# Sampling with double reduction of the response burden

Alina Matei [1] & Paul A. Smith [2]

[1] *University of Neuchâtel, Switzerland, alina.matei@unine.ch*
[2] *University of Southampton, UK, p.a.smith@soton.ac.uk*

**Résumé.**

Les méthodes d'échantillonnage coordonné cherchent à augmenter (en coordination positive) ou à diminuer (en coordination négative) la taille du chevauchement des échantillons. Les échantillons peuvent être sélectionnés à différents moments dans une enquête répétée dans le temps ou dans des enquêtes distinctes. En raison de la charge de réponse, certaines unités ne répondent pas aux questionnaires si elles sont sélectionnées dans de nombreux échantillons. Pour éviter ce problème, on utilise une coordination négative des échantillons. En général, les méthodes d'échantillonnage coordonné n'utilisent pas d'information sur la charge qu'une unité a supportée en répondant aux enquêtes précédentes. Nous utilisons une information de ce type dans une nouvelle approche, en modifiant une méthode d'échantillonnage étalé spatialement et en utilisant une méthode de coordination négative avec des nombres aléatoires permanents. L'objectif est de garantir que certaines unités connues comme des non-répondants 'notoires' seront resélectionnées moins souvent que dans le cas des échantillons indépendants. Nous réalisons avec cette méthode une double réduction de la charge de réponse des telles unités: une fois par l'utilisation d'une coordination négative des échantillons (avec des nombres aléatoires permanents) et une fois par l'utilisation d'une information sur la charge qu'une unité a supportée en répondant aux enquêtes précédentes.

**Mots-clés.** échantillonnage, coordination des échantillons, nombres aléatoires permanents.

**Abstract.**

Sample coordination methods seek to increase (in positive coordination) or decrease (in negative coordination) the sample overlap size. Samples can be selected on different occasions in a repeated survey or over different surveys. Due to response burden some units do not answer the survey questionnaires if they are selected in many samples. In order to reduce the unit response burden, negative sample coordination is applied. Thus, the response burden diminishes when sample overlap size diminishes. Usually, methods for sample coordination do not take into account the cumulated burden that a unit has already expended in responding to previous surveys. We introduce such a measure into a new method by modifying a spatially balanced sampling scheme, together with a negative coordination method with permanent random numbers. The goal is to ensure that units that are known to be 'notorious' non-respondents will be reselected less often than under independent selections. With this method we create a double reduction of the burden for these units: once by using permanent random numbers and once by using a measure of burden during the sampling process.

**Keywords.** sampling, sample coordination, permanent random numbers.

# 1 Sample coordination

We use two overlapping finite populations of units denoted by $U_1$ and $U_2$, respectively. From $U_1$ one selects a sample $s_1$, and from $U_2$ a sample $s_2$, using the sampling designs $p_1$ and $p_2$, respectively. The expected sample sizes for $s_1$ and $s_2$ are denoted respectively by $n_1$ and $n_2$. A bivariate sample $s = (s_1, s_2)$ is selected from $U_1 \times U_2$, using a joint sampling design $p$, with marginal sampling designs $p_1$ and $p_2$. The samples $s_1$ and $s_2$ are said to be *coordinated* if

$$p(s_1, s_2) \neq p_1(s_1)p_2(s_2),$$

that is, the samples are not drawn independently (see Cotton and Hesse, 1992; Mach et al., 2006). The main goal of sample coordination methods is to optimize the overlap between samples.

The size of the *overlap* between $s_1$ and $s_2$, denoted by $c$, represents the number of units common to $s_1$ and $s_2$. It is in general a random variable having expectation

$$E(c) = \sum_{k \in U} \pi_{k,12},$$

with $\pi_{k,12} = P(k \in s_1, k \in s_2) = \sum_{(s_1,s_2)} p(s_1, s_2), \forall k \in U$, and $U = U_1 \cup U_2$ is the so-called 'overall population'. In positive coordination, the goal is to maximize $E(c)$, while in negative coordination it is to minimize it. Let $\pi_{k1} = P(k \in s_1)$ and $\pi_{k2} = P(k \in s_2)$ be the first-order inclusion probabilities of unit $k \in U$ in the first and second samples, respectively. In order to manage new units ('births') in $U \setminus U_1$ and old units ('deaths') in $U \setminus U_2$, we assume that $\pi_{k1} = 0$ if $k \notin U_1$ and $\pi_{k2} = 0$ if $k \notin U_2$.

