is not necessarily required, but it could be if one wants to inform the units that they will be surveyed a certain number of times over a given period. Longitudinal estimators do not necessarily need to be optimized.

Yet, up to now, no algorithm enables us to combine all these requirements. That is why relatively simple cross-sectional designs such as optimal stratified designs are generally used in repeated business surveys. It would be interesting to be able to have cross-sectional sampling designs with nice properties such as balanced sampling.

Another requirement is that we may want to be able to draw a sample at time t without knowing how many other samples $\mathbf{s}^{t+1},\ldots\mathbf{s}^{T}$ will have to be drawn, or the future inclusion probabilities $\pi_k^u, k \in U$, $u = t+1,\ldots,T$. In order to do so, we need to have an adequate longitudinal sampling algorithm. One should not confuse the definition of a sampling design with that of a sampling algorithm. The sampling design is given by the probability measure $p(\cdot)$, while a sampling algorithm is a procedure that allows us to select a random sample. Usually, there are several algorithms that allow us to implement the same sampling design. For instance, in TILLÉ (2006, pp. 47–50), four sampling algorithms for simple random sampling design without replacement are proposed.

A sequential algorithm is a method that is applied to a list of units (or, in this case, occurrences) denoted $1,\ldots,t,\ldots,T$, which are sorted according to a particular order. TILLÉ (2006) gives two definitions of a sequential algorithm.

DEFINITION 5. *A longitudinal sampling algorithm, for a unit k, is said to be weakly sequential if at step* $t = 1,\ldots,T$ *of the procedure, the decision concerning whether the unit k is in the sample* $\mathbf{s}^t$ *is definitively taken.*

DEFINITION 6. *A longitudinal sampling algorithm is said to be strictly sequential if it is weakly sequential and if the decision concerning the unit k at time t does not depend on the inclusion probabilities of the unit k at times* $t+1,\ldots,T$ *and on the number T of the sampling occasions.*

A strictly sequential procedure may be necessary for the longitudinal design when we are sampling over time. This is the case when the inclusion probabilities for the future occasions are not known (e.g. they are proportional to a variable that is not available in advance), or when the total number of occasions is not known. Moreover, a strictly sequential algorithm allows for an indefinite number of sampling occasions, and can be used with dynamic populations. Indeed, with such an algorithm, the death of a unit has no influence on its previous selections and the unit can just stay in the population with null inclusion probabilities from then on. A newborn unit can always be added to the population and receive null inclusion probabilities for the previous sampling occasions.

A general schema for constructing a sequential algorithm consists in computing the conditional selection probabilities of a unit as described in Algorithm 1. First, a uniform random number is generated for each unit of the population. A unit is selected if its random number is at most equal to its inclusion probability. Next, for each consecutive wave, a uniform random number is generated for each unit $k \in U$. Then, a conditional probability of selection is computed. A unit is selected if its random number is at most equal to its conditional selection probability.

Algorithm 1. General longitudinal sequential algorithm

```
Generate $u$, a uniformly distributed random number in $[0,1)$.
if $u<\pi_k^1$ then
    $s_k^1=1$
else
    $s_k^1=0$
end if
for $t=2,\dots,T$ do
    Generate $u$, a uniformly distributed random number in $[0,1)$.
    Compute $p=\Pr(S_k^t=1 \mid S_k^{t-1}=s_k^{t-1},\dots,S_k^1=s_k^1)$.
    if $u<p$ then
        $s_k^t=1$
    else
        $s_k^t=0$
    end if
end for
```

The computation of the conditional selection probabilities can be intricate. In most cases, these probabilities depend on the inclusion probabilities at times $t+1$, $t+2,\dots,T$, and thus, in those cases there is no strictly sequential algorithm to implement the design. When the sampling design is such that these conditional selection probabilities do not depend on the future, Algorithm 1 is strictly sequential. In section 3, several strictly sequential algorithms, that are particular cases of Algorithm 1, along with the new algorithm that we propose, are presented.

2.3 *Average time out of the sample*

The distribution of the time between two selections of a given unit is an important characteristic for the coordination problem. Let ψ_k^t be the random variable defined for $t=1,\dots,T$ by

$$\psi_k^t(\mathbf{s}_k)=\begin{cases}\min(T-t,\min\{r\geq 1 \mid s_k^{t+r}=1\}) & \text{if } k\in\mathbf{s}^t,\\ 0 & \text{if } k\notin\mathbf{s}^t,\end{cases}$$

and $\psi_k^0(\mathbf{s}_k)=\min(T,\min\{r\geq 1 \mid s_k^r=1\})$. Let $\phi_k^t, t=1,\dots,T$, be a random variable with the same distribution as $\psi_k^t(\mathbf{s}_k)$ conditionally to $s_k^t=1$, so that, if $1\leq t<T$:

$$\phi_k^t=\begin{cases}1 & \text{with probability } \Pr(S_k^{t+1}=1 \mid S_k^t=1) \text{ if } t+1<T,\\ 2 & \text{with probability } \Pr(S_k^{t+2}=1,S_k^{t+1}=0 \mid S_k^t=1) \text{ if } t+2<T,\\ 3 & \text{with probability } \Pr(S_k^{t+3}=1,S_k^{t+2}=0,S_k^{t+1}=0 \mid S_k^t=1)\\ & \quad \text{if } t+3<T,\\ \vdots & \\ T-t & \text{with probability } 1-\sum_{r=1}^{T-t-1}\Pr(S_k^{t+r}=1,S_k^{t+r-1}=0,\dots,\\ & \quad S_k^{t+1}=0 \mid S_k^t=1).\end{cases}$$

We have the relation:

$$\sum_{t=0}^{T}\psi_k^t(\mathbf{s}_k)=T=\mathrm{E}(\psi_k^0)+\sum_{t=1}^{T}\pi_k^t\,\mathrm{E}\left(\phi_k^t \mid s_k^t=1\right). \tag{1}$$

The quantity $\mathrm{E}(\phi_k^t \mid s_k^t = 1)$ can be seen approximately as the expected time out of the sample for a unit that has just been selected at time t.

In the subsequent sections, we will give the distribution of ϕ_k^t for several sampling designs. We will also show that the control of ϕ_k^t is the main issue in sampling coordination. We will consider particular sampling designs such as simple random sampling, Poisson sampling, systematic sampling, Deville's sampling and give sequential algorithms for these designs.

3 Classical sampling designs

In this section, we will present a short summary of some of the classical sampling designs in the context of longitudinal sampling of a unit k at times $t = 1, \dots, T$ with inclusion probabilities $\pi_k^1, \dots, \pi_k^T$. We will also give sequential or strictly sequential procedures to implement these designs.

