Coordinated sampling: Theory, method and application at Insee

Caroline Imberti

National Institute of Statistics and Economic Studies

March 1, 2019

Overview

Insee
Measuring, understanding

Plan

- Each year, the National Institute of Statistics and Economic Studies (INSEE) conducts a lot of surveys of firms.
- The aim of negative coordination is to focus on firms that have not recently been sampled while getting unbiased samples.
- Negative coordination enables to reduce the response burden of the smallest firms, the biggest ones are systematically sampled in most surveys.

The method used by INSEE is based on random numbers given to the units. Each unit is given a random number once for all. After that, these numbers are transformed for coordination.

To simplify, we suppose that all the surveys are based on the same population of firms.

Plan

Insee Measuring, understanding

The first method we used to apply at INSEE enables to coordinate samples two by two.

Samples of firms are often selected by stratified simple random sampling.

The population of firms $U = \{1, ..., k, ...N\}$ is divided into nonoverlapping subpopulations called strata and denoted $U_1, ...U_h, ..., U_H$.

Such a partition is made thanks to well-known features of the firms (activity, headcount, location,...).

2 The method we used before

First of all, each unit k is given a permanent random number ω_k that is a realization of a random variable Ω_k with a uniform distribution on $[0,1[$.
$\Omega_1, \Omega_2, \ldots\Omega_N$ are N independent random variables.

Unit	Random number	Stratum for the first survey
A	0,02	1
B	0,62	1
C	0,10	1
D	0,18	1
E	0,04	1
F	0,66	1
I	0,42	2
J	0,10	2
K	0,28	2
L	0,35	2
M	0,08	2
N	0,44	2

Insee
Measuring, understanding

For the first survey, we select the units that correspond to the smallest random numbers ω_k within each stratum.

Unit	Random number	Stratum for the first survey	First selection
A	0,02	1	Yes
E	0,04	1	Yes
C	0,10	1	Yes
D	0,18	1	No
B	0,62	1	No
F	0,66	1	No
M	0,08	2	Yes
J	0,10	2	Yes
K	0,28	2	Yes
L	0,35	2	Yes
I	0,42	2	No
N	0,44	2	No

Insee
Measuring, understanding

2 The method we used before

After that, we switch the random numbers within each stratum in order to select the second sample.

Unit	Random number	Stratum for the first surv ey	First selection
A	0,02	1	Yes
E	0,04	1	Yes
C	0,10	1	Yes
D	0,18	1	No
B	0,62	1	No
F	0,66	1	No
M	0,08	2	Yes
J	0,10	2	Yes
K	0,28	2	Yes
L	0,35	2	Yes
I	0,42	2	No
N	0,44	2	No

⇒

Unit	Random number	Stratum for the first surv ey	First selection
A	0,66	1	Yes
E	0,62	1	Yes
C	0,18	1	Yes
D	0,10	1	No
B	0,04	1	No
F	0,02	1	No
M	0,44	2	Yes
J	0,42	2	Yes
K	0,35	2	Yes
L	0,28	2	Yes
I	0,10	2	No
N	0,08	2	No

2 The method we used before

We select the second sample by the same way: within each stratum, we select the units that correspond to the smallest random numbers.

Unit	Random number	Stratum for the first survey	First selection	Stratum for the second survey	Second selection
B	0,04	1	No	1	Yes
I	0,10	2	No	1	Yes
L	0,28	1	Yes	1	Yes
E	0,62	1	Yes	1	No
A	0,66	1	Yes	1	No
F	0,02	1	No	2	Yes
N	0,08	2	No	2	Yes
D	0,10	1	No	2	Yes
C	0,18	1	Yes	2	Yes
K	0,35	2	Yes	2	No
J	0,42	2	Yes	2	No
M	0,44	2	Yes	2	No

Coordination enables to select in priority the units that have not been selected before.

Plan

Such a method enables to coordinate samples two by two. We would like to generalize it to coordinate as many samples as we want.

In particular, we would like to weight the samples according to the burden of the surveys, which is impossible with the current method.

3.1 Generalization of the current method

Instead of switching the random numbers within each stratum, we would like to apply a function g that enables the units that have not been sampled yet to be selected.

The current method is a particular case insofar as we apply a permutation to the random numbers.

Unit	Random number	Stratum for the first survey	First selection	Transformation of the random number
A	0,02	1	Yes	$g(0{,}02)$
E	0,04	1	Yes	$g(0{,}04)$
C	0,10	1	Yes	$g(0{,}10)$
D	0,18	1	No	$g(0{,}18)$
B	0,62	1	No	$g(0{,}62)$
F	0,66	1	No	$g(0{,}66)$
M	0,08	2	Yes	$g(0{,}08)$
J	0,10	2	Yes	$g(0{,}10)$
K	0,28	2	Yes	$g(0{,}28)$
L	0,35	2	Yes	$g(0{,}35)$
I	0,42	2	No	$g(0{,}42)$

Let us consider a list of T surveys, S_t is the sample selected for the survey t.

Let $U = \{1, ..., k, ...N\}$ be the population of firms.

Each unit k is given a permanent random number ω_k that is a realization of a random variable Ω_k with a uniform distribution on [0,1[. $\Omega_1, \Omega_2, ...\Omega_N$ are N independent random variables.

