

# COORDINATION DES ENQUÊTES AUPRÈS DES ENTREPRISES

Paul-André Salamin<sup>1</sup>

## RÉSUMÉ

On décrit un algorithme de répartition de la charge qui a été développé dans le cadre de la révision du registre des entreprises et des établissements de l'Office fédéral de la statistique (OFS). Alors que la plupart des systèmes de coordination des enquêtes utilisent des nombres aléatoires permanents, dans l'approche suivie à l'OFS la coordination est réalisée uniquement par une mesure de la charge cumulée. Les distributions marginales des échantillons générés par l'algorithme de répartition de la charge sont celles du sondage aléatoire simple. De plus, l'algorithme est stable par rapport aux perturbations provoquées par une mise à jour du registre.

MOTS CLÉS : Coordination des enquêtes; Charge; Echantillonnage.

## ABSTRACT

We describe a co-ordination algorithm which has been developed within the framework of the revision of the business register of the Swiss Federal Statistical Office (SFSO). Whereas most systems for survey co-ordination use permanent random numbers, in the approach followed at the SFSO the spread of the response burden is achieved using only a measure of cumulative burden. The marginal distributions of the samples generated by the co-ordination algorithm are the same as for simple random sampling. Furthermore, the algorithm is stable with respect to the perturbations due to the up-dates of the register.

KEY WORDS: Survey co-ordination; Response burden; Sampling.

## 1. INTRODUCTION

L'Office fédéral de la statistique (OFS) a entamé en 1997 une révision complète de son Registre des entreprises et des établissements (REE). Dans le cadre de ce projet, il est prévu de développer un système de gestion des échantillons qui doit satisfaire à deux exigences principales : mettre à disposition des utilisateurs du système des outils de sondage flexibles et utiliser des algorithmes de sondage qui répartissent uniformément la charge, définie ici comme le nombre de fois qu'une unité a dû participer à une enquête.

Pour une suite de sondages indépendants, la charge n'est pas répartie uniformément puisque le recouvrement des échantillons est aléatoire. On parle de sondage coordonné lorsque l'on introduit une dépendance entre les échantillons. On distingue la coordination négative de la coordination positive, selon que l'on veut éviter ou forcer un recouvrement

des échantillons. Il existe essentiellement deux types d'approche pour réaliser une coordination des échantillons (Ohlsson, 1995) : on peut utiliser des nombres aléatoires permanents associés aux unités du registre, ou on peut utiliser une mesure de la charge cumulée des unités, ce qui est l'approche suivie à l'OFS.

On décrit dans la section 2 l'algorithme de coordination négative du système de gestion des échantillons du REE. On montre que les distributions marginales des échantillons générés par l'algorithme sont celles du sondage aléatoire simple. Dans la section 3, on présente le traitement des nouvelles unités. L'apparition de nouvelles unités lors de la mise à jour du registre provoque une perturbation du processus de sondage qui, pour une suite de sondages effectués selon l'algorithme de répartition de la charge du REE, converge vers zéro.

---

<sup>1</sup> Paul-André Salamin, Office fédéral de la statistique, Espace de l'Europe 10, CH-2010 Neuchâtel, Suisse, paul-andre.salamin@bfs.admin.ch.

## 2. ALGORITHME DE RÉPARTITION DE LA CHARGE

L'algorithme est similaire à celui utilisé à Statistics Netherlands (de Ree, 1999). La charge initiale de chaque unité est égale à 0. Le tirage d'un échantillon de taille  $n$  s'effectue selon l'algorithme suivant.

- On génère, indépendamment pour chaque unité, un nombre aléatoire uniformément distribué entre 0 et 1.
- On ordonne les unités par charge croissante et, en cas d'égalité de la charge, par nombre aléatoire croissant.
- On sélectionne les  $n$  premières unités.
- On augmente de 1 la charge des unités sélectionnées.

Une répartition globale de la charge conduirait à des biais. Prenons comme exemple une population, composée de deux branches économiques, dans laquelle on tirerait un échantillon exhaustif dans la première branche économique, puis un échantillon non-exhaustif dans l'ensemble de la population. Comme on accumule la charge et qu'on ordonne les unités par charge croissante avant un tirage, les unités de la première branche économique seront sous-représentées dans le second échantillon.

Pour réaliser une répartition locale de la charge on décompose le registre en micro-strates, définies par exemple en fonction de l'activité économique et d'une mesure de la taille des unités du registre. La décomposition du registre en micro-strates est fixe et le tirage d'un échantillon se fait au niveau des micro-strates. Les micro-strates doivent permettre de construire toutes les strates comme combinaisons des micro-strates. Le même argument que plus haut montre qu'on ne peut pas former des strates qui seraient plus petites qu'une micro-strate.

Comme un échantillon est tiré autant que possible parmi les unités qui ont la charge la plus faible, seules importent les différences entre les charges des unités d'une micro-strate. Avec notre définition de la charge, ces différences sont au plus de 1. En effet, comme les charges initiales sont égales à 0, nous avons les charges 0 ou 1 après le tirage du premier échantillon. Pour les tirages suivants, la seule façon d'avoir plus que deux valeurs consécutives de la charge serait de tirer un échantillon plus grand que la micro-strate, ce qui est bien sûr impossible. On peut décrire l'algorithme en se référant uniquement aux classes d'unités de même charge. Comme on l'a vu, ces classes sont au nombre de 2 au plus. Nous définissons  $G0$  comme l'ensemble des unités de charge minimale dans une micro-strate  $G$ . Soient  $M$  la taille de  $G0$  et  $n$  la

taille de l'échantillon. On peut décrire l'algorithme de répartition de la charge dans une micro-strate  $G$  comme suit.