3.1 Poisson sampling design

A longitudinal sampling design $p(\mathbf{s}_k)$ is said to be a Poisson sampling without replacement if it can be written as:

$$p(\mathbf{s}_k) = \prod_{t=1}^{T} (\pi_k^t)^{s_k^t} (1 - \pi_k^t)^{1 - s_k^t}.$$

The inclusion probabilities are equal to π_k^t and the joint inclusion probabilities are equal to $\pi_k^{tu} = \pi_k^t \pi_k^u$ when $t \neq u$.

The random variables $S_k^1, S_k^2, \dots, S_k^T$ are independent and thus the application of the general sequential Algorithm 1 to Poisson sampling gives the strictly sequential Algorithm 2.

Algorithm 2. Poisson strictly sequential

```
for $t = 1, \dots, T$ do
    Select unit $k$ at time $t$ with probability $\pi_k^t$.
end for
```

It is possible to compute the number of steps needed to select k again given that it has been selected at time t:

$$\phi_k^t = \begin{cases} 1 & \text{with probability } \pi_k^{t+1} \text{ if } t+1 < T, \\ 2 & \text{with probability } \pi_k^{t+2}(1 - \pi_k^{t+1}) \text{ if } t+2 < T, \\ 3 & \text{with probability } \pi_k^{t+3}(1 - \pi_k^{t+2})(1 - \pi_k^{t+1}) \text{ if } t+3 < T, \\ 4 & \text{with probability } \pi_k^{t+4}(1 - \pi_k^{t+3})(1 - \pi_k^{t+2})(1 - \pi_k^{t+1}) \text{ if } t+4 < T, \\ \vdots & \end{cases}$$

With Poisson sampling, the sample size $n(\mathbf{S}_k)$ is random and has a Poisson-binomial distribution (see, e.g. Hodges and LeCam, 1960). Its expected value and variance are, respectively, equal to:

$$\mathrm{E}[n(\mathbf{S}_k)] = \sum_{t=1}^{T} \pi_k^t \quad \text{and} \quad \mathrm{var}[n(\mathbf{S}_k)] = \sum_{t=1}^{T} \pi_k^t (1 - \pi_k^t).$$

When all the inclusion probabilities are equal to π_k, the Poisson sampling design is called a Bernoulli design. In this case, $n(\mathbf{S}_k)$ has a binomial distribution: $n(\mathbf{S}_k) \sim \mathcal{B}(T, \pi_k)$ and

$$\Pr(\phi_k^t = j) = (1 - \pi_k)^{j-1} \pi_k, \quad j = 1, \ldots, T - t - 1.$$

If T is not finite, ϕ_k^t has a geometric distribution. In this case,

$$\mathrm{E}(\phi_k^t) = \frac{1}{\pi_k} \quad \text{and} \quad \mathrm{var}(\phi_k^t) = \frac{1 - \pi_k}{(\pi_k)^2}.$$

3.2 *Simple random sampling*

A longitudinal sampling design $p(\mathbf{s}_k)$ is said to be a simple random sampling without replacement (SRSWOR; with fixed sample size n_k) if it can be written as:

$$p(\mathbf{s}_k) = \begin{cases} \binom{T}{n_k}^{-1} & \text{if } n(\mathbf{s}_k) = n_k, \\ 0 & \text{otherwise.} \end{cases}$$

The first-order inclusion probabilities are $\pi_k^t = n_k / T$, for all $t = 1, \ldots, T$, and the joint inclusion probabilities are $\pi_k^{tu} = n_k(n_k - 1)/[T(T-1)]$, if $t \neq u$.

This design can be implemented using several sampling algorithms. An application of the general sequential Algorithm 1 was proposed by Fan, Muller and Rezucha (1962) and is presented in Algorithm 3. First, a uniform random number u is generated. Then, we calculate the probability of selection p. If the random number is less than the selection probability, then a unit is selected. The algorithm ends when exactly n_k units are selected. This algorithm is sequential but not strictly sequential, as the inclusion probabilities depend on the number of sampling occasions.

Algorithm 3. SRSWOR sequential

```
Let $j = 0$.
Generate $u$, a uniformly distributed random number in $[0, 1)$.
for $t = 1, \ldots, T$ do
Calculate $p = \frac{n_k - j}{T - t + 1}$.
  if $u < p$ then
    Select unit $k$ in $\mathbf{s}^t$.
    $j = j + 1$
  end if
end for
```

It is possible to compute the number of steps needed to select unit k again, given that it has been selected at time t:

$$\Pr(\phi_k^t = j) = \frac{\binom{T-j-1}{n_k-2}}{\binom{T-1}{n_k-1}}, \quad j = 1, \ldots, \min(T - t - 1, T - n_k + 1).$$

For the first $n_k - 2$ sampling occasions, ϕ_k^t has a negative (or inverse) hypergeometric distribution (see, e.g. Johnson, Kotz and Kemp, 1992), and

$$\mathrm{E}(\phi_k^t) = \frac{T}{n_k} \text{ and } \operatorname{var}(\phi_k^t) = \frac{T(T-n_k)(n_k-1)}{(n_k+1)n_k^2}.$$

The moments of ϕ_k^t for larger values of t are not as easy to obtain, owing to the special treatment given to the last sampling occasion in the definition of ψ_k^t and ϕ_k^t.

3.3 Systematic sampling

Suppose that the longitudinal inclusion probabilities are such that $0 < \pi_k^t < 1$, $t = 1, \ldots, T$ with

$$\sum_{t=1}^{T} \pi_k^t = n_k.$$

Let V_k^t be the cumulated inclusion probabilities defined by:

$$V_k^t = \sum_{i=1}^{t} \pi_k^i, \text{ for all } t = 1, \ldots, T, \tag{2}$$

with $V_k^0 = 0$ and $V_k^T = n_k$. The usual selection procedure for systematic sampling is given in Algorithm 4. This algorithm is sequential but is not a direct application of the general sequential Algorithm 1 to systematic sampling. The procedure is as follows. A uniform random number $u \in [0, 1)$ is generated. For all $t = 1, \ldots, T$, unit k is selected in the sample $\mathbf{s}^t$ if there exists an integer j, $0 \leq j < n_k$, such that $u + j$ falls in the interval $[V_k^{t-1}, V_k^t)$.