3.1 Generalization of the current method

For the survey t, we would like to define a coordination function $g_{k,t}$ for each unit k.

Such a function is expected to encourage the selection of k if it has not recently been selected.

For the survey t, the population of firms $U = \{1, ..., k, ...N\}$ is divided into nonoverlapping subpopulations $U_{1,t}, ...U_{h,t}, ..., U_{H,t}$.

Within stratum $U_{h,t}$, we would like to apply simple random sampling (with the fixed size $n_{h,t}$) with the values $g_{k,t}(\omega_k)$, $k \in U_{h,t}$.

Let us suppose that the converted numbers $g_{k,t}(\omega_k)$, $k \in U_{h,t}$ are independent realizations of a uniform distribution on $[0,1[$.

Then we just have to select the $n_{h,t}$ units that correspond to the $n_{h,t}$ smallest values $g_{k,t}(\omega_k)$, $k \in U_{h,t}$ to apply simple random sampling.

The random numbers $(\omega_k)_{k \in U}$ are independent realizations of a uniform distribution on [0,1[.

Consequently, in order to have the converted numbers $(g_{k,t}(\omega_k))_{k \in U}$ be independent realizations of a uniform on [0,1[, a coordination function $g_{k,t}$ is expected to save the uniform distribution on [0,1[.

3.1 Generalization of the current method

In the following part, we demonstrate that it is possible to define a coordination function $g_{k,t}$ that enables the unit k to be selected in priority if it has not recently been sampled.

We denote:

- $\gamma_{k,t}$: the response burden of the unit k for the survey t
- $\omega = (\omega_k)_{k \in U}$: the vector of the random numbers of the firms
- $I_{k,t}(\omega)$:the sample membership indicator of the unit k for the survey t

$$I_{k,t}(\omega) = \begin{cases} 1 & \text{if} \quad k \in S_t \\ 0 & \text{if not} \end{cases}$$

The selection of the unit k in the sample S_t depends on the value $g_{k,t}(\omega_k)$ but also on all the values $(g_{i,t}(\omega_i))_{i \neq k}$ insofar as we sort the values $g_{i,t}(\omega_i)$ to select the sample.

We denote:

- $\gamma_{k,t}(\omega) = \gamma_{k,t}.I_{k,t}(\omega)$ the effective response burden of the unit k
- $\Gamma_{k,t}(\omega) = \sum_{u \leq t} \gamma_{k,u}.I_{k,u}(\omega)$ the cumulative response burden of the unit k

$\gamma_{k,t}(\omega)$ and $\Gamma_{k,t}(\omega)$ are both random variables that depend on ω.

3.2 Estimation of the response burden

Before selecting the sample t, we can know to what extent the unit k has been requested in the past surveys by looking at its cumulative response burden $\Gamma_{k,t-1}(\omega)$: The lower it is, the less the unit has been requested in the past.

The coordination function $g_{k,t}$ must enable the unit k to be selected if it has not recently been sampled.

As seen previously, the unit k is more likely to be selected if the value $g_{k,t}(\omega_k)$ is low. Consequently, the lower the cumulative response burden is, the lower the value $g_{k,t}(\omega_k)$ must be.

In other words, we would like to have the following property:

$$\Gamma_{k,t-1}(\omega_1) < \Gamma_{k,t-1}(\omega_2) \Longrightarrow g_{k,t}(\omega_{k,1}) < g_{k,t}(\omega_{k,2})$$

where $\omega_{k,i} (i = 1, 2)$ is the kth component of the vector ω_i

3.2 Estimation of the response burden

The cumulative response burden $\Gamma_{k,t-1}(\omega)$ is not easy to handle:

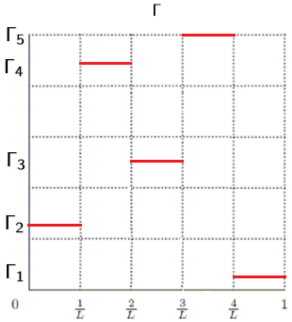
- It depends on all the random numbers $\omega = (\omega_k)_{k \in U}$ because of the sample membership indicators $I_{k,m}(\omega), m \in \{1, ..., t-1\}$.
- Its form is not easy to manipulate.

We can demonstrate that it is possible to approximate $I_{k,m}(\omega)$ by a constant piecewise function $I^{'}_{k,m}(\omega_k)$ that only depends on the kth component of ω.

Then the cumulative response burden $\Gamma_{k,t-1}(\omega)$ can be estimated by a constant piecewise function $\Gamma^{'}_{k,t-1}(\omega_k) = \sum_{u \leq t-1} \gamma_{k,u}.I^{'}_{k,u}(\omega_k)$ that only depends on ω_k.

We have to divide the interval [0,1[into sub intervals in order to define the constant piecewise function $\Gamma'_{k,t-1}$.

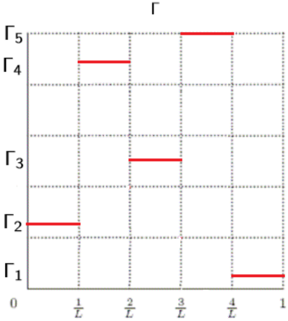
We decide to divide the interval [0,1[into L (≥ 100) sub intervals of equal length, $[\frac{l-1}{L}, \frac{l}{L}[$, $l = 1, ..., L$



3.3 Formulation of a coordination function

To simplify, labels k and t will be omitted.

