This module is part of the

Memobust Handbook

on Methodology of Modern Business Statistics

26 March 2014

Method: Sample Co-ordination Using Poisson Sampling with Permanent Random Numbers

Contents

General section

1.   Summary

The method described here allows for the selection of positively or negatively co-ordinated samples, that is to say samples with large or on the contrary small overlap with previous samples. Negative co-ordination is particularly useful in order to spread the response burden evenly on the population by avoiding that the same units get selected in different samples when this is not necessary. Positive co-ordination is desirable when one wants to update a panel sample, or to be able to compare accurately results of a new survey with those of a previous survey. The proposed method is suitable for unequal inclusion probability surveys and dynamic populations with births, deaths, mergers and splits of units.

2.   General description of the method

This is an extension of Brewer et al. (1972)'s method for the selection of two positively or negatively co-ordinated samples to the case of an indefinite number of surveys. It belongs to the family of Permanent Random Numbers (PRN) methods for selecting co-ordinated samples. These methods rely on the generation of random numbers for each unit that enters the population and on the principle that these random numbers govern the selection of future samples. In the method presented here, a unit is attached to its random number until it quits the population, so that the random numbers are actually permanent. In other methods, such as in Cotton and Hesse (1992) and Rivière (2001), random numbers are rotated among units after each sample selection, so that random numbers do indeed govern future sample selections, but they are not permanently attached to a given unit of the population. An account on sample co-ordination methods with PRN can be found in Ohlsson (1995),

In the method presented here, each unit of the population is dealt with independently from the others, so that obtained transversal samples, for each sampling occasion, are Poisson samples (see for example Tillé, 2006). An extensive description of the core of the method is available in Qualité (2009), and a shorter one in this document and in Qualité (2011). It can be summarised in the few following statements.

Each unit k of the population receives a permanent random number u_k generated uniformly in $(0,1)$ and independently from the random numbers of other units. If π_k^t denotes the inclusion probability of unit k at a given sampling occasion t, then unit k is selected in the sample if and only if its random number lies in a chosen subset of $(0,1)$ that has a total length equal to π_k^t. This ensures that the probability of selecting unit k in this sample is equal to π_k^t, as long as the choice of this subset is made without information on u_k. The co-ordination between surveys is obtained by a careful choice of the different selection subsets for all sampling occasions. A maximal positive co-ordination between two surveys is obtained when the corresponding selection subsets have the largest possible overlap. A maximal negative co-ordination is obtained when selection subsets are non-overlapping, if possible, or have the smallest possible overlap. The construction of these selection subsets for all sampling occasions is thus the main point of the co-ordination method.

This description, and the method used at the Swiss Federal Statistical Office (SFSO) to construct the selection intervals are better understood recursively and on an example. Since each unit of the

population is dealt with independently from the others, it is sufficient to examine what may happen for a generic unit k.

Suppose that this unit k has inclusion probabilities π_k^1, at the first sampling occasion, π_k^2 at the second sampling occasion and π_k^3 at the third sampling occasion.

1. At the first sampling occasion, the selection set is naturally defined to be $(0, \pi_k^1)$, for all k (see figure 1). This selection set has a correct length of π_k^1, and the probability that u_k is in this set, and thus that unit k is selected in the first sample is equal to π_k^1.
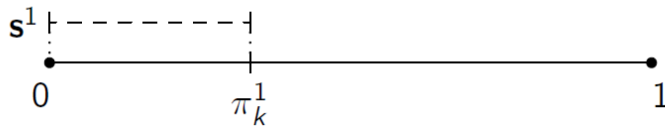

Figure 1. First sampling occasion

2. The second survey may be either positively co-ordinated or negatively co-ordinated with the first survey. We consider only maximal co-ordination. If the desired co-ordination is positive, the selection subset for the second survey is defined as $(0, \pi_k^2)$, in figure 2. If u_k is in $(0, \min(\pi_k^1, \pi_k^2))$, unit k is selected in both samples. If it is in $(\min(\pi_k^1, \pi_k^2), \max(\pi_k^1, \pi_k^2))$, then unit k is selected in one sample but not the other, and if it is in $(\max(\pi_k^1, \pi_k^2), 1)$, then unit k is selected in neither sample. Consequently, the probability of selecting unit k in both samples is maximal and equal to $\min(\pi_k^1, \pi_k^2)$, the minimum of π_k^1 and π_k^2.
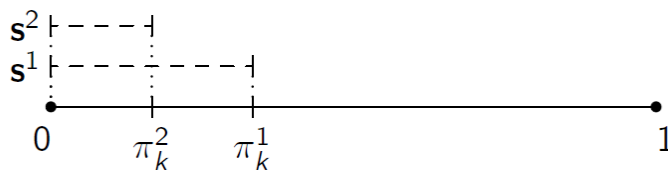