Algorithm 4. Usual strictly sequential algorithm for systematic sampling

```
Generate $u$, a uniformly distributed random number in $[0, 1)$.
for $t = 1, \ldots, T$ do
  if there is an integer $j > 0$ such that $V_k^{t-1} \leq u + j - 1 < V_k^t$ then
    $s_k^t = 1$
  else
    $s_k^t = 0$
  end if
end for
```

The sampling design can be computed exactly:

$$p(\mathbf{s}_k) = \Lambda\left(\bigcap_{t \mid s_k^t = 1} A_k^t\right),$$

where $r_k^t = V_k^t \bmod 1$,

$$A_k^t = \begin{cases} [r_k^{t-1}, r_k^t) & \text{if } r_k^{t-1} < r_k^t \\ [r_k^{t-1}, 1) \cup [0, r_k^t) & \text{otherwise,} \end{cases}$$

and $\Lambda(\bigcap_{t \mid s_k^t = 1} A_k^t)$ is the sum of the length of the intervals in $\bigcap_{t \mid s_k^t = 1} A_k^t$.

We propose an alternative to Algorithm 4 that is a direct application of the general sequential Algorithm 1 to systematic sampling. Algorithm 5 is strictly sequential, and as such it is practical for longitudinal sampling. It gives a simple procedure to compute the conditional probabilities of selecting the unit k at time t given the past. This algorithm can easily be adapted to the case of an indefinite number of sampling occasions.

If the sampling design is systematic with equal inclusion probabilities $\pi_k^t = \pi_k$, if T is infinite, and if c is the smallest integer such that $c\pi_k > 1$, then

$$\phi_k^t = \begin{cases} c-1 & \text{with probability } (c\pi_k - 1)/\pi_k \\ c & \text{with probability } 1 - (c\pi_k - 1)/\pi_k. \end{cases}$$

$$\mathrm{E}(\phi_k^t) = \frac{1}{\pi_k} \quad \text{and} \quad \mathrm{var}(\phi_k^t) = \frac{(c\pi_k - 1)(1 + \pi_k - c\pi_k)}{\pi_k^2}.$$

If $1/\pi_k$ is an integer, then $\mathrm{var}(\phi_k^t) = 0$.

3.4 Deville's systematic sampling

Deville (1998) presented a variant of the systematic algorithm that gives a new sampling design with unequal probabilities (see also Tillé, 2006, p. 128). Deville's technique gives a fixed-size sampling design with a larger support than systematic sampling and is based on a relatively simple algorithm. While only one random number is used for systematic sampling and its position relative to the cumulative inclusion probabilities V_k^t defined by expression (2) determines the whole sample, Deville's sampling uses several random numbers. The position of a random number has repercussions on a limited number of selection variables s_k^t. As a consequence,

Algorithm 5. Sequential systematic algorithm

```
Define $\boldsymbol{\pi}_k = (\pi_k^1, \ldots, \pi_k^T)'$, a vector of inclusion probabilities in $[0, 1]^T$.
Define $\mathbf{s}_k = (s_k^1, \ldots, s_k^T)' = (0, \ldots, 0)'$, the empty sample.
Define $[a, b] = [0, 1]$.
Generate $u$, a uniformly distributed random number in $[0, 1)$.
for $j = 1, \ldots, T$ do
  Calculate $p = \frac{\max(\min(\pi_k^j, b) - \max(0, a), 0)}{b - a}$.
  if $p > 0$ then
    if $u < \Pr(S_k^j = 1 \mid S_k^{j-1} = s_k^{j-1}, \ldots, S_k^1 = s_k^1) = p$ then
      $s_k^j = 1$
      $b = \min(\pi_k^j, b)$
    else
      $a = \pi_k^j$
    end if
  end if
  $a = (a - \pi_k^j) \bmod 1$
  $b = (b - \pi_k^j) \bmod 1$
  if $b < a$ then $b = b + 1$ end if
end for
```

the conditional selection probabilities of a unit given its past are less constrained than in systematic sampling. A random number is selected uniformly in each interval between two integers. Its position relative to the V_k^t that are also in this interval determines the values of the corresponding s_k^t. A slight adaptation has to be made so as to take into account the t such that $[V_k^{t-1}, V_k^t)$ contains an integer. Deville's sampling can be implemented with Algorithm 6.

Deville's systematic sampling can also be implemented in the form of a strictly sequential algorithm. Algorithm 7 is a particular case of the general sequential Algorithm 1. At each step of the algorithm, a conditional selection probability p, is computed. This algorithm can easily be adapted to the case of an indefinite number of sampling occasions.

For this sampling design, in the particular case where T is infinite, the π_k^t are equal to π_k, and $K = 1/\pi_k$ is an integer, we can compute the distribution of the variables ϕ_k^t. Let $r = K - (t \bmod K)$, for all t. Then,

$$\phi_k^t = \begin{cases} r+1 & \text{with probability } 1/K \\ r+2 & \text{with probability } 1/K \\ \vdots & \\ r+K & \text{with probability } 1/K. \end{cases}$$

In this case,

$$\Pr(\phi_k^t = j) = \frac{1}{K}, \quad j = r+1, \dots, r+K,$$

$$\mathrm{E}(\phi_k^t) = r + \frac{K+1}{2} \text{ and } \mathrm{var}(\phi_k^t) = \frac{(K-1)(K+1)}{12}.$$

The variance of ϕ_k^t is larger than in the case of systematic sampling.

ALGORITHM 6. Deville's systematic sampling

```
Generate $u_1$, a realization of a uniform random variable in $[0, 1)$.
if $V_k^{t-1} \le u_1 < V_k^t$ then $s_k^t = 1$ end if
for $i = 2, \dots, n_k$ do
  if $\ell$ is such that $V_k^{\ell-1} \le i - 1 < V_k^\ell$ then
    if $s_k^\ell = 1$ then
      $f(x) = \begin{cases} \dfrac{1}{i - V_k^\ell} & \text{if } x \ge V_k^\ell - (i-1) \\ 0 & \text{if } x < V_k^\ell - (i-1) \end{cases}, \quad x \in [0, 1).$
    else
      $f(x) = \begin{cases} 1 - \dfrac{(i - 1 - V_k^{\ell-1})(V_k^\ell - i + 1)}{[1 - (i - 1 - V_k^{\ell-1})][1 - (V_k^\ell - i + 1)]} & \text{if } x \ge V_k^\ell - (i-1) \\ \dfrac{1}{1 - (i - 1 - V_k^{\ell-1})} & \text{if } x < V_k^\ell - (i-1). \end{cases}$
    end if
  end if
  Generate $u_i$, a random variable with density $f(x)$.
  if $V_k^{t-1} \le u_i + i - 1 < V_k^t$ then $s_k^t = 1$ end if
end for
```

Systematic sampling and Deville's systematic sampling share the property that if

$$j \leq V_k^t = \sum_{i=1}^{t} \pi_k^i,$$

for a given integer j, then

$$j \leq \sum_{i=1}^{t} s_k^i.$$

This property enables us to implement a design sequentially with a controlled size over an indefinite period of time.