Let us suppose that $L = 5$ and the constant piecewise approximation Γ' has the following form:
(to simplify, we suppose that all the values of Γ' are not the same).



3.3 Formulation of a coordination function

Let P^{Γ} be the probability distribution of the cumulative response burden Γ defined as :
For any interval I included in R, $P^{\Gamma}(I) \stackrel{def}{=} P[\Gamma^{-1}(I)]$.

Let F_{Γ} be the distribution function of the burden Γ.

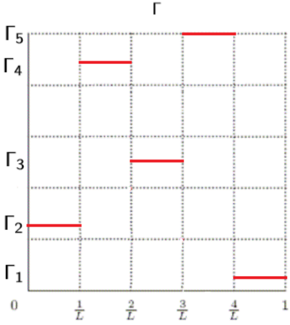
Let us define $G_{\Gamma} = F_{\Gamma}(\Gamma)$. Then we have:

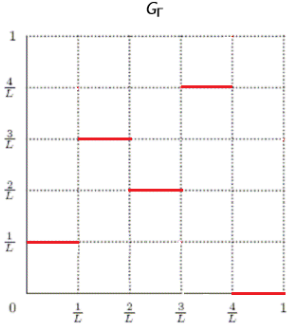
$$G_{\Gamma}(\omega) = P^{\Gamma}(]-\infty, \Gamma(\omega)[) = P(\Gamma^{-1}]-\infty, \Gamma(\omega)[)$$

$G_{\Gamma}(\omega)$ corresponds to the probability that a unit of the population has a burden lower than $\Gamma(\omega)$

Insee
Measuring, understanding

3.3 Formulation of a coordination function





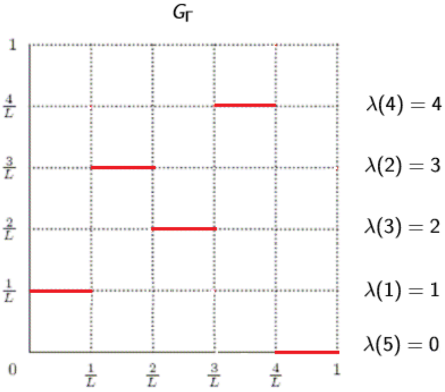
It is possible to calculate the values of G_Γ thanks to Γ.
Indeed, $G_\Gamma(\omega)$ corresponds to the probability that a unit has a burden lower than $\Gamma(\omega)$.

We have to look at the intervals that correspond to such units on the graph of Γ: $G_\Gamma(\omega)$ is equal to the sum of the lengths of these intervals.

3.3 Formulation of a coordination function

All the intervals $[\frac{l-1}{L}, \frac{l}{L}[$ have the same length of $\frac{1}{L}$. As a result, on the interval $[\frac{l-1}{L}, \frac{l}{L}[$, the value of G_Γ that we denote $G_\Gamma(l)$ is a multiple of $\frac{1}{L}$: we denote $G_\Gamma(l) = \frac{\lambda(l)}{L}$.

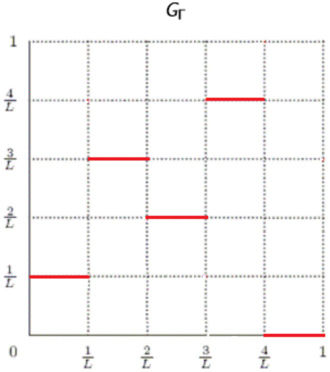
$\lambda(l)$ corresponds to the rank (between 0 and $L-1$) of the value $G_\Gamma(l)$.



Insee
Measuring, understanding

We supposed that all the values of Γ (so are the values of G_Γ) are not the same.

As a result, there is a permutation σ on $\{1,2,3,...,L\}$ that verifies:
$\lambda(l) = \sigma(l) - 1$



$\lambda(4) = 4 \quad \sigma(4) = 5$

$\lambda(2) = 3 \quad \sigma(2) = 4$

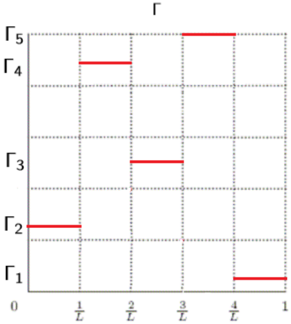
$\lambda(3) = 2 \quad \sigma(3) = 3$

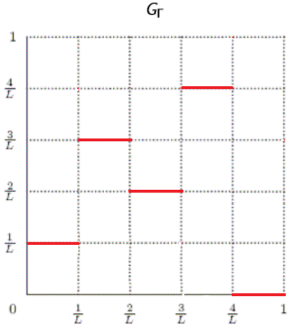
$\lambda(1) = 1 \quad \sigma(1) = 2$

$\lambda(5) = 0 \quad \sigma(5) = 1$

Insee
Measuring, understanding

3.3 Formulation of a coordination function

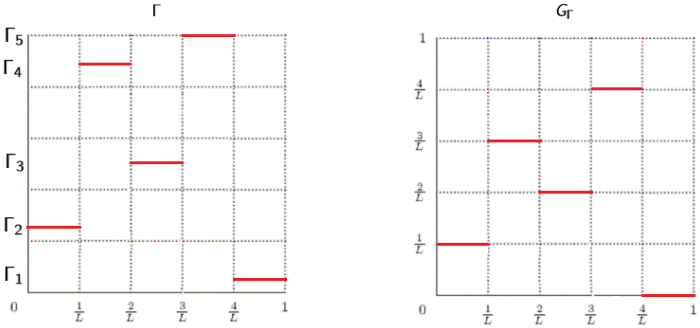




The function G_Γ has interesting properties:

