

Algorithmes d'affectation d'un échantillon de logements entre différents enquêteurs.

Marc CHRISTINE (*),

Louise DE FERRAN, Rémi BARRAUD, Paul BOISSEAU, Luigi ROMAIN, Dorian TESSIER (**)

(*) Ancien Insee, mchristine7577@gmail.com

(**) Mines Paris-PSL

Mots-clés (6 maximum) : enquêtes en face à face, réseau d'enquêteurs, coût de déplacement, allocation optimale de l'échantillon, minimisation du coût.

Domaine concerné : Modélisation - Algorithmes

Résumé

Un problème pratique courant, dans l'organisation des enquêtes en face à face - principalement auprès des ménages – est celui de la répartition de l'échantillon entre les différents enquêteurs, qui constituent un réseau fixe, qui se partageront la collecte auprès des unités de l'échantillon.

L'un des défauts des enquêtes en face à face – au regard de leurs multiples avantages – est son **coût**, notamment celui qui est lié aux déplacements des enquêteurs. Il est donc opportun de chercher à réduire ce coût autant que possible. La question se pose donc de chercher à affecter les logements de l'échantillon au mieux entre les différents enquêteurs disponibles, de manière à **minimiser ce coût lié aux déplacements**.

Ce petit papier présente donc différents algorithmes, soit empiriques, soit analytiques, permettant de réaliser cette affectation dans une optique d'optimalité.

Formalisation

On dispose d'un échantillon de n logements, notés $i = 1, \dots, n$, qu'il s'agit de répartir entre K enquêteurs, notés $j = 1, \dots, K$, localisés de manière prédéterminée.

Les paramètres.

- Chaque enquêteur j a une charge d'enquête (égale au nombre de logements qu'il peut ou doit interroger), notée C_j . Les charges doivent être a priori adaptées à la taille de l'échantillon. Elles doivent donc satisfaire la contrainte : $\boxed{\sum_{j=1}^K C_j = n}$.

- Les logements échantillonnés et les enquêteurs sont localisés dans l'espace. On peut résumer cette localisation par la distance $d_{i,j}$ entre le logement i et l'enquêteur j . Il s'agit soit d'une vraie distance, soit d'un indicateur généralisé de coût.

L'optimalité recherchée.

Notons E_j l'ensemble des logements attribués à l'enquêteur j ; les E_j doivent constituer une partition de l'échantillon. Posons, pour tout $i \in \{1, 2, \dots, n\}$ et pour tout $j \in \{1, 2, \dots, K\}$:

$$a_{i,j} = \begin{cases} 1 & \text{si } i \in E_j \\ 0 & \text{sinon} \end{cases} \in \{0, 1\}.$$

On cherchera alors à minimiser un critère de type « coût total », soit :

$$\text{Min}_{E_1, \dots, E_K} \sum_{j=1}^K \sum_{i \in E_j} d_{i,j} = \text{Min}_{a_{i,j}} \sum_{j=1}^K \sum_{i=1}^n a_{i,j} d_{i,j}$$

sous les conditions :

$$\begin{aligned} \forall i \in \{1, \dots, n\} : \sum_{j=1}^K a_{i,j} &= 1 \text{ [un logement } i \text{ est attribué à un seul enquêteur]} \\ \forall j \in \{1, \dots, K\} : \sum_{i=1}^n a_{i,j} &= C_j \text{ [contrainte de charge pour chaque enquêteur } j] \end{aligned}$$

Méthodologie de résolution proposée

On testera tout d'abord des algorithmes empiriques qui ne nécessitent pas de modélisation mathématique. On construira ensuite un problème approché dont les solutions seront explicites analytiquement, ce qui permettra d'approcher l'optimalité.

La méthode analytique se déroulera génériquement en trois temps :

- formulation du problème approché
- résolution de ce problème, permettant de définir des coefficients $a_{i,j}^*$ approchant les solutions cibles $a_{i,j}$
- affectation définitive des logements aux enquêteurs à partir de ces coefficients $a_{i,j}^*$ selon un processus de choix approprié.

On obtiendra et présentera donc différents algorithmes selon la formulation retenue du problème approché. D'autres procédures plus complexes pourront être testées (programmation linéaire en nombres entiers).

Bibliographie

Ardilly, P. : Efficacité / coût de l'échantillon-maître Insee, 13^e colloque international francophone sur les sondages, novembre 2024.

Christine M., Davezies L., Morando M., Rousseau S. : Rénovation de l'enquête Camme : de l'échantillonnage au calcul de précision, en passant par l'allocation optimale entre les enquêteurs, 9^{es} Journées de méthodologie statistique de l'Insee, mars 2005.