3.5 A new sampling algorithm for unequal probability sampling

In this algorithm, we define a fixed number of steps during which a unit, once selected, is not selected anymore. At each step, a conditional probability is calculated. However, this is only possible if the sum of the inclusion probabilities over this number of consecutive steps does not exceed 1. If r is the number of steps such that $s_k^t = 1$ implies $s_k^{t+1} = 0, \ldots, s_k^{t+r} = 0$, and if $\sum_{j=t}^{t+r-1} \pi_k^j < 1$, for $t = 1, \ldots, T - r + 1$, we consider Algorithm 8.

This algorithm is strictly sequential and can easily be adapted to an indefinite number of sampling occasions. In the Appendix, we give a modified version of this algorithm in which there is no condition on the inclusion probabilities. In that case, the fixed minimum time out of sample cannot always be respected.

When T is infinite, if all the π_k^t are equal, that is, $\pi_k^t = \pi_k$, and if $r\pi_k < 1$, then the distribution of the ϕ_k^t is as follows:

$$\Pr(\phi_k^t = j) = \begin{cases} 0, & j = 1, \ldots, r, \\ (1 - v_k)^{j-r-1} v_k, & j = r+1, r+2, r+3, \ldots, \end{cases}$$

where

$$v_k = \frac{\pi_k}{1 - r\pi_k}.$$

The variable ϕ_k^t has a shifted geometric distribution. We have

$$\mathrm{E}(\phi_k^t) = r + \frac{1}{v_k} = \frac{1}{\pi_k} \quad \text{and} \quad \mathrm{var}(\phi_k^t) = \frac{[(r+1)\pi_k - 1][(r+1)\pi_k - \pi_k - 1]}{(\pi_k)^2}.$$

The minimum time out of sample design can be viewed as a compromise between Poisson sampling and systematic sampling. On the one hand, if all the π_k^t are equal to π_k and $\pi_k = (r+1)^{-1}$, then $v_k = 1$ and the sampling design is systematic. On the other hand, if $r = 0$, then we obtain a Poisson design. The Poisson design maximizes the entropy while the systematic design has a very small entropy because it is a minimum support design. Between these two extreme situations, the minimum time out of sample design provides a large range of intermediate solutions.

Algorithm 7. Deville's systematic sequential

```
Define $\boldsymbol{\pi}_k=(\pi_k^1,\dots,\pi_k^T)'$, a vector of inclusion probabilities of length $T$.
Define $\mathbf{s}_k=(s_k^1,\dots,s_k^T)'=(0,\dots,0)'$, the empty sample.
Define $v=0$, the cumulated inclusion probability.
Define $f=0$.
Generate $u$, a uniformly distributed random number in $[0,1)$.
for $j=1,\dots,T$ do
  $p=0$
  if $v+\pi_k^j<1$ then
    if $f=0$ then $p=\pi_k^j/(1-v)$ end if
  else
    if $f=0$ then
      $p=1$
    else
      $p=(v+\pi_k^j-1)/v$
    end if
  end if
  $v=v+\pi_k^j$
  if $u<\Pr(S_k^j=1\mid S_k^{j-1}=s_k^{j-1},\dots,S_k^1=s_k^1)=p$ then $s_k^j=1$ end if
  if $v>1$ then
    $v=v-1$
    if $s_k^j=0$ then $f=0$ end if
  else
    if $s_k^j=1$ then $f=1$ end if
  end if
end for
```

Algorithm 8. Minimum time out of sample

```
Define $\boldsymbol{\pi}_k=(\pi_k^1,\dots,\pi_k^T)'$, a vector of inclusion probabilities of length $T$.
Define $\mathbf{s}_k=(s_k^1,\dots,s_k^T)'=(0,\dots,0)'$, the empty sample.
Generate $u^1$, a uniformly distributed random number in $[0,1)$.
if $u^1\le\pi_k^1$ then $s_k^1=1$ end if
for $t=2,\dots,r$ do
  Generate $u^t$, a uniformly distributed random number in $[0,1)$.
  if $s_k^1=0,\dots,s_k^{t-1}=0$ and $u^t\le p=\pi_k^t/(1-\sum_{i=1}^{t-1}\pi_k^i)$ then $s_k^t=1$ end if
end for
for $t=r+1,\dots,T$ do
  Generate $u^t$, a uniformly distributed random number in $[0,1)$.
  if $s_k^{t-r}=0,\dots,s_k^{t-1}=0$ and $u^t\le p=\pi_k^t/(1-\sum_{i=t-r}^{t-1}\pi_k^i)$ then $s_k^t=1$ end if
end for
```

3.6 Remark on the variables ϕ_k^t

For the sampling designs we just viewed, the expectation of the variables ϕ_k^t do not vary much. Expression (1), which is valid for any T, implies that in most cases this expectation will be close to $1/\pi_k^t$. For instance, if all the inclusion probabilities are equal to π_k, and T is infinite, then systematic sampling, minimum time out of sample and Poisson sampling all have the same expectation for ϕ_k^t. The variance of ϕ_k^t, however, varies greatly from one sampling design to the other. When $\pi_k^t=\pi_k$ is constant and $1/\pi_k$ is an integer, the only sampling design that gives a null variance for ϕ_k^t is the systematic sampling design.

The variable ϕ_k^t counts the number of waves that a unit stays out of the sample after having been selected. The expectation of this variable does not depend much on the sampling design. Hence, a good method for negative coordination cannot be a method that maximizes the number of times out of the sample after the selection of a unit. Instead, we can look for a method that organizes the rotation in a regular way, that is, which minimizes the variance of the ϕ_k^t. In this respect, systematic sampling is an interesting longitudinal design as it can give, in a very special case, perfect control over the frequency with which a unit is sampled.

In the next section, we examine the most usual coordination methods and show that their longitudinal designs match the ones we have just described. In the simplest cases, we compute the cross-sectional, longitudinal and joint designs exactly.

4 Usual methods of coordination

There are several simple algorithms that allow us to draw coordinated samples with simple random or Poisson cross-sectional sampling designs. In this section, we describe three well-known methods. We give the corresponding longitudinal designs and compute the joint sampling designs resulting from these algorithms.