- $G_\Gamma([0,1[) \subset [0,1[$
- G_Γ verifies the property: $\Gamma(\omega_1) < \Gamma(\omega_2) \Longrightarrow G_\Gamma(\omega_1) < G_\Gamma(\omega_2)$

Insee
Measuring, understanding



G_Γ can't be a coordination function insofar as it doesn't save the uniform distribution on [0,1[. Indeed, it only has L distinct values. We would like to transform G_Γ into a coordination function.

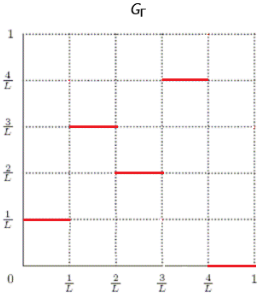
We replace every level of G_Γ with a linear function with a slope of 1.

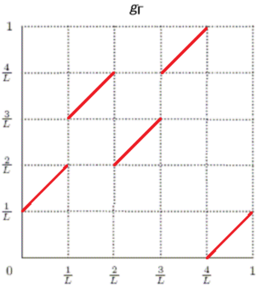
On the interval $[\frac{l-1}{L}, \frac{l}{L}[$, we define g_Γ as :

$$g_\Gamma(\omega) = G_\Gamma(l) + (\omega - \frac{l-1}{L}) = \frac{\sigma(l)-1}{L} + (\omega - \frac{l-1}{L})$$

Then, we have:

$$g_\Gamma(\frac{l-1}{L}) = G_\Gamma(l) = \frac{\sigma(l)-1}{L} \quad \text{and} \quad g_\Gamma(\frac{l}{L}) = G_\Gamma(l) + \frac{1}{L} = \frac{\sigma(l)-1}{L} + \frac{1}{L}$$

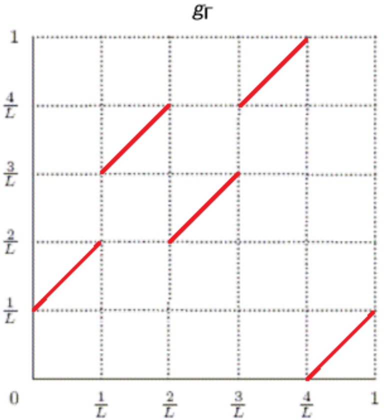




3.3 Formulation of a coordination function

The coordination function g_Γ is totally based on a permutation σ on {1,2,3,...,L}. It is defined as:

$$\forall \omega \in [\frac{l-1}{L}; \frac{l}{L}[, \; g_\sigma(\omega) = \frac{\sigma(l)-1}{L} + (\omega - \frac{l-1}{L})$$



First sample S_1:

We initialize the cumulative response burden of each unit k with zero:
$\forall\, k \in U,\ \Gamma_{k,0}(\omega_k) = 0$

There is no coordination to realize. Within stratum $U_{h,1}$, we select the $n_{h,1}$ units that correspond to the $n_{h,1}$ smallest random numbers ω_k, $k \in U_{h,1}$.

It is a particular case of coordination with $g_{k,1} = Id$ for every unit k.
The expected cumulative response burden that only depends on ω_k is
$\Gamma^{'}_{k,1}(\omega_k) = \gamma_{k,1} I^{'}_{k,1}(\omega_k)$.

Insee
Measuring, understanding

Second sample S_2:

We formulate a coordination function $g_{k,2}$ based on the constant piecewise function $\Gamma'_{k,1}$ in order to select the second sample S_2.

Within stratum $U_{h,2}$, we select the $n_{h,2}$ units that correspond to the $n_{h,2}$ smallest values $g_{k,2}(\omega_k)$, $k \in U_{h,2}$.

The expected cumulative response burden is
$\Gamma'_{k,2}(\omega_k) = \Gamma'_{k,1}(\omega_k) + \gamma_{k,2}.I'_{k,2}(\omega_k)$.

Sample S_t:

More generally, we formulate a coordination function $g_{k,t}$ based on the constant piecewise function $\Gamma'_{k,t-1}$.

Within stratum $U_{h,t}$, we select the $n_{h,t}$ units that correspond to the $n_{h,t}$ smallest values $g_{k,t}(\omega_k)$, $k \in U_{h,t}$.

The expected cumulative response burden is
$\Gamma'_{k,t}(\omega_k) = \Gamma'_{k,t-1}(\omega_k) + \gamma_{k,t}.I'_{k,t}(\omega_k)$.

Insee
Measuring, understanding

Plan

We realized a simulation of coordinated sampling. We considered a population of length 100. We successively selected 20 samples by simple random sampling.

