This module is part of the

Memobust Handbook

on Methodology of Modern Business Statistics

26 March 2014

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

Contents

General section

1. Summary

The purpose of this module is to introduce the reader to the Permanent Random Number (PRN) technique for co-ordinating (stratified) simple random samples. The concept of co-ordination based on PRNs using simple random sampling was introduced in Sweden in the early 70s (see, e.g., Ohlsson 1992). This technique provides a practical solution for controlling the overlap (the number of common units) between different samples.

The methods developed and used by Statistics Sweden (Ohlsson,1992) and the Institut National de la Statistique et des Études Économiques (INSEE) of France (Cotton and Hesse, 1992) are both methods of co-ordination of simple random samples using PRNs and will be briefly described.

2. General description of the method

The National Statistical Institutes (NSIs) publish economic statistics based on business surveys. Often, different surveys use the same sampling frame (a register or a list frame) which leads to a need to co-ordinate sampling for the surveys. We distinguish two main kinds of co-ordination: positive and negative co-ordination which have different aims, i.e., maximising or minimising the overlap between samples. The general aspects of the sample co-ordination problem are described in detail in the theme module "Sample Selection – Sample Co-ordination".

With the PRN technique for co-ordinating samples both positive and negative co-ordination can be obtained. Several national agencies use variations of this technique. A review of basic PRN techniques and a comparison by countries can be found in Ohlsson (1995) and Hesse (1999).

2.1 Sequential simple random sampling

Consider a population of size N. This population could be also a stratum, the principle is the same. Each unit in the population is assigned a random number, uniformly distributed over the interval [0,1]. The units are sorted in ascending order of the random numbers. The sample is composed of the first n units in the ordered list. This technique is described in Fan et al. (1962) who called it "sequential", thus the name *sequential simple random sampling*. Ohlsson (1992) shows that this technique produces a simple random sampling without replacement (*srswor*). Due to the symmetry of the uniform distribution, the selection of the last n units in the ordered list of units also gives a sequential *srswor*. More precisely, selecting the first n units to the left, or to the right, of any fixed point a in [0,1] also yields a *srswor*. If there are not enough units to the right (or left) of the starting point a, then, depending on the chosen direction, the selection can continue to the right (or left) of the point 0 (the point 1). Thus, the sampling frame can be viewed as a circular list.

2.2 The "JALES" method

The Swedish SAMU system uses sequential *srswor* to co-ordinate samples across surveys and over time (SAMU is an acronym for "co-ordinated samples" in Swedish). The used method is referred to as "JALES". A full description of the system can be found in Ohlsson (1992) and Lindblom (2003). Following Ohlsson (1995), we present a brief overview of the method.

A PRN, uniformly distributed over the interval [0,1], is associated with each unit in the frame. Units who stay in the frame have the same random number on each sampling occasion. Every new business in the frame (a birth) is assigned a new PRN while closed-down businesses (deaths) are withdrawn from the sampling frame together with their assigned PRN.

Co-ordination over time is done in the following way. On each sampling occasion a new sequential *srswor* using PRNs is drawn from the updated (for births and deaths) frame. However, a large overlap with the previous sample can be expected, since persistent businesses have the same PRN on each occasion. This type of co-ordination enables good precision in estimates of change over time.

In order to co-ordinate samples with desired sample sizes n_1 and n_2 for two different surveys, we choose two constants a_1 and a_2 in [0,1]. Then we take the units with the n_1 PRNs closest to the right (or left) of a_1 to obtain the first sample and the n_2 units closest to the right or left of a_2 to obtain the second sample. Positive co-ordination of two surveys is obtained by using the same starting point and direction for both surveys. Negative co-ordination can be achieved if the starting points and sampling directions are properly chosen, e.g., different starting points and the same direction.

2.3 The Cotton & Hesse method

The Cotton & Hesse method is a PRN method based on sequential *srswor*. It is fully described in Cotton and Hesse (1992). The method, which allows only for negative co-ordination, could be used for co-ordinating samples for different surveys drawn the same year or for co-ordinating samples for a single survey over time. Each unit of the population receives a PRN from a uniform distribution U[0,1]. Then a sequential *srswor* of size n is drawn choosing the units with the n smallest random numbers (starting point $a = 0$). Negative co-ordination is obtained by permutation of the random numbers. The reordering is done in such a way that the selected units receive the largest PRNs and the non-selected units receive the smallest PRNs. Within the two subsets of selected and non-selected units, the order of the permuted PRNs must remain unchanged. This means that if a unit was assigned the smallest of the first n ordered PRNs, after reordering it will be assigned the smallest of the n ordered largest PRNs. Then a sequential *srswor* is drawn in the reordered list of units.