4.1 The systematic-Poisson (or Brewer) repeated design

Brewer *et al.* (1972) suggested a very simple procedure to draw negatively coordinated Poisson samples. It gives a very convenient method to negatively coordinate samples with unequal probabilities. However, the cross-sectional samples do not have a fixed size. First, a uniform random number is generated for each unit of the population. A unit is selected if its random number is at most equal to its inclusion probability. Next, for each consecutive wave, we calculate a new uniform random number for all $k \in U$, which depends on the random number and on the inclusion probability at the previous wave. A unit is then selected if its new random number is at most equal to its new inclusion probability. The selection procedure is given in Algorithm 9.

Algorithm 9. Coordination of Poisson samples in the case of a static population

```
At time 1, assign a uniform random number $u_k^1$ to each unit $k \in U^1$.
if $u_k^1 \leq \pi_k^1$ then $s_k^1 = 1$ end if
for $t = 2, \ldots, T$ do
    Compute $u_k^t = (u_k^{t-1} - \pi_k^{t-1}) \bmod 1$.
    if $u_k^t \leq \pi_k^t$ then $s_k^t = 1$ end if
end for
```

The cross-sectional design given by this algorithm is a Poisson design:

$$p(\mathbf{s}^t) = \prod_{k \in U} \{(\pi_k^t)^{s_k^t}(1 - \pi_k^t)^{1-s_k^t}\}.$$

The longitudinal design is a systematic sampling design with unequal probabilities:

$$p(\mathbf{s}_k) = \Lambda\left(\bigcap_{t \mid s_k^t = 1} A_k^t\right),$$

with the notations of section 3.3. The selection of different units of the population being totally independent, the complete design is given by:

$$p(\mathbf{s}) = \prod_{k \in U} p(\mathbf{s}_k) = \prod_{k \in U} \Lambda\left(\bigcap_{t \mid s_k^t = 1} A_k^t\right).$$

From the complete design, we can derive all the properties of the sampling design. For example, we have

$$\pi_{k\ell}^t = \pi_k^t \pi_\ell^t, \text{ with } k \neq \ell, \text{ for all } t,$$

and

$$\pi_k^{t,t+j} = \Lambda(A_k^t \cap A_k^{t+j})$$

This method can easily be adapted for dynamic populations. If a newborn unit enters the population at a given time $t > 1$, then it receives an inclusion probability π_k^t and a uniform random number u_k^t. If its random number is not greater than its inclusion probability, then it is selected in the sample $\mathbf{s}^t$. At the following waves, its random number is subject to the same transformations as those of the other units of the population, described in line 4 of the algorithm. If a unit leaves the population at time t, then its inclusion probability becomes equal to zero for times $t, t+1, \ldots, T$. The adjustment of Algorithm 9 to dynamic populations is straightforward. We just need to replace U by U^t and add the following line between lines 4 and 5 of the algorithm:

```
4b: Add newborn units to the sampling frame with their $u_k^t$ and $\pi_k^t$.
```

As stated in the preceding section, a longitudinal systematic design can be desired in order to control the rotation of units in the sample. While Brewer's repeated design enjoys this property, it has a drawback: the cross-sectional design does not have a fixed sample size.

4.2 The systematic-simple repeated design

Suppose that the inclusion probabilities of the units in the population are constant at each wave, that is, $\pi_k^t = \pi^t$, and that $\sum_{k \in U} \pi_k^t = n^t$ is an integer. The following well-known procedure (see, e.g. Cotton and Hesse, 1992a,b), given in Algorithm 10, can be used to negatively coordinate simple random samples without replacement. Its main drawback is that it can only be used in the case of simple random sampling or stratified sampling with fixed strata.

First, a uniform random number is generated for each unit of the population. In order to obtain the sample $\mathbf{s}^1$, the n^1 units having the smallest random numbers are selected. At the following waves, permute the uniform random numbers so that the selected units at the previous wave receive the largest random numbers and the non-selected receive the smallest. Select the n^t units having the smallest random numbers to obtain the sample $\mathbf{s}^t$.

Algorithm 10. Coordination of SRSWOR using random numbers in the case of a static population.

```
1: At time 1, assign a uniform random number, $u_k^1$, to each unit $k \in U$, that is, construct the vector $\mathbf{u}^1 = (u_1^1, \dots, u_N^1)$.
2: Select the units that have the $n^1$ smallest $u_k^1$ to obtain the sample $\mathbf{s}^1$.
3: for $t = 2, \dots, T$ do
4:    Construct $u^t$ as a permutation of $\mathbf{u}^{t-1}$ so that the selected units at wave $t-1$ receive the largest $u_k^{t-1}$, the non-selected units receive the smallest $u_k^{t-1}$ and the ranks of the permuted random numbers remain unchanged within the subsets of the selected and non-selected units.
5:    Select the units that have the $n^t$ smallest $u_k^t$ to obtain the sample $\mathbf{s}^t$.
6: end for
```

This sampling algorithm results in a systematic longitudinal design. All the cross-sectional designs are simple and without replacement:

$$p(\mathbf{s}^t) = \begin{cases} \binom{N^t}{n^t}^{-1} & \text{if } n(\mathbf{s}^t) = n^t, \\ 0 & \text{otherwise.} \end{cases}$$

Moreover, for a static population, if

$$n(\mathbf{s}^1) = n^1, \dots, n(\mathbf{s}^t) = n^t \text{ and } \sum_{j=1}^{t} n^j \leq N,$$

then

$$p(\mathbf{s}) = \left\{ \frac{N!}{n^1! n^2! \dots n^t! (N - n^1 - \dots - n^t)!} \right\}^{-1}.$$

This method can easily be adapted for dynamic populations. If a newborn unit k enters the population at a given time $t > 1$, then a uniform random number u_k^t is generated for this unit. The vector $\mathbf{u}^t$ is permuted as before and the new random number is added to this vector. Again we select the n^t units having the smallest random numbers among the living N^t units of the population at time t. If a unit leaves the population at time $t > 1$, then we simply remove its random number from the vector $\mathbf{u}^t$. The adjustment of Algorithm 10 is straightforward. We just need to replace U by U^t and add the following line between lines 4 and 5 of the algorithm:

```
4b: Add newborn units' random numbers $u_k^t$ to the vector $\mathbf{u}^t$ at their relative positions.
```

4.3 Use of a measure of burden or the Deville's systematic-simple repeated design

Another method of coordination with simple random cross-sectional designs, based on the use of permanent random numbers for the selection of the sample, is used by *Statistics Netherlands* for their business surveys (see, e.g. De Ree, 1983; Koeÿers and Willeboordse, 1995; Van Huis *et al.*, 1994a,b). This method, called EDS, gives stratified cross-sectional samples with fixed size. The cumulative response burden of the units is factored in the selection process, and the surveys can have unequal response burdens. However, the choice of the stratification is not completely free. Indeed, the strata are constituted of response burden control groups, which are basic blocks of units that need to be defined once and for all. Rivière (2001a) proposed another method that takes into account the response burden, and that does not require predefined strata. However, it uses the intersection of all the previous stratifications. Consequently, it is not practical for a large number of sampling occasions. These methods, along with the Cotton and Hesse (1992a,b) method, can be used with dynamic populations.