Every sample but the third has a length of 25. The third has a length of 50.

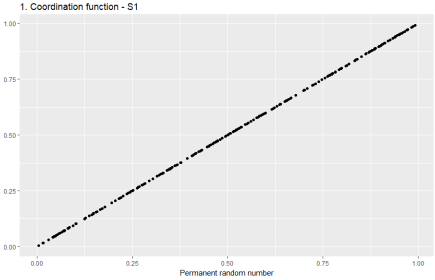
The response burden γ is equal to 1 for every unit and for every sample but the third.

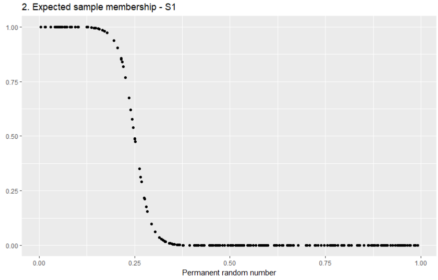
For the third sample, the response burden γ is equal to 3.

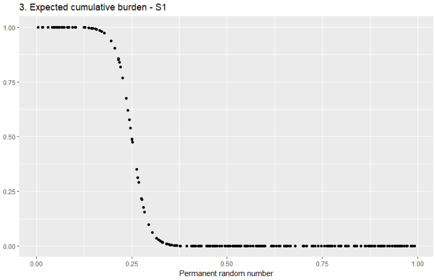
Such a case is very simple insofar as there is no strata and all the units have the same coordination function and the same cumulative response burden.

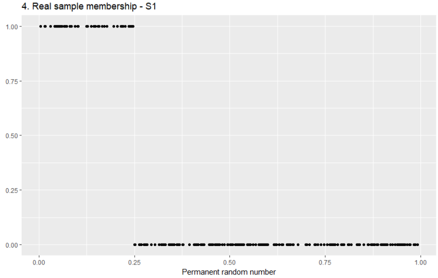
We will analyze the coordination function g_t, the expected sample membership indicator $\mathbb{E}\left[I_{k,t}(\Omega)\,|\Omega_k = \omega_k\right]$, the expected cumulative response burden and the real sample membership indicator for the first four samples and the sample 20.

