
ÉCHANTILLONNAGE SPATIAL : L'ÉTAT DE L'ART

Yves TILLÉ

Université de Neuchâtel, Institut de Statistique

yves.tille@unine.ch

Mots-clés : auto-corrélation, échantillonnage, étalement, équilibrage, probabilités d'inclusion

Résumé

L'échantillonnage spatial est particulièrement important pour la statistique environnementale (voir entre autres Marker et Stevens Jr., 2009; Thompson, 2012). Cependant son champ d'application peut être étendu aux cas où la distance n'est pas géographique. Par exemple, on peut calculer des distances entre des entreprises en se basant sur leurs caractéristiques comme le chiffre d'affaires, le bénéfice ou le nombre de travailleurs. Lorsque deux unités sont proches, elles sont en général similaires, ce qui induit une dépendance spatiale entre les unités. Si l'on sélectionne deux unités proches dans l'échantillon, on aura tendance à récolter une information partiellement redondante. Un simple raisonnement développé dans Grafström et Tillé (2013) montre que sous un modèle linéaire auto-corrélé, il est plus efficace de sélectionner un échantillon bien étalé dans l'espace. Grafström et Lundström (2013) discutent longuement du concept d'étalement appelé aussi équilibrage spatial et de l'implication sur l'estimation. En particulier, ils montrent sous certaines hypothèses qu'un plan d'échantillonnage bien étalé est approprié quand la population est régie par un modèle corrélé.

Le plan de sondage par pavage randomisé stratifié généralisé (Generalized Random Tessellation Sampling GRTS) a été proposée par Stevens Jr. et Olsen (1999, 2004, 2003) pour sélectionner des échantillons étalés. La méthode est basée sur la construction d'une grille sur l'espace. Les cellules de la grille doivent être suffisamment petites pour qu'il y ait au maximum une unité dans chaque case. Les cellules sont ensuite ordonnées de manière aléatoire de sorte que les relations de proximité soient préservées. Ensuite, un échantillonnage systématique est appliqué le long des cellules ordonnées.

La méthode du pivot a été proposée par Deville et Tillé (2000). L'algorithme commence avec un vecteur de probabilités d'inclusion. A chaque étape, seulement deux composantes de ce vecteur sont modifiées aléatoirement en respectant en espérance les probabilités d'inclusion initiales et la taille fixe. Grafström et al. (2012) ont proposé d'utiliser la méthode du pivot pour l'échantillonnage spatial. Cette méthode, appelée méthode du pivot local, consiste, à chaque étape, à confronter deux unités voisines. Si la probabilité de l'une de ces deux unités est augmentée, la probabilité de l'autre est diminuée, ce qui à son tour induit une certaine répulsion entre les unités. L'extension naturelle de cette idée consiste à faire se confronter un groupe d'unités. La méthode du pivot locale a été généralisée par Grafström et Tillé (2013) pour fournir un échantillon à la fois bien étalé dans l'espace et équilibré sur des variables auxiliaires. Cette méthode, appelée méthode du cube local, consiste à lancer la phase de vol de la méthode du cube sur un sous-ensemble d'unités voisines.

Un polygone de Voronoï est l'ensemble des éléments de la population qui sont plus proches d'un point donné que tous les autres points de la population. Stevens Jr. et Olsen (2004) ont proposé d'utiliser la variance de la somme des probabilités d'inclusion des unités de population incluses dans les polygones de Voronoï autour des points sélectionnés comme un indicateur de la qualité de l'étalement de l'échantillon. Malheureusement, cet indicateur a des limites. Il est en effet possible de

construire des échantillons dont les unités sont toutes agrégées pour lesquels la variance des polygones de Voronoï est nulle. Tillé et al. (2018) ont proposé une autre mesure d'étalement basée sur la corrélation entre chaque point et son voisinage. Cette mesure peut être positive ou négative et vaut -1 en cas d'étalement parfait, c'est-à-dire quand l'échantillon est périodique.

Aucune des méthodes proposées ne permet d'obtenir un plan composé d'échantillons périodiques quand on sait que ce plan existe. Les échantillons périodiques peuvent être construits sur des grilles quand les inverses des probabilités sont constants et égaux à un nombre entier. Nous montrerons plusieurs exemples de tels plans. Ensuite, nous proposerons une nouvelle méthode qui échantillonne en étalant mieux les points que toutes les méthodes existantes et qui permet de construire des plans de sondages périodiques quand ces plans existent.

Bibliographie

- [1] Deville, J.-C. et Tillé, Y. (2000). Selection of several unequal probability samples from the same population. *Journal of Statistical Planning and Inference*, 86:215–227.
- [2] Grafström, A. et Lundström, N. L. P. (2013). Why well spread probability samples are balanced? *Open Journal of Statistics*, 3(1):36–41.
- [3] Grafström, A., Lundström, N. L. P., et Schelin, L. (2012). Spatially balanced sampling through the pivotal method. *Biometrics*, 68(2):514–520.
- [4] Grafström, A. et Tillé, Y. (2013). Doubly balanced spatial sampling with spreading and restitution of auxiliary totals. *Environmetrics*, 14(2):120–131.
- [5] Marker, D. A. et Stevens Jr., D. L. (2009). Sampling and inference in environmental surveys. In *Sample surveys: design, methods and applications*, volume 29 of Handbook of Statistics, pages 487–512. Elsevier/North-Holland, Amsterdam.
- [6] Stevens Jr., D. L. et Olsen, A. R. (1999). Spatially restricted surveys over time for aquatic resources. *Journal of Agricultural, Biological, and Environmental Statistics*, pages 415–428.
- [7] Stevens Jr., D. L. et Olsen, A. R. (2003). Variance estimation for spatially balanced samples of environmental resources. *Environmetrics*, 14(6):593–610.
- [8] Stevens Jr., D. L. et Olsen, A. R. (2004). Spatially balanced sampling of natural resources. *Journal of the American Statistical Association*, 99(465):262–278.
- [9] Thompson, S. K. (2012). *Sampling*. Wiley, New York.
- [10] Tillé, Y., Dickson, M. M., Espa, G. et Giuliani, D. (2018). Measuring the spatial balance of a sample: A new measure based on the Moran's I index. *Spatial Statistics*, Accepted for publication:1–20.