We describe another simple method of coordination that uses a measure of burden but does not use permanent random numbers. At each wave, every unit receives a measure of burden equal to the number of times it has previously been selected. At time t, the sample of size n^t is selected among the units with the lowest burden measure. More precisely, the sample $\mathbf{s}^t$ at time t is drawn with a simple random sampling of size n^t in the set of units with the smallest burden, if this set is large enough. Or, if this set is too small, it is entirely selected and a sample is drawn with a simple random design in its supplement, in order to complete $\mathbf{s}^t$.

With this method, the burden measure, at any time t, can only take two values. It splits the population into a set of units with the lowest burden measure, denoted by $\mathcal{M}^{t-1}$ and a set of units with the largest burden measure, denoted by $U \backslash \mathcal{M}^{t-1}$. The procedure is given in Algorithm 11.

The cross-sectional design resulting from this algorithm is a SRSWOR, and the longitudinal design is the Deville's systematic sampling design, presented in Algorithm 7. Indeed, if the cumulated sum (over t), V_k^ℓ, of the $\pi_k^t = n^t/N$ is such that there is an integer $i-1 \geq 1$ between $V_k^{\ell-1}$ and V_k^ℓ, then the population at time $\ell-1$ is divided between units that have been selected $i-2$ times and units that have been selected $i-1$ times. Depending on its burden at time $\ell-1$, unit k is automatically

Algorithm 11. Coordination of SRSWOR using a measure of burden in the case of a static population

```
At time 1, assign a burden equal to 0 to each unit $k \in U$, that is, $b_k^1 = 0$.
Select a SRSWOR of size $n^1$.
**if** $s_k^1 = 1$ **then** $b_k^1 = b_k^1 + 1$ **end if**
**for** $t = 2, \ldots, T$ **do**
  Define $\mathcal{M}^{t-1}$, the set of units with the smallest burden.
  **if** card$(\mathcal{M}^{t-1}) > n^t$ **then**
    Select a SRSWOR of size $n^t$ from $\mathcal{M}^{t-1}$.
  **else**
    Select all the units from $\mathcal{M}^{t-1}$.
    Complete the sample by a SRSWOR with the units from $U \backslash \mathcal{M}^{t-1}$.
  **end if**
  **if** $s_k^t = 1$ **then** $b_k^t = b_k^{t-1} + 1$ **end if**
**end for**
```

selected in $\mathbf{s}^\ell$ (case $s_k^\ell = 1$ at line 5 in Algorithm 6) or it is drawn with equal probability among the units that have a burden equal to $i-1$. At the following occasions, its conditional selection probability satisfies the equations in lines 6 and 8 of Algorithm 6.

This method can easily be adapted for dynamic populations. Following the idea of De Ree (1999), we randomly assign to the newborn units a measure of burden. For example, at the beginning of wave $t>1$, there are $\text{card}(\mathcal{M}^{t-1})$ units with the smallest burden, denoted by b, and $N^{t-1} - \text{card}(\mathcal{M}^{t-1})$ units with the largest burden, equal to $b+1$. A newborn unit will be inserted into $\mathcal{M}^{t-1}$ and receive a burden equal to b with probability $\text{card}(\mathcal{M}^{t-1})/N^{t-1}$, and, with probability $1-\text{card}(\mathcal{M}^{t-1})/N^{t-1}$, it will receive a burden equal to $b+1$. The adjustment of Algorithm 11 is straightforward. We just need to replace U by U^t, define $\mathcal{M}^{t-1}$ as the set of living units with the smallest burden, and add the following line between lines 5 and 6 of the algorithm:

```
5b: Add each newborn unit to $\mathcal{M}^{t-1}$ with probability $\text{card}(\mathcal{M}^{t-1})/N^{t-1}$ and to $U^t \setminus \mathcal{M}^{t-1}$ with probability $1-\text{card}(\mathcal{M}^{t-1})/N^{t-1}$.
```

The coordination is not as good as with longitudinal systematic sampling as the measure of burden does not reflect the time spent out of the sample. For example, suppose that $N=4$, all the inclusion probabilities are $\pi_k^t = 1/4$ and $t=1,\ldots,T$. After four waves, all the units are selected once and they have a burden $b_k^4 = 1$. Hence, at the fifth wave, any unit can be selected again with probability 1/4. The same unit can thus be consecutively selected at $t=4$ and 5. Contrariwise, in this case, the systematic-simple design provides a strict rotation, in such a way that once a unit is selected it remains out of the sample during three waves. The EDS method by Van Huis *et al.* (1994a,b) is not affected by this problem.

5 Other repeated sampling designs

5.1 General method for the coordination of samples

The usual algorithms described in the previous section result either in a systematic or in a Deville's systematic longitudinal design. While systematic sampling seems to be a good choice for the longitudinal design, these algorithms do not allow for a wide selection of cross-sectional designs. We can wonder if there is a general way of obtaining a repeated sampling design with a given sequential longitudinal design and a given cross-sectional design. A weaker solution is possible if we are prepared to have imperfect control over the cross-sectional sampling designs.

- For each unit k of the population at time t, compute the conditional inclusion probabilities

$$\pi_k^t(s_k^{t-1},\ldots,s_k^1) = \Pr(S_k^t = 1 \mid S_k^{t-1} = s_k^{t-1},\ldots,S_k^1 = s_k^1),$$

according to the chosen strictly sequential algorithm (systematic, Deville's systematic, minimum time out of sample).

- When all the conditional probabilities are computed, apply to them any cross-sectional design in order to select $\mathbf{s}^t$. This design can be stratified, with unequal inclusion probabilities and fixed sample size, or even balanced (see, e.g. Deville and Tillé 2004).

With this method, conditionally to the past, we can choose any cross-sectional design provided that it is compatible with the conditional inclusion probabilities. This may in itself be a limiting factor, especially in the case of a systematic longitudinal design where these inclusion probabilities can rapidly become close or equal to 0 or 1. Moreover, the choice of the conditional cross-sectional design at time t is not the same as the choice of the marginal (unconditional) design for $\mathbf{s}^t$. This method is perfectly applicable to dynamic populations.