First sample S_1

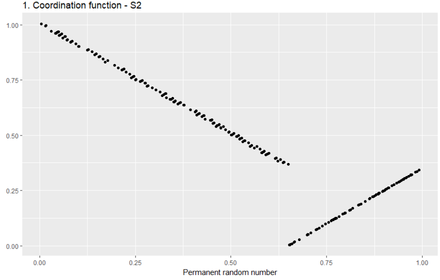


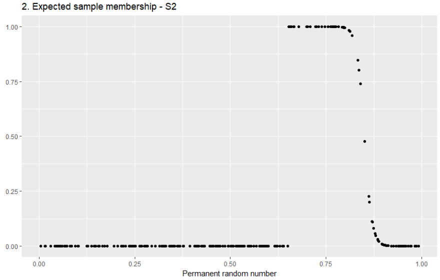


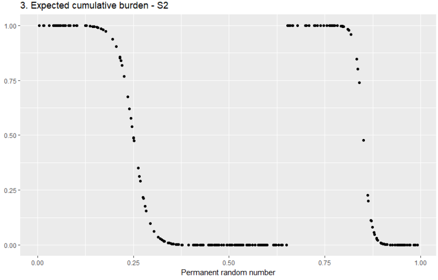


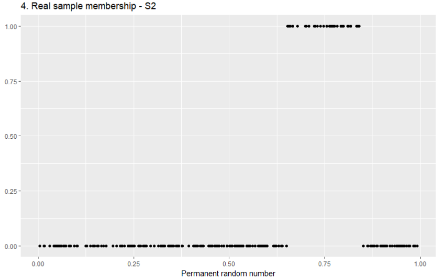


Second sample S_2

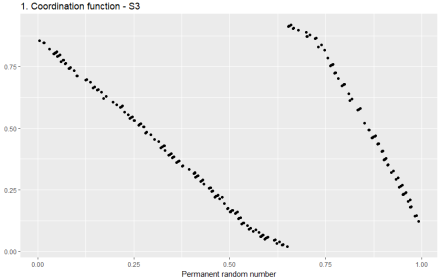


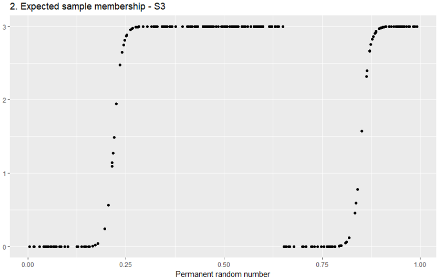


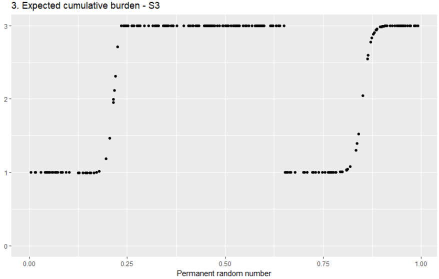


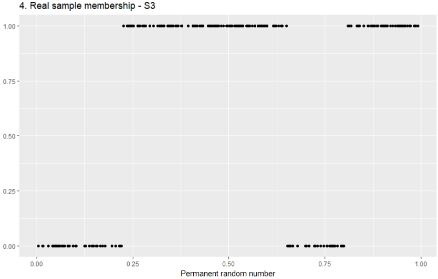


Third sample S_3

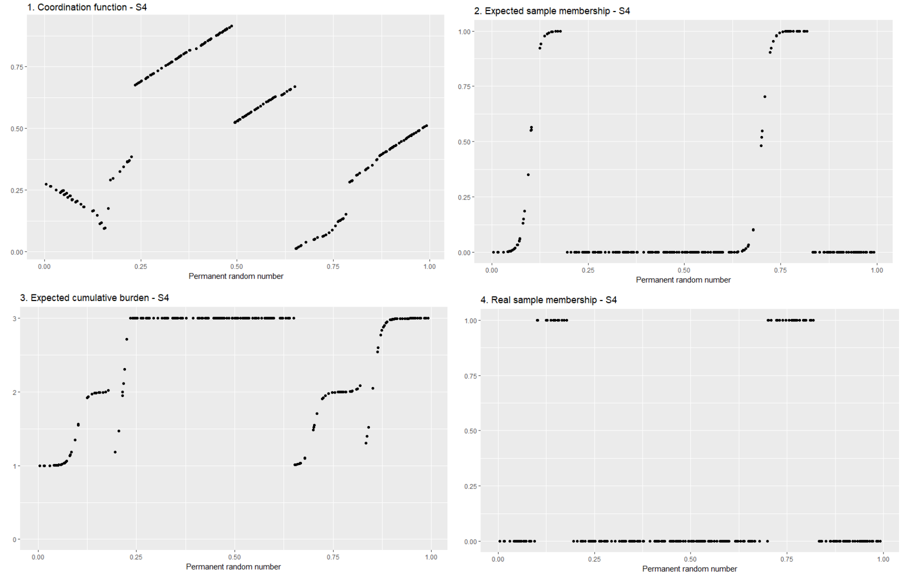




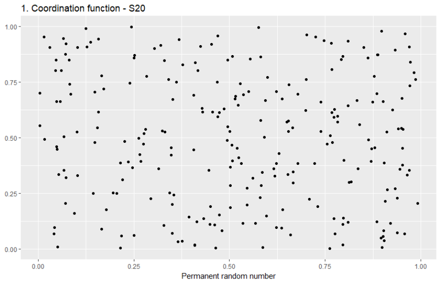


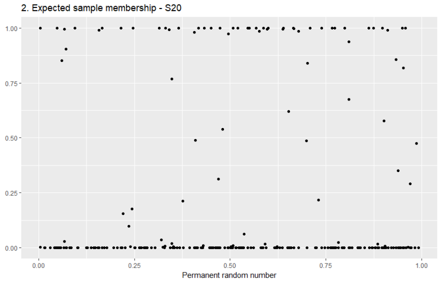


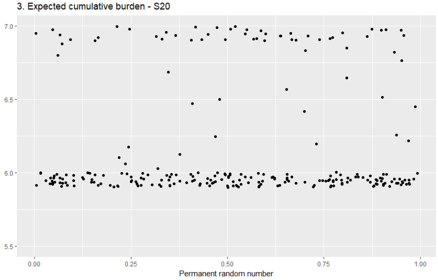
Fourth sample S_4

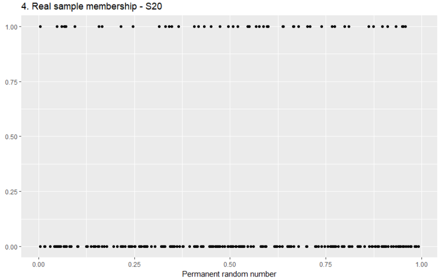


Twentieth sample S_{20}









Plan

There are two principal steps to select a sample S_t using negative coordination. First, we calculate the cumulative response burden of each unit k. Secondly, we calculate the coordination function $g_{k,t}$.

Calculation of $\Gamma'_{k,t-1}$

For each sample S_m, $m \in \{1, 2, ..., t-1\}$, a coordination table exists with the following columns:

- the ID of the unit
- the stratum $U_{h,m}$ the unit belongs to
- the size $N_{h,m}$ of the stratum
- the size of the sample $n_{h,m}$
- the L values of the permutation $\sigma_{k,m}$

.

	sigma 1	sigma 2	sigma 3	ID	Nh	nech	Stratum
1	96	97	100	000325175	1540	26	3212Z00E
2	80	81	84	005420021	734	110	4669B11E
3	96	97	100	005450093	17182	252	4778C00E

Measuring, understanding

Thanks to the L values of the permutation $\sigma_{k,m}$, we calculate the L values of the function $I^{'}_{k,m}$.

Then we concatenate the t-1 coordination tables. Then, for the unit k, we calculate the sum $\sum_{m\leq t-1}\gamma_{k,m}.I^{'}_{k,m}(\omega_k)$ to estimate the cumulative response burden $\Gamma_{k,t-1}$.