- Si  $1 \leq n < M$ , on tire un échantillon aléatoire simple dans  $G0$ .
- Si  $M \leq n \leq N$ , l'échantillon est formé de  $G0$  complété par un échantillon aléatoire simple de taille  $M - n$  tiré dans le complémentaire de  $G0$ .

En examinant comment  $G0$  change après chaque tirage d'un échantillon, on peut montrer que  $G0$  est un échantillon aléatoire simple (EAS) de la micro-strate  $G$ . Alors, un échantillon obtenu par l'algorithme de répartition de la charge est soit un EAS d'un EAS, soit un EAS complété par un EAS tiré dans son complémentaire, d'où il suit que cet échantillon est aussi un échantillon aléatoire simple de  $G$ . On peut aussi utiliser ce résultat comme une condition sur l'algorithme de répartition de la charge : pour que les échantillons tirés dans une micro-strate  $G$  soient des échantillons aléatoires simples de  $G$ , il faut que  $G0$ , l'ensemble des unités de charge minimale, soit lui-même un échantillon aléatoire simple de  $G$ .

En général, une strate est une combinaison de micro-strates. L'ensemble des unités de charge minimale d'une strate est donné par  $G0\_strate = G0\_1 + \dots + G0\_H$ , où  $H$  est le nombre de micro-strates qui composent la strate. Comme  $G0\_strate$  n'est pas un échantillon aléatoire simple de la strate, on ne peut pas, au risque d'introduire des biais qui peuvent être importants, appliquer l'algorithme de répartition de la charge à  $G0\_strate$ . Dans le système de gestion des échantillons du REE, on a résolu ce problème en introduisant les micro-strates comme un niveau de stratification supplémentaire. L'allocation aux micro-strates est proportionnelle, ce qui donne les mêmes probabilités d'inclusion de premier ordre que pour un sondage aléatoire simple.

## 3. MISE À JOUR DU REGISTRE

La composition des micro-strates change lors d'une mise à jour du registre. Ces changements peuvent être dus aux migrations entre les micro-strates et aux créations ou disparitions d'unités. Les nouvelles unités nécessitent un traitement spécial. En effet, comme les échantillons sont tirés autant que possible parmi les unités qui ont la charge la plus faible, on ne peut pas simplement attribuer la même charge à toutes les nouvelles unités.

On note  $G(t1)$  et  $G(t2)$ ,  $t1 < t2$ , les états d'une micro-strate immédiatement avant et après une mise à jour du registre. Pour simplifier les notations, on suppose qu'il n'y a pas de disparition d'unité. Nous pouvons alors

écrire  $G(t2) = G(t1) + \Delta G$ , où  $\Delta G$  est l'ensemble des nouvelles unités (créations ou immigrations) de la micro-strate  $G$ . Soit  $G0(t1)$  l'ensemble des unités de charge minimale de  $G(t1)$ . Une charge est attribuée aux nouvelles unités par une allocation proportionnelle de  $\Delta G$  aux classes  $G0(t1)$  et  $G(t1) \setminus G0(t1)$ . Après la mise à jour du registre, l'ensemble des unités de charge minimale est donné par  $G0(t2) = G0(t1) + \Delta G0$ . On a montré dans la section 2 que  $G0(t1)$  est un échantillon aléatoire simple de  $G(t1)$ . Il suit que  $G0(t2)$  est un échantillon stratifié de  $G(t2)$ . Comme l'allocation est proportionnelle, on a les mêmes probabilités d'inclusion d'ordre 1 que si  $G0(t2)$  était un échantillon aléatoire simple de  $G(t2)$ .

Partant de l'algorithme de répartition de la charge, on peut définir une chaîne de Markov qui permet de calculer l'évolution, au fil des sondages, des probabilités d'inclusion d'ordre 2 liées à  $G0$ . On peut exprimer les probabilités d'inclusion d'ordre 2 après la mise à jour du registre comme une perturbation des probabilités d'inclusion d'ordre 2 pour un sondage aléatoire simple, qui sont celles que l'algorithme "voit".

La chaîne de Markov déduite de l'algorithme de répartition permet de montrer que cette perturbation converge vers zéro.

#### 4. CONCLUSIONS

On a décrit un algorithme de coordination négative pour le sondage aléatoire simple, stable par rapport aux perturbations provoquées par une mise à jour du registre. L'approche présentée ici peut être adaptée à la coordination négative ou positive pour le sondage de Bernoulli et peut donc aussi être utilisée pour la gestion des panels.

#### RÉFÉRENCES

- de Ree, J. (1999). "Co-ordination of Business Samples using Measured Response Burden". *Bulletin of the International Statistical Institute*, 52nd session of ISI, Proceedings Tome LVIII, Book 2, Invited Topic 64, 289-292.
- Ohlsson, E. (1995). "Coordination of Samples Using Permanent Random Numbers", Chapter 9 in *Business Survey Methods*, edited by Cox, Binder, Chinnappa, Colledge, and Kott. New York: John Wiley.