5.2 Application to a systematic longitudinal design

We have seen that the systematic longitudinal design is well suited for sampling coordination. It is thus of interest to know which cross-sectional designs can be implemented with a longitudinal systematic sampling design. We have seen in the previous section that it is the case of the Poisson and simple random sampling designs.

The sampling design at the first wave can always be chosen at will. If we apply the sequential algorithm presented in Algorithm 5, then we can compute, at wave 2, the conditional inclusion probabilities for each unit as follows:

$$\pi_k^2(s_k^1) = \Pr(S_k^2 = 1 \mid S_k^1 = s_k^1),$$

which, given that the longitudinal design is systematic, are such that

$$\pi_k^2(s_k^1) = \begin{cases} \pi_k^2/(1-\pi_k^1) & \text{if } s_k^1 = 0 \text{ and } \pi_k^1 + \pi_k^2 \leq 1 \\ 0 & \text{if } s_k^1 = 1 \text{ and } \pi_k^1 + \pi_k^2 \leq 1 \\ 1 & \text{if } s_k^1 = 0 \text{ and } \pi_k^1 + \pi_k^2 > 1 \\ (\pi_k^1 + \pi_k^2 - 1)/\pi_k^1 & \text{if } s_k^1 = 1 \text{ and } \pi_k^1 + \pi_k^2 > 1. \end{cases}$$

Then, any sampling design can be applied with the conditional inclusion probabilities $\pi_k^2(s_k^1)$. It must be noted that this free choice of conditional sampling design $p(\mathbf{s}^2 \mid \mathbf{s}^1)$ does not mean that we know how to obtain a given marginal sampling design $p(\mathbf{s}^2)$ for the second wave. The identity $p(\mathbf{s}^2) = \sum_{\mathbf{s}^1} p(\mathbf{s}^2 \mid \mathbf{s}^1) p(\mathbf{s}^1)$ is not readily reversible in a way that would enable us to select an adequate conditional design for a given marginal design.

Moreover, the conditional sampling design must respect the conditional inclusion probabilities. These constraints prevent us from using some conditional sampling designs. For instance, suppose that $p(\mathbf{s}^1)$ and $p(\mathbf{s}^2)$ have unequal inclusion probabilities π_k^1 and π_k^2 such that $\pi_k^1 + \pi_k^2 < 1$ for all $k \in U$. Then, even if $\mathbf{s}^1$ is selected with a fixed sample size n^1, there is no particular reason why

$$\sum_{k\in U} \pi_k^2(s_k^1) = \sum_{k\,|\,s_k^1=0} \pi_k^2/(1-\pi_k^1)$$

would be an integer, and there is even less reason for it to be equal to

$$\sum_{k\in U} \pi_k^2.$$

Hence, with a longitudinal systematic design, it is not possible to have complete control over the cross-sectional design of the second wave. The size of the conditional sampling design for the second wave may be random. This question is also discussed in TILLÉ and FAVRE (2004).

The method described in this section can be generalized for any number of sampling occasions. We compute the conditional inclusion probabilities at time t:

$$\pi_k^t(s_k^{t-1},\dots,s_k^1) = \Pr(S_k^t = 1 \mid S_k^{t-1} = s_k^{t-1},\dots,S_k^1 = s_k^1),$$

for a systematic longitudinal design. These conditional probabilities are computed at line 6 in Algorithm 5 for any time t. After several waves, more and more conditional inclusion probabilities are equal to 0 or 1, as the interval $[a,b]$ of Algorithm 5 becomes smaller. Hence, the conditional inclusion probabilities become more and more deterministic and the conditional sampling design cannot be freely chosen. This problem sheds light on the antagonism between the desire for a good rotation system and the control over the cross-sectional sampling designs.

6 Other solutions to the coordination problem

6.1 The dilemma of sampling coordination

In the previous section we have seen that, while the systematic design is a good longitudinal sampling design, its use leads to a considerable loss of control over the cross-sectional sampling design. This is becuase of the fact that systematic sampling has a very small support. In order to avoid this problem, we must choose a longitudinal design that gives more freedom to the user, such as Deville's design or the minimum time out of sample design, and possibly cope with deteriorated coordination.

Until now, there were two main approaches to the coordination problem. They are as follows.

- To choose the cross-sectional design and try to get the best coordination. This is the aim of the COTTON and HESSE (1992b) method, and of the RIVIÈRE (1998, 1999, 2001a,b) method. However, a simple example shows that these methods do not always provide the best longitudinal design for a fixed cross-sectional design on a population with changing strata. Unfortunately the only way, that we know of, to identify this optimal design consists in applying a linear programme on all the possible samples. In most situations, this method is not practical.

- To choose a longitudinal systematic design and accept the progressive loss of control over the cross-sectional design.

We propose a new solution that makes a compromise between the control of the longitudinal and of the cross-sectional sampling designs. As in Deville's systematic-simple repeated design, we use a longitudinal design that has a larger support than systematic sampling.

6.2 The minimum time out of sample method

We have seen that any longitudinal design can be applied provided that there is a strictly sequential algorithm to implement it. Being able to inform a unit that it will not be sampled for a fixed minimum number of waves after it has been sampled is a nice feature, so we propose to use the minimum time out of sample method for the longitudinal design. Moreover, this method allows us to set the number of previous waves that can have an influence on the present. If r is the fixed time out of sample, we have that

$$\Pr(S_k^{t+r}=1 \mid S_k^1,\dots,S_k^{t+r-1})=\Pr(S_k^{t+r}=1 \mid S_k^t,\dots,S_k^{t+r-1}).$$

At the first wave, any cross-sectional sampling design can be applied on the inclusion probabilities π_k^1 for $k\in U$. Then, using Algorithm A1 or Algorithm 8, it is possible to compute the conditional inclusion probabilities $\Pr(S_k^2=1 \mid S_k^1=s_k^1), k\in U$. At the second wave, any compatible sampling design can be applied on these conditional inclusion probabilities.

At time t, the conditional inclusion probabilities $\Pr(S_k^t=1 \mid S_k^1=s_k^1,\dots,S_k^{t-1}=s_k^{t-1})$ can again be computed with Algorithm 8 or Algorithm A1. Any compatible cross-sectional sampling design can then be applied using these inclusion probabilities. If $t\geq r+1$, we just need to know $\pi_k^{t-r},\dots,\pi_k^t$ and $s_k^{t-r},\dots,s_k^{t-1}$ in order to compute the conditional probabilities. The implementation of this algorithm is thus relatively simple and practical.