Calculation of $g_{k,t}$

We create a matrix with the following columns:

- the random number ω_k
- the stratum $U_{h,t}$
- the size $N_{h,t}$ of the stratum
- the size of the sample $n_{h,t}$
- the L values of the function $\Gamma'_{k,t-1}$ we estimated before

	ω_k	$U_{h,t}$	$N_{h,t}$	$n_{h,t}$	$\Gamma'_{k,t-1}(1)$	$\Gamma'_{k,t-1}(2)$	$\Gamma'_{k,t-1}(3)$
1	0,66	1	500	250	0,29	0,70	0,54
2	0,35	1	500	250	0,08	0,82	0,24
3	0,08	2	300	150	0,49	0,31	0,12

5 Algorithm for coordinated sampling

We replace the line $[\Gamma'_{k,t-1}(1)\ \ \Gamma'_{k,t-1}(2)\ \ ...\Gamma'_{k,t-1}(L)]$ with the line that contains the rank of each value $\Gamma'_{k,t-1}(l)$. Such a line corresponds to the different values of $\sigma_{k,t}(l)$.

	ω_k	$U_{h,t}$	$N_{h,t}$	$n_{h,t}$	$\Gamma'_{k,t-1}(1)$	$\Gamma'_{k,t-1}(2)$	$\Gamma'_{k,t-1}(3)$
1	0,66	1	500	250	0,29	0,70	0,54
2	0,35	1	500	250	0,08	0,82	0,24
3	0,08	2	300	150	0,49	0,31	0,12

⇩

	ω_k	$U_{h,t}$	$N_{h,t}$	$n_{h,t}$	$\Gamma'_{k,t-1}(1)$	$\Gamma'_{k,t-1}(2)$	$\Gamma'_{k,t-1}(3)$
1	0,66	1	500	250	1	3	2
2	0,35	1	500	250	1	3	2
3	0,08	2	300	150	3	2	1

Then we calculate the value $g_{k,t}(\omega_k)$ thanks to the definition:
$\forall \omega \in [\frac{l-1}{L}; \frac{l}{L}[, \; g_{k,t}(\omega) = \frac{\sigma_{k,t}(l)-1}{L} + (\omega - \frac{l-1}{L})$

Then we select the sample S_t depending on the values $(g_{k,t}(\omega_k))_{k \in U}$.

Finally, we create a coordination table that contains the ID of the unit, the stratum $U_{h,t}$ it belongs to, the size $N_{h,t}$ of the stratum, the size of the sample $n_{h,t}$ and the L values of the permutation $\sigma_{k,t}$ that corresponds to the coordination function $g_{k,t}$.

Plan

- We have used coordination to select samples of firms for five years.
- The aim of negative coordination is to reduce the burden of the firms while getting unbiased samples.
- The general method of coordination enables to coordinate as many samples as we want. What is more, it is possible to weight the samples according to the burden of the surveys.

References

Emmanuel Gros, Ronan Le Gleut
Sample coordination and response burden for business surveys: methodology and practice of the procedure implemented at Insee
The Unit Problem and Other Current Topics in Business Survey Methodology, page 139, Cambridge Scholars Publishing, 2018

Fabien Guggemos, Olivier Sautory
Sampling coordination of business surveys conducted by insee
Proceedings of the Fourth International Conference of Establishment Surveys, 2012

Thank you for your attention !

Appendix: Property of a coordination function

The random numbers $(\omega_k)_{k \in U}$ are independent realizations of a uniform distribution on [0,1[.

Consequently, in order to have the converted numbers $(g_{k,t}(\omega_k))_{k \in U}$ be independent realizations of a uniform on [0,1[, a coordination function $g_{k,t}$ is expected to save the uniform distribution on [0,1[:

if P is the uniform distribution on [0,1[then $P^{g_{k,t}}$ is the uniform distribution too.
For any interval I=[a,b[included in [0,1[,
$P^{g_{k,t}}(I) \stackrel{def}{=} P[g_{k,t}^{-1}(I)] = P(I) = b - a$.

Insee
Measuring, understanding

The best approximation (within the meaning of the L_2 norm) of the sample membership indicator $I_{k,m}(\omega)$ of unit k that only depends on ω_k is the conditional expectation:
$I'_{k,m}(\omega_k) = \mathbb{E}\left[I_{k,m}(\Omega)\,|\Omega_k = \omega_k\right] = P(k \in S_m|\Omega_k = \omega_k)$

If we suppose that coordination functions are bijectives then we have:

$$I'_{k,m}(\omega_k) = P(k \in S_m | \Omega_k = \omega_k) = P(k \in S_m | g_{k,m}(\Omega_k) = g_{k,m}(\omega_k))$$

$I'_{k,m}(\omega_k)$ can be seen as a function of $g_{k,m}(\omega_k)$ that we denote $b_{k,m}$.

