

# La coordination d'échantillons d'enquêtes auprès des entreprises mise en place à l'Insee

*Fabien GUGGEMOS<sup>1</sup>, Olivier SAUTORY<sup>2</sup>*

La question de la coordination des échantillons d'enquêtes entreprises a donné lieu à de nombreux travaux théoriques. Elle est souvent vue comme une méthode pour réduire la charge statistique (coordination négative), mais elle permet aussi d'obtenir un recouvrement entre échantillons pour les exploiter de façon conjointe, ce qui est par exemple le cas des panels renouvelés par partie (coordination positive). Concernant la charge statistique, ces méthodes sont destinées à traiter principalement le cas des petites entreprises, car les grandes entreprises sont en général incluses d'office dans les échantillons de la plupart des enquêtes.

La méthode actuellement utilisée à l'Insee pour la coordination négative est fondée sur l'utilisation de nombres aléatoire, tirés selon la loi uniforme sur  $[0, 1[$ . Dans une strate donnée, les  $n$  unités ayant les plus petits numéros aléatoires seront sélectionnés. À l'issue du tirage, une rotation des numéros aléatoires est opérée, qui donne ainsi aux unités qui viennent d'être tirées une probabilité plus faible d'être sélectionnées dans le tirage suivant, tout en assurant de bonnes propriétés aux estimateurs statistiques classiques servant à produire les résultats. D'autres schémas sont utilisés pour assurer une coordination positive. Ces méthodes ne sont utilisées que pour coordonner deux à deux certains échantillons.

La pression politique de plus en plus forte sur la question de la charge conduit à revoir le dispositif existant, afin de prendre en compte l'ensemble des enquêtes de la statistique publique : il s'agit d'intégrer la dimension "charge cumulée d'enquêtes sur les dernières années", cette charge cumulée pouvant tenir compte de pondérations différentes attribuées aux enquêtes, selon la longueur du questionnaire d'enquête par exemple.

Ch. Hesse (document de travail E0101, Insee 2001) a proposé une méthode qui généralise la technique actuelle : à chaque unité sont associées un nombre aléatoire initial appartenant à  $[0, 1[$  et une "fonction de coordination". Cette fonction, qui transforme les nombres aléatoires, a la propriété de préserver la densité uniforme sur  $[0, 1[$ . Elle change à chaque tirage d'échantillon, et dépend du type de coordination souhaité. Cette méthode conduit à un processus automatique de sélection d'échantillons séparés, ou de d'échantillons partiellement renouvelés (mise à jour de panels), qui prend en compte la charge cumulée de réponse sur plusieurs enquêtes.

Cette méthode pourra être opérationnelle à l'Insee après la mise en place d'un répertoire statistique contenant pour chaque enquête les caractéristiques du plan de sondage (liste des unités appartenant au champ de l'enquête, liste des strates, nombre d'unités sélectionnées par strate).

La communication présentera les premières expérimentations de cette méthode.

---

<sup>1</sup> Insee ([fabien.guggemos@insee.fr](mailto:fabien.guggemos@insee.fr))

<sup>2</sup> Insee ([olivier.sautory@insee.fr](mailto:olivier.sautory@insee.fr))