Various sample coordination methods have been proposed in the literature. For an overview on sample coordination methods, one can see, for instance, Matei and Smith (2022). An important approach in sample coordination is the use of *permanent random numbers* (PRNs). PRNs were introduced by Brewer et al. (1972) to coordinate Poisson samples, and have been widely used afterwards in other, different methods (Ohlsson, 1995).

Brewer et al. (1972) provided the following method: one associates a uniform random number $u_k$ drawn independently from the $U(0, 1)$ distribution with each unit $k \in U$, which is used in the selection process for samples across all surveys or different time occasions. The numbers $u_k$ are called 'permanent random numbers' since they are kept over time and over surveys for units which persist in the population. For a 'birth', a new PRN is assigned; for a 'death', the unit and its associated PRN are deleted from the corresponding survey frame.

Usually, a coordination method does not use any measure of response burden in the sampling process. If a negative coordination method is applied, the unit response burden diminishes since the size of the sample overlap is minimized. Our goal is to produce a double reduction of the unit response burden: first by using a method for negative sample coordination, and second by using a measure of the response burden in the sampling process. To do this, we modify spatially correlated Poisson sampling (Grafström, 2012), and use PRNs to coordinate samples, similarly to the method of Grafström and Matei (2018).

# 2 Sampling with double reduction of the response burden

In spatial sampling, the units in $U$ have associated geographical coordinates. It is thus possible to compute distances between units usually using Euclidean distance. Units close in distance provide in general similar information. Spatially balanced sampling allows the selection of units that are spread over the space, and thus avoids to collect similar information. A sampling design useful for selecting spatially balanced samples is spatially correlated Poisson (SCP) sampling (Grafström, 2012).

Due to the response burden, some units do not answer the survey questionnaires if they are selected in many samples. Over the time, some such units become 'notorious' non-respondents. It is possible to classify the units in $U$ in two categories: potential respondents and 'notorious' non-respondents. This information represents a measure of the response burden in the sampling process. We propose to replace the matrix of geographical coordinates in SCP sampling by the matrix formed by this measure together with the vector of inclusion probabilities. The Euclidean distances between units are next computed using this new matrix. If a 'notorious' non-respondent is selected in the current sample, SCP sampling avoids selecting a similar unit. Next, we use a negative coordination with PRNs of different samples, similar to the approach of Grafström and Matei (2018). We create with this method a double reduction of the response burden of the 'notorious' non-respondents: once by using PRNs and once by using a measure of burden during the sampling process.

# References


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

Cotton, F. and Hesse, C. (1992). Tirages coordonnés d'échantillons. Technical Report E9206, Direction des Statistiques Économiques, INSEE, Paris, France.

Grafström, A. (2012). Spatially correlated Poisson sampling. *J. Statist. Plann. Inference*, 142:139–147.

Grafström, A. and Matei, A. (2018). Coordination of spatially balanced samples. *Survey Methodology*, 44(2):215–238.

Mach, L., Reiss, P. T., and Şchiopu-Kratina, I. (2006). Optimizing the expected overlap of survey samples via the northwest corner rule. *Journal of the American Statistical Association*, 101(476):1671–1679.

Matei, A. and Smith, P. A. (2022). Sample coordination methods and systems for establishment surveys. In Snijkers, G., Bavdaž, M., Bender, S., Jones, J., MacFeely, S., Sakshaug, J., Thompson, K., and van Delden, A., editors, *Advances in Business Statistics, Methods and Data Collection*. Wiley, New York, USA.

Ohlsson, E. (1995). Coordination of samples using permanent random numbers. In Cox, B. G., Binder, D. A., Chinnapa, B. N., Christianson, A., Colledge, M. J., and Kott, P. S., editors, *Business Survey Methods*, chapter 9, pages 153–169. Wiley, New York, USA.