Note that, after permutation the PRNs remain independent uniform random numbers. So, the successive samplings remain *srswor*. As for SAMU, a closed-down unit loses its PRN and a new unit receives a new PRN. An interesting property of the method is that the minimum of the expected overlap between two successive simple random samples is obtained. The joint sampling design of each pair of subsequent samples is the same as this obtained by sequential *srswor* when the same starting point and opposite directions are chosen (Hesse, 1999).

2.4 Co-ordination of stratified samples

Many business surveys use stratified *srswor* in order to improve the accuracy of estimates by dividing the frame population into homogenous sub-populations (strata). The PRN technique for co-ordination can be easily adapted to this environment. In SAMU a sequential *srswor* is drawn in each stratum. For a specific survey, if the same direction and starting point are used in all strata, then we obtain positive co-ordination. As before, negative co-ordination for two surveys can be obtained by choosing different starting points or directions. SAMU allows for positive or negative co-ordination when different stratifications are used. The Cotton and Hesse method is easily adapted to stratified *srswor*. It also allows for different stratifications (Cotton and Hesse, 1992).

2.5 *Co-ordination of surveys based on different unit types*

The methods of co-ordination of simple random samples using PRNs can be used for surveys based on different unit types, e.g., establishments and enterprises. The module "Sample Selection – Assigning Random Numbers When Co-ordination of Surveys Based on Different Unit Types is Considered" explains in more detail how random numbers can be assigned to the different unit types and how co-ordination is done. This kind of co-ordination has been implemented and used in SAMU (Lindblom, 2003). The method of Cotton and Hesse also allows for co-ordinating sampling units belonging to different levels (Hesse, 1999).

2.6 *Rotation*

The main purpose of rotation is to spread the response burden among small businesses. Rotation is meaningful mainly for units with small inclusion probabilities. Sample rotation ensures that a unit will rotate out of the sample for a certain number of years and will not be included immediately in the sample of another survey. Using methods based on PRNs, we can handle sample rotation quite easily.

Let us assume, for example, that a sample should be rotated every year and persisting units with inclusion probabilities of less than 0.10 should leave the sample after five years. The Random Rotation Cohort (RRC) method which can deal with this situation was introduced in the Swedish SAMU in 1989 (see, e.g., Ohlsson, 1992; Lindblom, 2003). The principle is to designate randomly and permanently each unit in the frame to one of five rotation cohorts (groups). The random numbers are then shifted by 0.10 only in one rotation group each year. After the first year all units in rotation group one will shift PRNs 0.10 to the left. The second year, the PRNs in rotation group two are shifted 0.10 to the left and so on. Units with inclusion probabilities less than 0.10 can be expected to be out of sample after five years . Thus we achieve an expected rotation rate of 1/5. Units with larger inclusion probabilities are also rotated but it takes more time for them, depending on the size of the inclusion probability, to leave the sample. With the RRC method the sampling procedure remains sequential *srswor* which ensures the positive and negative co-ordination between surveys.

3. Preparatory phase

4. Examples – not tool specific

4.1 *Negative co-ordination with the "JALES" method of two surveys using two different starting points*

A simple example which illustrates how negative co-ordination for two samples works will be given in the following table. Suppose we have a population of 10 units. Each unit has been assigned a PRN drawn independently from U[0,1]. The desired sample size for both samples is equal to 6. In order to reduce the overlap between the two samples we choose two different starting points, $a_1=0$ and $a_2=0.6$, and the same direction (right of the starting point). The units are ordered in ascending order of their PRNs and then the first 6 units starting from $a_1=0$ are selected to compose the first sample, and the first 6 units starting from $a_2=0.6$ are selected to compose the second sample. Thus the first sample, S_1, is composed of units {5, 8, 1, 2, 4, 3} and the second sample, S_2, is composed of units {3, 10, 6, 7, 9, 5}. Units 3 and 5 are in both samples.

