Autour de la coordination d'échantillons Poissoniens

Lionel Qualité

Université de Neuchâtel

janvier 2012

Coordination d'échantillons

- Contrôle du recouvrement entre échantillons,
- pour répartir la charge d'enquête sur la population (coordination négative),
- but: obtenir des meilleurs taux de réponse,
- ou pour estimer plus précisément des évolutions et différences (coordination positive),
- ne réduit pas la charge d'enquête.

Echantillonnage de Poisson coordonné

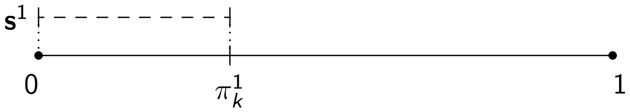
- Extension de la méthode de Brewer et al. (1972) pour la sélection de deux échantillons avec des numéros aléatoires permanents,
- permet de sélectionner des échantillons coordonnés pour des enquêtes ponctuelles, des panels ou des panels rotatifs,
- dans une population dynamique avec des naissances, décès mais aussi des fusions et des scissions,
- fournit des échantillons transversaux poissoniens (sélections indépendantes à probabilités inégales des participants),
- permet une coordination optimale dans un certain sens du terme.

Situation - notations

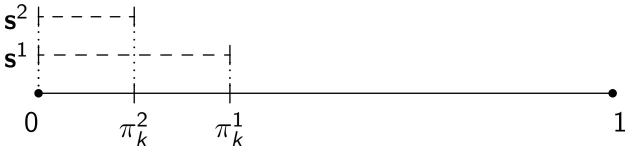
- Population à différentes dates t, U^t, $t = 1, \ldots,$
- probabilité d'inclusion de l'unité k dans le t^{ieme} échantillon : π_k^t,
- probabilités d'inclusion jointes aux temps t et s: π_k^{ts},
- tirages indépendants : $\pi_k^{ts} = \pi_k^t \pi_k^s$,
- coordination positive k si $\pi_k^{ts} > \pi_k^t \pi_k^s$, et négative sinon,
- coordination "optimale" si l'une des bornes $max(0, \pi_k^t + \pi_k^s - 1) \leq \pi_k^{ts} \leq min(\pi_k^t, \pi_k^s)$ est atteinte.

Méthode de Brewer pour deux échantillons - 1

- Premier tirage

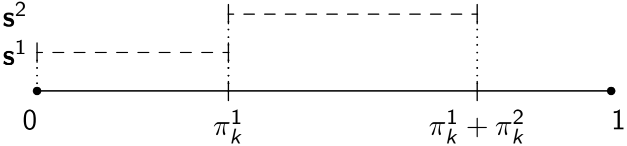


- Coordination positive avec $\pi_k^2 \leq \pi_k^1$

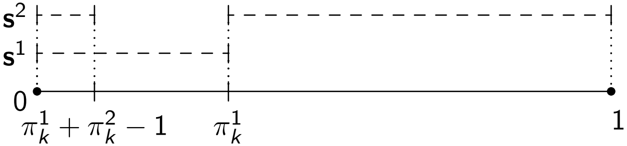


Méthode de Brewer pour deux échantillons - 2

- Coordination négative quand $\pi_k^1 + \pi_k^2 \leq 1$

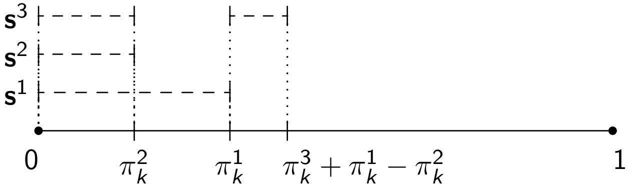


- Coordination négative quand $\pi_k^1 + \pi_k^2 \geq 1$



Généralisation à 3 enquêtes et plus

1. Mettre un ordre sur les sous-intervalles de $[0,1]$ définis au tirage précédent selon les règles de coordination choisies,
2. construire une zone de sélection pour la nouvelle enquête,
3. exemple : troisième enquête coordonnée positivement avec la seconde puis négativement avec la première.



Implementation

- Nécessite de donner un ordre de priorité sur les coordinations,
- croissance linéaire des données et du temps de calcul,
- utilisé depuis 10/2009 pour les enquêtes entreprise et 11/2010 pour les enquêtes population,
- 930'000 unités sélectionnées dans une population de 8'000'000, en 21 enquêtes : 10 unités sélectionnées deux fois contre plus de 42'000 attendues si tirages indépendants.