Figure 2. Second sampling occasion, positive co-ordination

If, on the contrary, the desired selection in negative, then two cases may occur:

4

a. If $\pi_k^1 + \pi_k^2 \leq 1$, then the selection subset for the second survey is defined as $\left(\pi_k^1, \pi_k^1 + \pi_k^2\right)$, in figure 3. Unit k is selected in at most one of the two samples.
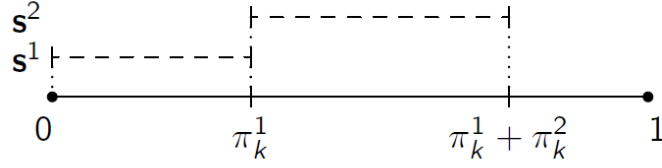

Figure 3. Second sampling occasion, negative co-ordination, first case

b. If $\pi_k^1 + \pi_k^2 > 1$, then the selection subset for the second survey is defined as $\left(\pi_k^1, 1\right) \cup \left(0, \pi_k^1 + \pi_k^2 - 1\right)$, in figure 4. Unit k can be selected in both samples, with probability $\pi_k^1 + \pi_k^2 - 1$, which is the theoretical minimal bound in this case.
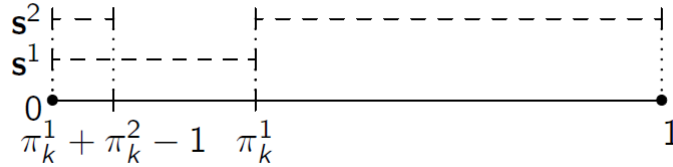

Figure 4. Second sampling occasion, negative co-ordination, second case

3. The third survey may be positively or negatively co-ordinated with the first or the second survey. The method described here allows to mix positive co-ordination with some surveys and negative co-ordination with others, but requires an order of priority for these co-ordinations. Exploring all possibilities for the third sampling occasion would be extremely tedious, but it is enough to investigate a simple example in order to understand the general idea. Suppose that the two first surveys were positively co-ordinated, so that the situation is that of figure 2, and that the third survey must be positively co-ordinated with the second, and, with a lesser priority, negatively co-ordinated with the first. Suppose moreover that π_k^3 is larger than π_k^2. The first objective is obtained by choosing a selection subset that overlaps the most the selection set of the second survey. As π_k^3 is larger than π_k^2, the whole subset $\left(0, \pi_k^2\right)$ is included into the selection subset of the third survey. Then, it needs to be completed with an additional subset of length $\pi_k^3 - \pi_k^2$ that respects the most the remaining co-ordination rules: this additional subset should not overlap, if possible, the selection subset of the first survey. The solution is thus, in that case, to define the selection subset for the third sample as $\left(0, \pi_k^2\right) \cup \left(\pi_k^1, \pi_k^1 + \pi_k^3 - \pi_k^2\right)$, in figure 5.
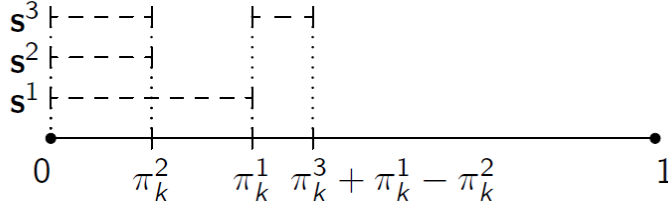
Figure 5. Third sampling occasion

4. Recursively, after t sampling occasions, we want to select a $(t+1)^{th}$ survey co-ordinated negatively with none, some, or all previous surveys and positively with the other surveys. A strict order of priority for these co-ordinations is required. Usually, but not necessarily, the reverse chronological order is used. The interval $(0,1)$ is subdivided into a collection of subsets that are the intersections of all selection subsets of previous surveys. These subsets are then ranked by strict order of compliance with the desired co-ordination rules, so that subsets that respect co-ordinations with higher priority are ranked higher than intervals that do not respect them. The selection subset for the $(t+1)^{th}$ survey is then obtained by including the top ranked subsets and, when necessary, part of one of these subsets, up to a total length of π_k^{t+1}. The progression of the number of subsets to consider is thus linear. For each unit, it is at most equal to $t+1$ after t sampling occasions. This last point is the feature that ensures that the system does not exceed computation capacities too rapidly, which is the main difficulty with multidimensional sampling designs.