Unit number	PRN
1	0.25
2	0.4
3	0.7
4	0.5
5	0.1
6	0.8
7	0.9
8	0.2
9	1
10	0.75

Ordered list of the PRNs	Uunits after ordering	Sample 1	Sample 2
0.1	5	x	x
0.2	8	x	
0.25	1	x	
0.4	2	x	
0.5	4	x	
0.7	3	x	x
0.75	10		x
0.8	6		x
0.9	7		x
1.0	9		x

4.2 Positive co-ordination with the "JALES" method of a single survey over time

In this example we illustrate how positive co-ordination with the "Jales" method works for a single survey over time, the sampling frame being updated between times t=1 and t=2. We suppose that we have a population of 10 units at time t=1. Each unit has been assigned a PRN drawn independently from U[0,1]. The desired sample size is equal to 6. The units are ordered in ascending order of their PRNs and then the first 6 units starting from $a_1=0$ are selected to compose the sample at time t=1. Thus the sample, S^1, is composed of units {5, 8, 1, 2, 4, 3}. For simplification, we suppose that the population at time t=2 has the same number of units, thus we have as many births as deaths in the population. Deaths are denoted by - and births by +. The births receive a new PRN. The units are the ordered in ascending order of their PRNs and the first 6 units starting from $a_1=0$ are selected. Thus the sample at time t=2, S^2, is composed of units {5, 11, 13, 2, 4, 12}. Units 2, 4 and 5, which are persistent units in the sampling frame, are in the sample at both times t=1 and t=2.

Unit	PRN
1	0.25
2	0.4
3	0.7
4	0.5
5	0.1
6	0.8
7	0.9
8	0.2
9	1
10	0.75

Ordered PRNs	Unit	Sample t=1	Births/ Deaths
0.1	5	x	-
0.2	8	x	
0.25	1	x	-
0.4	2	x	
0.5	4	x	
0.7	3	x	
0.75	10		
0.8	6		-
0.9	7		
1.0	9		
			+
			+
			+

Unit	PRN
1	
2	0.4
3	
4	0.5
5	0.1
6	0.8
7	0.9
8	
9	1
10	0.75
11	0.18
12	0.58
13	0.28

Ordered PRNs	Unit	Sample t=2
0.1	5	x
0.18	11	x
0.28	13	x
0.4	2	x
0.5	4	x
0.58	12	x
0.75	7	
0.8	8	
0.9	9	
1	10	
	11	
	12	
	13	

4.3 Negative co-ordination over time with the Cotton and Hesse method

In this simple example, we consider co-ordinated sampling over time of three samples of size 3 in a population U of size 5. The initial PRNs are 0.5, 0.2, 0.4, 0.6, 0.8. The table below shows how the algorithm is executed at times t =1, 2 and 3 and what the selected samples are. At time t=1, we order the units by ascending PRN. We select the sample by taking the first three units in the ordered list of units. Thus, $S^1 = \{2,3,1\}$. Next, at time t=2, the PRNs are reassigned to the units of the population in the following way: Units 2, 3 and 1, which had the smallest three PRNs at t=1 receive the largest PRNs, by respecting the ranks of the PRNs, respectively 0.5, 0.6 and 0.8, which are the largest three PRNs in ascending order. Units 4 and 5, which were not selected at time t=1, receive the smallest PRNs, respectively, 0.2 and 0.4. PRN 0.2 (the smallest of 0.2 and 0.4) is assigned to unit 4 which had a PRN equal to 0.6 at time t=1 (the smallest of 0.6 and 0.8). Then the units are ordered by ascending order of their new PRNs and the 3 units with the smallest new PRNs are selected to form the sample at time t=2. Thus, $S^2 = \{4,5,2\}$. Following the same procedure, we obtain the sample $S^3 = \{3,1,4\}$. Thus, we obtain a minimum overlap between S^1 and S^2 (unit 1) and between S^2 and S^3 (unit 4).

t=1

Unit	PRNs t=1
1	0.5
2	0.2
3	0.4
4	0.6
5	0.8

Ordered PRNs at time t=1	Units after ordering at t=1	Sample at time t=1
0.2	2	x
0.4	3	x
0.5	1	x
0.6	4	
0.8	5	

t=2

Units after ordering at t=1	New PRN t=2
2	0.5
3	0.6
1	0.8
4	0.2
5	0.4

Ordered PRNs at time t=2	Units after ordering at t=2	Sample at time t=2
0.2	4	x
0.4	5	x
0.5	2	x
0.6	3	
0.8	1	

t=3

Units after ordering at t=2	New PRN t=3
4	0.5
5	0.6
2	0.8
3	0.2
1	0.4

Ordered PRNs at time t=3	Units after ordering at t=3	Sample at time t=3
0.2	3	x
0.4	1	x
0.5	4	x
0.6	5	
0.8	2	

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

Cotton, F. and Hesse, C. (1992), Co-ordinated selection of stratified samples. *Proceedings of Statistics Canada Symposium 92*, 47–54.