$I'_{k,m}(\omega_k) = b_{k,m}(g_{k,m}(\omega_k))$ where $b_{k,m}(x) = P(k \in S_m | g_{k,m}(\Omega_k) = x)$

Insee
Measuring, understanding

$(g_{k,m}(\Omega_k))_{k \in U}$ are N independent realizations of the uniform distribution on [0,1[. As a result, $b_{k,m}(x)$ corresponds to the probability of the following event:
Among the N-1 values $g_{i,m}(\Omega_i)$ $(i \neq k)$, *there are at most n-1 values that are below x.*

In other words, the event $(k \in S_m | g_{k,m}(\Omega_k) = x)$ equates to the following event:
Let $X_1, X_2, ..., X_N$ *be N independent realizations of the uniform distribution on [0,1[, the n th random variable* $X_{(n)}$ *is higher than x*

What is more, it is a well known fact that $X_{(n)}$ has a beta distribution with the parameters n and N-n+1. Then we have:

$$b_{k,m}(x) = 1 - P(X_{(n)} \leq x) = 1 - \frac{1}{B(n,N-n+1)} \int_0^x u^{n-1}(1-u)^{N-n}\,\mathrm{d}u$$

where $B(n, N-n+1) = \frac{(n-1)!(N-n)!}{N!}$

If we know the coordination functions $g_{k,1}, ..., g_{k,t-1}$ for the surveys $1, ..., t-1$ then we can calculate $I'_{k,1}(\omega_k) = b_{k,1}(g_{k,1}(\omega_k)), ..., I'_{k,t-1}(\omega_k) = b_{k,t-1}(g_{k,t-1}(\omega_k))$ thanks to the beta distribution.

There are three steps to built an constant piecewise function that estimates $I'_{k,m}(\omega_k) = b_{k,m}(g_{k,m}(\omega_k))$.

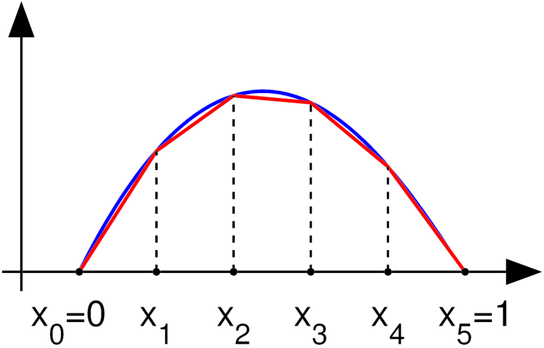
- The interval [0,1[will be divided into L (that is supposed to be an integer greater than 100) sub intervals of equal length, $[\frac{l-1}{L}, \frac{l}{L}[, l = 1, ..., L$

Appendix: Estimation of the response burden

- We replace the function $b_{k,m}$ with a linear piecewise function $\tilde{b}_{k,m}$ on each interval $[\frac{l-1}{L}, \frac{l}{L}[$

$$\forall x \in [\frac{l-1}{L}; \frac{l}{L}[, \ b_{k,m}^{\sim}(x) = L.(b_{k,m}(\frac{l}{L}) - b_{k,m}(\frac{l-1}{L})).(x - \frac{l}{L}) + b_{k,m}(\frac{l}{L})$$

We notice that $b_{k,m}(\frac{l-1}{L}) = b_{k,m}^{\sim}(\frac{l-1}{L})$ and $b_{k,m}(\frac{l}{L}) = b_{k,m}^{\sim}(\frac{l}{L})$



Insee
Measuring, understanding

- On each interval $[\frac{l-1}{L}, \frac{l}{L}[$, we approximate $I'_{k,m}(\omega_k) = b_{k,m}(g_{k,m}(\omega_k))$ by the mean value of the function $\tilde{b}_{k,m} \circ g_{k,m}$

  $$\beta_{k,m}(l) = \frac{1}{1/L} \cdot \int_{\frac{l-1}{L}}^{\frac{l}{L}} \tilde{b_{k,m}}(g_{k,m}(\omega))\mathrm{d}\omega$$

  The function $\beta_{k,m}$ defined as:

  $$\forall \omega \in [\frac{l-1}{L}; \frac{l}{L}[, \beta_{k,m}(\omega) = \beta_{k,m}(l)$$

  is a constant piecewise estimation of the function $I'_{k,m}(\omega_k)$.
  As a result, $\Gamma'_{k,t-1}(\omega_k) = \sum_{u \leq t-1} \gamma_{k,u} . \beta_{k,u}(\omega_k)$ is a constant piecewise estimation of $\Gamma_{k,t-1}(\omega)$.

Appendix: Coordinated sampling with different types of units

ID of the firm	Random number of the firm
A	0,75
C	0,06
G	0,82
H	0,55
I	0,32
J	0,03
K	0,67

Population for the survey *t-1*

ID of the firm	Random number of the firm
A	0,75
B	0,23
C	0,06
D	0,36
E	0,12
F	0,47
G	0,82
H	0,55
I	0,32

Population for the survey *t*

ID of the firm	Random number of the firm
A	0,75
C	0,06
G	0,82
H	0,55
I	0,32

Coordination table for the survey *t-1*

Appendix: Coordinated sampling with different types of units

ID of the local firm	Random number of the local firm
A-2	0,11
A-3	0,79
C-1	0,36
C-2	0,40
D-1	0,23
D-2	0,60
D-3	0,03

Population for the survey *t-1*

ID of the firm	ID of the principal local firm	Random number of the firm
A	A-3	0,79
B	B-1	0,33
C	C-2	0,40
D	D-1	0,23
E	E-1	0,86

Population for the survey *t*

ID of the principal local firm	Random number
A-3	0,79
C-2	0,40
D-1	0,23

Coordination table for the survey *t-1*