

# Tirages coordonnés d'échantillons poissonniens

*Desislava NEDYALKOVA, Lionel QUALITÉ<sup>1</sup> et Yves TILLÉ*

Nous proposons une généralisation de la méthode de Brewer et coll., (1972) pour tirer des échantillons coordonnés positivement ou négativement. Cette méthode répond aux besoins exprimés par l'Office Fédéral de la Statistique (OFS, Neuchâtel, Suisse) pour organiser ses enquêtes auprès des entreprises. L'OFS souhaite pouvoir sélectionner des panels, des panels rotatifs et des enquêtes ponctuelles, en contrôlant au mieux la charge de réponse des entreprises. La méthode doit être basée sur des numéros aléatoires permanents et un tirage indépendant des unités (échantillonnage transversal de Poisson). Pour organiser une nouvelle enquête, le sens de coordination avec les enquêtes précédentes, ainsi que l'ordre de priorité de ces coordinations, doit être fourni.

Après une première enquête selon un plan de Poisson, pour chaque unité  $k$  avec la probabilité d'inclusion  $\pi_k^1$ , l'intervalle  $[0,1]$  est partitionné en deux sous-intervalles,  $[0, \pi_k^1]$  et  $[\pi_k^1, 1]$ . L'appartenance du nombre aléatoire  $u_k$  à l'un ou l'autre de ces intervalles indique la sélection ou non de l'unité  $k$ . On constate que, pour coordonner négativement une deuxième enquête avec la probabilité d'inclusion  $\pi_k^2$ , on est naturellement amené à partitionner l'un de ces intervalles, au point  $(\pi_k^1 + \pi_k^2) \text{ modulo } 1$ . L'intervalle  $[0,1]$  est donc subdivisé en trois intervalles, et l'appartenance de  $u_k$  à ces intervalles détermine la sélection de  $k$  aux deux enquêtes. Si l'on veut au contraire coordonner positivement la deuxième enquête, on va encore partitionner  $[0, \pi_k^1]$  ou  $[\pi_k^1, 1]$ , mais, cette fois-ci, la coupure sera en  $\pi_k^2$ .

Sur ce modèle, on définit pour chaque unité une collection d'intervalles, partition de  $[0,1]$ , qui fournit son plan de sondage longitudinal. Pour passer de l'enquête t-1 à l'enquête t, on va classer les intervalles définis à la vague t-1 en fonction des sens et priorités de coordination voulus. Puis on va « répartir » la nouvelle probabilité d'inclusion  $\pi_{k,t}$  parmi ces intervalles. Les intervalles les mieux classés seront des zones de sélection de l'unité au temps t. L'un des intervalles sera subdivisé de manière à ce que la zone de sélection ait la longueur totale  $\pi_{k,t}$ .

Ce système permet de gérer les naissances et décès d'unités (le choix des probabilités d'inclusion, en fonction des objectifs, est libre). Il permet d'organiser des panels rotatifs, ceux-ci seront en fait des assemblages de plusieurs sous-échantillons. Enfin, les scissions et fusions d'entreprises sont également possibles. La méthode permet d'avoir une coordination optimale entre l'enquête en cours et celle avec laquelle elle doit être coordonnée en priorité. L'inconvénient principal est que les tailles d'échantillon ne sont pas fixes.

Cette méthode nécessite de stocker et de manipuler des données dont la taille n'augmente que linéairement avec le nombre d'enquêtes. Une première implémentation en SAS montre

---

<sup>1</sup> Université de Neuchâtel - lionel.qualite@unine.ch

qu'elle est tout à fait utilisable dans le cas de quelques dizaines d'enquêtes et d'une base de sondage de plusieurs centaines de milliers d'unités, sur un ordinateur standard.

Références :

Brewer, K., Early, L., et Joyce, S. (1972). Selecting several samples from a single population. *Australian Journal of Statistics*, 3:231--239.