

Tirages coordonnés d'échantillons poissonniens

Desislava Nedyalkova, Lionel Qualité, Yves Tillé

Université de Neuchâtel

mars 2009

Motivation

Coordination

Méthode proposée

Utilisation

Références

Besoins de l'OFS

- ▶ Méthode qui apporte des résultats pour chaque unité :
coordination individuelle la meilleure possible,
- ▶ permet de gérer toutes les enquêtes : coordinations positives
et négatives en même temps,
- ▶ adaptée à une population dynamique : naissances, décès mais
aussi scissions, fusions,
- ▶ fournit des échantillons poissoniens : probabilités inégales,
formules de variance simples.

Méthodes existantes

- ▶ On trouve dans Hesse (1999) un tour d'horizon;
- ▶ La plupart des méthodes fournissent des plans simples ou stratifiés, au prix de pas mal de difficultés lorsque la population est dynamique;
- ▶ Ne permettent pas toujours d'avoir des enquêtes coordonnées positivement et d'autres négativement.

Introduction

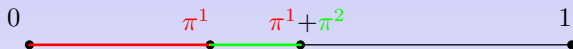
- ▶ Population envisagée à différents instants t , U^t , $t = 1, \dots$;
- ▶ Probabilité d'inclusion de l'individu k à l'enquête t : π_k^t ;
- ▶ Probabilité d'inclusion aux dates t et s : π_k^{ts} ;
- ▶ Si deux enquêtes sont indépendantes, $\pi_k^{ts} = \pi_k^t \pi_k^s$;
- ▶ Il y a coordination positive pour l'unité k si $\pi_k^{ts} > \pi_k^t \pi_k^s$ et négative si c'est dans l'autre sens;
- ▶ $\max(0, \pi_k^t + \pi_k^s - 1) \leq \pi_k^{ts} \leq \min(\pi_k^t, \pi_k^s)$.

Charge d'enquête

- ▶ Sur T enquêtes, l'unité k sera sélectionnée en moyenne $\pi_k^1 + \dots + \pi_k^T$ fois;
- ▶ Pour la coordination négative, les différentes méthodes utilisées en pratique donnent toutes le même temps moyen hors échantillon (à peu près);
- ▶ Dans ce cas un plan longitudinal systématique a de bonnes propriétés (voir Nedyalkova et al., 2009);

Méthode de Brewer

- ▶ Numéro aléatoire permanent;
- ▶ Zones de sélection :
 - ▶ pour la coordination négative,



- ▶ pour la coordination positive,

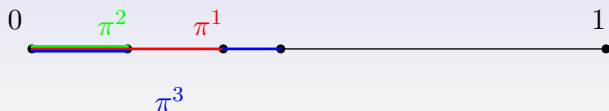


- ▶ Avantages :
 - ▶ les naissances et les décès ne posent pas de problème,
 - ▶ la coordination est optimale.

Généralisation

1. Attribuer un score à chaque intervalle en fonction du sens de coordination voulu et des priorités,
2. au temps t , déterminer une zone de sélection de longueur π^t qui contient les intervalles avec les scores les plus élevés,
3. après la vague t , $[0, 1]$ est coupé en $t + 1$ intervalles.

Exemple : 3ème enquête coordonnée d'abord positivement avec la 2ème puis négativement avec la 1ère.



Propriétés

- ▶ Plan transversal poissonien;
- ▶ Coordination optimale entre une enquête et celle avec laquelle elle devait être coordonnée en priorité;
- ▶ Dans le cas d'enquêtes coordonnées négativement :
 - ▶ plan longitudinal systématique,
 - ▶ nombre de sélections pour chaque unité égal à $E(\sum \pi_k^t)$ ou $E(\sum \pi_k^t) + 1$;

Application aux panels rotatifs - 1

Tirage initial : par exemple, pour un panel dont un cinquième est renouvelé chaque année, tiré après $t - 1$ enquêtes.

1. Définir les règles de coordination avec les enquêtes précédentes;
2. Tirer un premier échantillon s^t avec ces règles de coordination et les probabilités d'inclusion $\pi^{panel}/5$;
3. Tirer un deuxième cinquième, s^{t+1} , coordonné en priorité négativement avec le premier cinquième, puis avec les autres règles de coordination;
4. Tirer les trois derniers cinquièmes selon le même principe;
5. La première vague du panel est $s^t \sqcup \dots \sqcup s^{t+4}$.