Points problématiques

1. Taille d'échantillon aléatoire
   - coûts incontrolés
   - variance accrue
2. Sélections indépendantes
   - pas de coordination "à plusieurs niveaux" ex: enquêtes ménages et individus,
   - tirages multiples dans des ménages possibles.

Taille aléatoire - 1

- Normalement très faible variation de la taille totale,
- mais peut poser des problèmes pour de petites enquêtes ou pour contrôler la charge de travail des enqueteurs
- la variance de l'estimateur calé après un tirage poissonien est proche de celle de l'estimateur de Horvitz-Thompson après un tirage de taille fixe (*pour des probabilités d'inclusion égales et après approximations*).

Taille aléatoire - 2

- Reste un problème pour de petits domaines dans lesquels on voudrait avoir un nombre minimal d'unités,
- pas un nouveau problème : était déjà présent à cause de la non-réponse,
- solution : calculer la probabilité de sélectionner un échantillon trop petit dans des domaines et augmenter les probabilités d'inclusion pour limiter ce risque,
- le coût peut être élevé si l'on veut avoir un risque vraiment faible, ou l'allocation peut être inefficace.

Sélections indépendantes - 1

- Pas de bonne solution pour la coordination multi-niveaux,
- le plus simple (pour la coordination négative) serait de coordonner au niveau le plus élevé et d'utiliser un plan à plusieurs degrés,
- mais cela ne permet pas de tirer des panels au niveau le plus bas,
- difficulté pour suivre des ménages plus grande que pour suivre des individus.

Sélections indépendantes - 2

- Pour certaines enquêtes (ex: l'enquête “de recensement” suisse), on veut sélectionner au maximum une unité par ménage
- solution : plusieurs phases de tirage. Un échantillon coordonné est d'abord tiré avec un certain jeu de probabilités d'inclusion, puis, dans chaque ménage avec plusieurs sélections, on retient au hasard une des unités sélectionnées.
- En général les probabilités de tirage de première phase doivent être approchées numériquement, mais dans des cas simples on peut les calculer directement.
- exemple : probabilités d'inclusion π égales dans un ménage de taille m, si $m\pi \leq 1$, on utilise $p = 1 - (1 - m\pi)^{\frac{1}{m}}$.
- Inconvénient : la coordination est dégradée pour les grands ménages.

Sélections indépendantes - 3

- Une autre enquête devait être sélectionnée en plusieurs vagues et avec la même contrainte,
- si les ménages ne changent pas entre les tirages et si les probabilitéss d'inclusion sont égales dans un ménage, le paramètre de tirage de première phase est simple :
  - au temps 2, $q = 1 - p - \left[(1-p)^m - m\pi^2\right]^{\frac{1}{m}}$,
  - au temps 3, $r = 1 - p - q - \left[(1-p-q)^m - m\pi^3\right]^{\frac{1}{m}} \ldots$
- pour les ménages qui changent, on a besoin de procédures numériques,
- environ 2% des ménages ont changé entre chaque vague de tirage.

Conclusion

- La variabilité de la taille d'échantillon n'est pas un problème nouveau,
- le problème des sélections multiples a pu être résolu jusqu'à présent pour les enquêtes de l'OFS,
- la simplicité du plan de Poisson permet de conserver le plan longitudinal complet et les calculs pour répondre à des exigences imprévues.

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

Grab, E. & Savage, I. (1954).
Tables of the expected value of 1/x for positive bernoulli and poisson variables.
Journal of the American Statistical Association **49**, 169–177.

Marciniak, E. & Wesolowski, J. (1999).
Asymptotic eulerian expansions for binomial and negative binomial reciprocals.
Proceedings of the American Mathematical Society **127**, 3329–3338.

Qualité, L. (2009).
Unequal probability sampling and repeated surveys.
Thèse de doctorat, Université de Neuchâtel, Neuchâtel, Suisse.

Thionet, P. (1963).
Sur le moment d'ordre (-1) de la distribution tronquée. application à l'échantillonnage de hájek.
Publ. Inst. Statist. Univ. Paris 12 **31:827**, 93–102.