7 Conclusions

We made an attempt to develop a general theory for the problem of units rotation in repeated sampling. The main methods that are currently in use have well-known cross-sectional designs and we derived their longitudinal designs. Longitudinal systematic sampling plays a fundamental role in sampling coordination because it provides a good coordination. However, it results in a rapid loss of control over the cross-sectional sampling design. This problem highlights the deep antagonism between control of the cross-sectional design and coordination. Whatever the adopted solution may be, it is not possible to have at the same time the best coordination and a complete choice of cross-sectional design. We offer a compromise that allows us to have a relatively free choice of cross-sectional design while providing a good coordination between the samples.

Acknowledgements

The authors wish to thank the editor and anonymous referees for their helpful suggestions and comments that helped in improving this article. The authors are also grateful to the Swiss National Science Foundation (grant FN205121-105187/1) and the Swiss Federal Statistical Office for the financial support. The views expressed herein are those of the authors and not necessarily those of the Swiss Federal Statistical Office or of the Swiss National Science Foundation.

Appendix: A new algorithm for unit rotation

The aim of Algorithm 8 was to impose a fixed number of steps during which a unit, once selected, is not selected anymore. However, this is only possible if the sum of the inclusion probabilities for any r successive occasions does not exceed 1. If this condition is not verified, Algorithm 8 cannot be applied. One solution that allows for any vector of inclusion probabilities is to use Algorithm A1. However, in this case the minimum time out of sample cannot always be respected. Algorithm A1 gives the exact results as Algorithm 8 when the sums of the inclusion probabilities do not exceed 1.

Algorithm A1. Minimum time out of sample sequential algorithm, without conditions on the inclusion probabilities

```
Define $\boldsymbol{\pi}_k = (\pi_k^1, \ldots, \pi_k^T)'$, a vector of inclusion probabilities.
Define $\mathbf{s}_k = (s_k^1, \ldots, s_k^T)' = (0, \ldots, 0)'$, the empty sample.
Fix $r$, the number of times that a unit must stay out of the sample.
for $t = 1, \ldots, T$ do
  Generate $u^t$, a uniformly distributed random number in $[0, 1)$.
  $p = 0$
  if $\pi_k^t \geq 1$ then
    $j = t$
  else
    $j = \max(t - r - 1, 1)$
    while $\sum_{i=j}^{t} \pi_k^i > 1$ do
      $j = j + 1$
    end while
  end if
  if $j = t$ then
    $p = \pi_k^t$
  else
    if $\sum_{i=j}^{t-1} s_k^i = 0$ then $p = \pi_k^t / (1 - \sum_{i=j}^{t-1} \pi_k^i)$ end if
  end if
  if $u^t < p$ then $s_k^t = 1$ end if
end for
```

This method can easily be adapted to dynamic populations. A newborn unit, at time t, will receive an imaginary past, that is, $\pi_k^i = s_k^i = 0, i \leq t$. If a unit leaves the population at time t, then its inclusion probability becomes equal to zero at times $t, t+1, \ldots, T$. The adjustment of Algorithm A1 to dynamic populations is straightforward. We just add the following line between lines 1 and 2 of the algorithm:

1b: Each newborn unit at time t receives a vector of inclusion probabilities $\boldsymbol{\pi}_k = (0, \ldots, 0, \pi_k^t, \ldots, \pi_k^T)'$.

References

Brewer, K., L. Early and M. Hanif (1984), Poisson, modified Poisson and collocated sampling, *Journal of Statistical Planning and Inference* **10**, 15–30.

Brewer, K., L. Early and S. Joyce (1972), Selecting several samples from a single population, *Australian Journal of Statistics* **3**, 231–239.

Chen, S., A. Dempster and J. Liu (1994), Weighted finite population sampling to maximize entropy, *Biometrika* **81**, 457–469.

Cotton, F. and C. Hesse (1992a), Co-ordinated selection of stratified samples, *Proceedings of Statistics Canada Symposium 1992*, 47–54.

Cotton, F. and C. Hesse (1992b), *Tirages coordonnes d'échantillons. Document de travail de la Direction des Statistiques Économiques E9206*, Technical Report, INSEE, Paris.

Deville, J.-C. (1998). *Une nouvelle (encore une!) méthode de tirage à probabilités inégales*, Technical Report 9804, *Méthodologie Statistique*, INSEE, Paris.

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

Fan, C., M. Muller and I. Rezucha (1962), Development of sampling plans by using sequential (item by item) selection techniques and digital computer, *Journal of the American Statistical Association* **57**, 387–402.

Hodges, J. and L. LeCam (1960), The Poisson approximation to the Poisson binomial distribution, *Annals of Mathematical Statistics* **31**, 737–740.

Johnson, N., S. Kotz and A. Kemp (1992), *Univariate discrete distributions* (2nd ed.), John Wiley and Sons, New York.

Koeijers, E. and A. Willeboordse (1995), *Reference manual on design and implementation of business surveys*, Technical Report, Statistics Netherlands.

Péa, J., L. Qualité and Y. Tillé (2007), Systematic sampling is a minimum support design, *Computational Statistics and Data Analysis* **51**, 5591–5602.

De Ree, S. J. M. (1983), A system of co-ordinated sampling to spread response burden of enterprises, Contributed Paper, *Proceedings: 44th Session of the ISI*, Madrid, 673–676.

De Ree, S. J. M. (1999), Co-ordination of business samples using measured response burden, Invited Paper, *Proceedings: 52nd Session of the ISI*, Helsinki.

Rivière, P. (1998), *Description of the chosen method: deliverable 2 of supcom 1996 project (part "co-ordination of samples")*, Report, EUROSTAT, 11–33.

Rivière, P. (1999), Coordination of samples: the microstrata methodology, *Proceedings of the 13th Roundtable on Business Survey Frames*, INSEE, Paris.

Rivière, P. (2001a), Coordinating samples using the microstrata methodology, *Proceedings of Statistics Canada Symposium 2001.*

Rivière, P. (2001b), *Random permutations of random vectors as a way of co-ordinating samples*, Working Paper, University of Southampton.

Tillé, Y. (2006), *Sampling algorithms*, Springer-Verlag, New York.

Tillé, Y. and A.-C. Favre (2004), Co-ordination, combination and extension of optimal balanced samples, *Biometrika* **91**, 913–927.

Van Huis, L., C. Koeijers and S. J. M. De Ree (1994a), *EDS, sampling system for the Central Business Register at Statistics Netherlands*, Internal Report, Statistics Netherlands.

Van Huis, L., C. Koeijers and S. J. M. De Ree (1994b), *Response burden and co-ordinated sampling for economic surveys*, Internal Report, Vol. 9, Statistics Netherlands.