Application aux panels rotatifs - 2

Renouvellement : la deuxième vague du panel arrive alors que $u - 1$ enquêtes ont été tirées.

1. On attribue des probabilités d'inclusion aux nouvelles unités de la population;
2. Sélection du nouveau cinquième s^u coordonné négativement avec s^t puis s^{t+1}, \dots, s^{t+4} , puis avec les règles choisies quant aux autres enquêtes;
3. Mise à jour du premier cinquième conservé s^t : on tire un échantillon s^{u+1} coordonné positivement avec s^t , puis négativement avec $s^u, s^{t+1}, \dots, s^{t+4}$;
4. On fait de même pour mettre à jour s^{t+1}, \dots, s^{t+3} ;
5. La deuxième vague du panel est $s^u \sqcup \dots \sqcup s^{u+4}$.

Naissances - décès - scissions

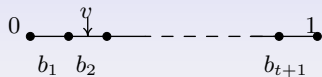
- ▶ Une unité qui naît après t enquêtes est ajoutée avec ses t premières zones de sélection de longueurs nulles;
- ▶ Une unité qui disparaît reçoit une probabilité d'inclusion nulle à chaque enquête suivante;
- ▶ Deux (ou plus) unités qui proviennent d'une scission sont ajoutées à la population
 - ▶ soit avec un passé vierge, comme des unités qui viennent de naître,
 - ▶ soit en héritant le passé de l'unité dont elles descendent.

Fusions - 1

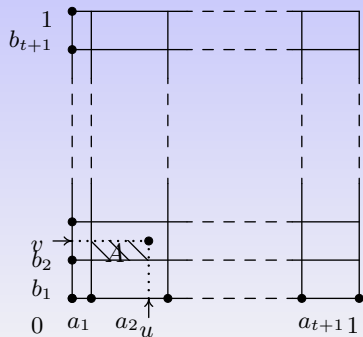
Première unité :



Deuxième unité :



(a) plans marginaux



(b) plan joint

Fusions - 2

But : intégrer cette unité particulière dans le système d'enquête.

1. *Gratuit* : se ramener à des intervalles de $[0, 1]$ et un nombre aléatoire;
2. *Nécessaire pour pouvoir calculer des scores* : choisir une fonction de passage des couples d'échantillons longitudinaux des deux unités à un seul échantillon longitudinal,

$$f : (\mathbf{s}_k, \mathbf{s}_\ell) \mapsto \mathbf{s}_m$$

3. *Optionnel mais plus prudent* : se ramener de $(t + 1)^2$ échantillons possibles à $t + 1$ (comme pour une unité "normale"),
 - ▶ préserver l'échantillon tiré,
 - ▶ préserver la longueur de l'échantillon tiré,
 - ▶ dégrader le moins possible...

Mise en oeuvre

- ▶ Nécessité de conserver tout le plan joint des t vagues précédentes pour pouvoir coordonner la suivante (et non pas seulement les échantillons sélectionnés), et les numeros aléatoires permanents : taille des données $O(tN)$;
- ▶ Possibilité d'effacer d'anciennes enquêtes (avec lesquelles on ne veut plus coordonner), de réorganiser les zones de sélection;
- ▶ Calcul rapide des probabilités d'inclusion π_k^{ts} ;

- [1] Brewer, K., Early, L., and Joyce, S. (1972). Selecting several samples from a single population. *Australian Journal of Statistics*, 3:231–239.
- [2] Cotton, F. et Hesse, C. (1992). Tirages coordonnées d'échantillons. Document de travail de la Direction des Statistiques Économiques E9206. Technical report, INSEE, Paris.
- [3] Hesse, C. (1999). Sampling co-ordination: a review by country. Technical Report E9908, INSEE.
- [4] Nedyalkova, D., Qualité, L. et Tillé, Y. (2009). General framework for the rotation of units in repeated survey sampling. *Accepté pour publication dans Statistica Neerlandica*.
- [5] Rivière, P. (2001). Coordinating samples using the microstrata methodology. Proceedings of Statistics Canada Symposium 2001.