3.   Preparatory phase

Before any new survey is selected, different tasks have to be accomplished. First, if relevant, the sampling frame can be updated using available information in the business register. Second, the inclusion probabilities, or sampling design for the new survey need to be determined. Finally, co-ordination rules must be decided so that the selection subsets defined in section 2 may be computed.

3.1   Updating the sampling frame

When the sampling frame is updated, newborn units are added to the population with an empty selection history, trivial selection subsets and an independently generated PRN. Deceased units are removed from the frame. Parent units from splits and mergers can transmit their history and PRN to child units. However, in order to keep independence between units selection, each history and PRN can only be transmitted to one unit.

3.2   Computing inclusion probabilities for the new survey

With the introduction of this co-ordinated sampling method, all transversal sampling designs have been replaced by Poisson designs. These were previously, at the SFSO, mostly stratified sampling designs. Optimal sample allocation procedures need not necessarily be modified, but special care must be given to small sampling strata, for which Poisson sampling entails the risk of selecting an empty

sample, as is already the case when non-response is possible. When such small strata are present, it may be considered to deviate a bit from the optimal allocation, in order to keep this risk acceptable.

3.3   Choosing co-ordination rules between the new survey and previous surveys

For most applications, reverse chronological order is used, but when a panel or a rotating panel is updated, it makes sense to ensure the co-ordination first between the previous panel sample and the new one, and only then with other surveys.

4.   Examples – not tool specific

4.1   One-occasion business surveys, panels and rotating panels in Switzerland

The SFSO has been using a co-ordinated sampling system for business surveys since October 2009. One occasion surveys as well as panels and rotating panels have been selected through this system. Most notably, the survey on value-added, a rotating panel survey with 20% rotation rate, has been selected and thrice updated. The rotation is achieved by considering the survey as a collection of five non-overlapping co-ordinated surveys: the rotation groups. Each year a fresh rotation group is selected, negatively co-ordinated with those surveyed the previous year, and four of the previous rotation groups are updated. This updating is obtained by selecting sample for these four groups anew, with maximal positive co-ordination with the samples they are supposed to replace. Other ongoing business surveys are progressively being integrated in this co-ordinated sampling system, when they undergo planned extensive revision and maintenance.

5.   Examples – tool specific

6.   Glossary

For definitions of terms used in this module, please refer to the separate "Glossary" provided as part of the handbook.

7.   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), Coordinated Selection of Stratified Samples. *Proceedings of Statistics Canada Symposium*.

Ohlsson, E. (1995), Coordination of samples using permanent random numbers. In: *Business Survey Methods* (eds. Cox, B. G., Binder, D. A., Chinnappa, D. N., Christianson, A., Colledge, M. J., and Kott, P. S.), John Wiley and Sons, New York, 153–169.

Qualité, L. (2011), Developments on Coordinated Poisson Sampling. Presented at the 58th congress of the International Statistical Institute, Dublin.
http://isi2011.congressplanner.eu/pdfs/950958.pdf

Qualité, L. (2009), *Unequal probability sampling and repeated surveys*. Ph.D. thesis, University of Neuchâtel, Switzerland (http://doc.rero.ch/record/12284).

Rivière, P. (2001), Coordination Sample using the Microstrata Methodology. *Proceedings of Statistics Canada Symposium*.

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

Specific section

8.   Purpose of the method

This method provides a co-ordinated sampling system. Poisson samples can be selected with positive or negative co-ordination with previous samples that were selected within the co-ordinated sampling system. One occasion surveys, panels and rotating panels can be selected while spreading the response burden as evenly as possible on the population.

9.   Recommended use of the method

1. This method is recommended for the selection of moderate to large sample surveys, or of a large number of small surveys, when co-ordination between samples is an important feature, and when correlation between selections of different units of the population at a given sampling occasion is not needed.

10.   Possible disadvantages of the method

1. This method provides Poisson transversal designs, with random size. It is usually not a real problem since the sampling-related variability is secondary compared to the variability due to non-response and to its possibly inaccurate anticipation. The implied increase in estimation variance is negated by the use of calibrated estimators. However, precautions must be taken to limit the risk of having empty or too small samples in interest domains with modest planned sample size.

2. The independence between units selection makes it impossible to select samples where a strong dependence is required, such as for face-to-face surveys where geographic proximity is an important cost factor. For the same reason, this method does not provide a global co-ordinated sampling system for businesses and local units surveys, or for households and population surveys.