Fan, C., Muller, M., and Rezucha, I. (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.

Hesse, C. (1999), Sampling co-ordination: A review by country. Technical Report E9908, Direction des Statistique d'Entreprises, INSEE, Paris.

Lindblom, A. (2003), SAMU - The system for coordination of frame populations and samples from the Business Register at Statistics Sweden. Background Facts on Economic Statistics 2003:3, Statistics Sweden.

Ohlsson, E. (1992), SAMU. The system for Co-ordination of samples from the Business Register at Statistics Sweden. R&D report 1992:18, Statistics Sweden.

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

Specific section

8. Purpose of the method

The purpose of the method is to allow for positive or negative co-ordination of (stratified) simple random samples, when co-ordination of the same survey over time is desired or co-ordination between different surveys.

9. Recommended use of the method

1. To obtain positive co-ordination of samples over time for the same survey.
2. Co-ordination of samples for different surveys.
3. Co-ordination of samples with different designs.
4. Co-ordination of surveys based on different unit types.

10. Possible disadvantages of the method

1. Information from the surveys should not be used to update the list frame, as this could introduce bias in the estimates.
2. If they are different levels in the surveys, e.g., enterprise and establishment levels, the co-ordination of samples selected at different levels is less efficient for multiple-location enterprises, unless clustered (or two-stage) sampling of local units (establishments) is used.

11. Variants of the method

1. Rotation of samples is obtained by shifting the permanent random numbers of the sampling units which are divided into rotation groups.

12. Input data

1. Ds_input1: Sampling frame, essentially a business register.
2. Ds_input2: Auxiliary information, i.e., size measures for stratification can come from additional sources, e.g., an administrative register.

13. Logical preconditions

1. Missing values
   1.
2. Erroneous values
   1.
3. Other quality related preconditions
   1.
4. Other types of preconditions
   1.

14. Tuning parameters

1. Starting points and directions for co-ordination.

15. Recommended use of the individual variants of the method

1. Sample rotation.

16. Output data

1. Ds_output1: Selected co-ordinated sample for the current sampling occasion.

17. Properties of the output data

1. The selected sample is selected with a sequential *srswor*, thus is of fixed sample size.

18. Unit of input data suitable for the method

Incremental processing of frame units is possible since the units are treated independently.

19. User interaction - not tool specific

1. Delimitation of the sampling frame.
2. Determination of co-ordination rules, e.g., negative or positive co-ordination.
3. Determination of stratification and allocation.

20. Logging indicators

1.

21. Quality indicators of the output data

1. Number of repeated selections of an enterprise as a measure of response burden.
2. Size of the overlap between the current survey and the previous surveys.

22. Actual use of the method

1. The PRN technique for co-ordinating (stratified) simple random samples is actually used in the system for co-ordinating business surveys (SAMU) of Statistics Sweden (Lindblom, 2003) and in INSEE (Hesse, 1999). Other countries also use methods based on sequential *srswor*.

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 Poisson Sampling with Permanent Random Numbers

2. Sample Selection – Assigning Random Numbers When Co-ordination of Surveys Based on Different Unit Types is Considered

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

1.

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

1.

27. Tools that implement the method described in this module

1. SAMU (Ohlsson,1992)

28. Process step performed by the method

Sample selection

Administrative section

29. Module code

Sample Selection-M-PRN Using Simple Random Sampling

30. Version history

Version	Date	Description of changes	Author	Institute
0.1	03-05-2013	first version	Desislava Nedyalkova	SFSO (Switzerland)
0.2	07-06-2013		Desislava Nedyalkova	SFSO (Switzerland)
0.3	17-06-2013		Desislava Nedyalkova	SFSO (Switzerland)
0.3.1	18-06-2013		Desislava Nedyalkova	SFSO (Switzerland)
0.3.2	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 17:43