11.   Variants of the method

1. Some dependence between units selection at a given occasion can be introduced by using the co-ordinated sample selection as one phase of a multiphase sampling design. For example, if the system is used to select local units samples, one may not want that more than one or a given number of local units of any business is selected. In the case of population surveys, one may want to avoid multiple selections within households. Inclusion probabilities at each sampling phase must then be computed, and co-ordination between actual surveys on field is necessarily degraded.

2. A rotating panel is obtained by splitting the survey into a collection of smaller surveys corresponding to rotation groups. In order to do so, the rotation rate must be a fraction of one, and kept constant over time.

12.   Input data

1. Ds_input1: Sampling frame, essentially a list of identifiers, but other variables are useful in order to produce monitoring indicators;

2. Ds_input2: Inclusion probabilities for all units in the sampling frame, with zeroes for units outside of the target population;

3. Ds_input3: Sign of the desired co-ordination with previous surveys, and an order of priority for these co-ordinations;

4. Ds_input4: History of all units in the sampling frame, with their PRN and collection of selection subsets for all previous surveys.

13.   Logical preconditions

1. Missing values

    1. No missing values are allowed.

2. Erroneous values

    1. Erroneous values for auxiliary variables of the sampling frame will not prevent the selection from occurring, but monitoring indicators will be incorrect. Other tables must be clean.

3. Other quality related preconditions

    1. Ds_input1, Ds_input3 and Ds_input4 can be joined with the unique identifier present in all three tables.

4. Other types of preconditions

    1.

14.   Tuning parameters

1. None.

15.   Recommended use of the individual variants of the method

1. When using the co-ordinated sampling as one phase of a multiphase design, organisation of the sampling phases should be such that the inclusion probabilities for the co-ordinated sampling phase are as close as possible to the true final inclusion probabilities. If this is the case, the co-ordination is still efficient. Usually, this can be achieved by using the co-ordinated sampling for the first phase of selection and then tailoring the sample to verify all requirements.

2. When selecting rotating panels, special attention has to be given to the co-ordination priorities in order to ensure that rotation groups at a given sampling occasion do not overlap and that units of the exiting rotation group do not re-enter the panel except when necessary.

16.   Output data

1. Ds_output1: updated history of each unit of the population, after the current sampling occasion;

2. Ds_output2: selected sample for the current sampling occasion.

17. Properties of the output data

1. The selected sample is selected with a Poisson sampling design, with specified inclusion probabilities.

2. The co-ordination recorded in Ds_output1 and obtained between the current sample and the previous ones is optimal, in the sense that the current survey is optimally co-ordinated with the one with highest priority. Then, within the remaining wiggle room, it is optimally co-ordinated with the second survey by order of priority, and so on.

18. Unit of input data suitable for the method

Incremental processing of population units is possible since units are treated independently.

19. User interaction - not tool specific

1. Computation of inclusion probabilities.

2. Determination of co-ordination rules (sign and priority).

20. Logging indicators

1. Processing time.

2. Size of the population and expected size of the sample.

3. Number of selections in the population, within the system, as a measure of survey burden.

21. Quality indicators of the output data

1. Comparison between expected and obtained size of the sample.

2. Size of overlaps between the current survey and all previous surveys, number of repeated selections within the system.

22. Actual use of the method

1. Population surveys issued by the SFSO in Switzerland, starting in November 2010.

2. Business surveys issued by the SFSO in Switzerland, starting in October 2009.

Interconnections with other modules

23. Themes that refer explicitly to this module

1. Sample Selection – Sample Co-ordination

24. Related methods described in other modules

1. Sample Selection – Sample Co-ordination Using Simple Random Sampling with Permanent Random Numbers

25. Mathematical techniques used by the method described in this module

1. Basic arithmetic

26.   GSBPM phases where the method described in this module is used

1.   12

27.   Tools that implement the method described in this module

1. Unnamed SAS Macro at the SFSO.

28.   Process step performed by the method

Sample selection and co-ordination

Administrative section

29.   Module code

Sample Selection-M-PRN Using Poisson Sampling

30.   Version history

Version	Date	Description of changes	Author	Institute
0.1	12-06-2012	initial version	L. Qualité	SFSO (Switzerland)
0.2	15-05-2013	first revision	L. Qualité	SFSO (Switzerland)
0.2.1	18-09-2013	preliminary release		
1.0	26-03-2014	final version within the Memobust project		

31.   Template version and print date

Template version used	1.0 p 4 d.d. 22-11-2012
Print date	21-3-2014 17:44