

PLANS SYSTÉMATIQUES QUASI-ÉQUILIBRÉS

Lionel Qualité ¹ & Yves Tillé ¹

¹ *Institut de Statistique, Université de Neuchâtel, Pierre à Mazel 7, 2000 Neuchâtel, Suisse,*
lionel.qualite@unine.ch, yves.tille@unine.ch

Nous revenons sur le problème de l'échantillonnage équilibré, qui a été pour l'essentiel résolu dans Deville et Tillé (2004). Pour une population finie U de taille N , il s'agit de trouver un plan de sondage $p(\cdot)$ tel que

$$\boldsymbol{\pi} = \sum_{\mathbf{s}} \mathbf{s}p(\mathbf{s}) \quad \text{et} \quad p(\mathbf{s}) > 0 \Rightarrow \sum_{k \in \mathbf{s}} \frac{\mathbf{x}_k}{\pi_k} = \sum_{k \in U} \mathbf{x}_k,$$

où $\boldsymbol{\pi}$ est un vecteur de probabilités d'inclusion spécifié, et \mathbf{x}_k est un vecteur de données auxiliaires connues pour toutes les unités k de la population.

Ce problème admet des solutions dans des cas standards : les plans de taille fixe ($\mathbf{x}_k = \pi_k$), les plans "stratifiés". Lorsque la matrice d'information auxiliaire est plus compliquée, par exemple lorsqu'elle contient des variables numériques, le problème n'a en général pas de solution exacte. Cela se traduit dans la méthode du cube par l'utilisation d'une procédure appelée "phase d'atterrissage" pour régler la sélection de quelques unités de la population en ne respectant plus les contraintes d'équilibrage.

Lorsque le problème n'est pas exact, on cherche à minimiser une fonction de coût, par exemple la somme des coefficients de variation des estimateurs de Horvitz-Thompson des totaux des variables d'équilibrage. Cette fonction de coût est la restriction d'une fonction linéaire en $p(\cdot)$, et ses extrema sont des plans à support minimal. Dans Deville et Tillé (2004), il est proposé d'utiliser l'algorithme du simplexe pour trouver ce minimum. Du fait de la grande dimensionnalité du problème cela n'est envisageable que pour un nombre de contraintes faible (une vingtaine). Une autre option est d'abandonner les contraintes successivement de manière à réduire la taille du problème résiduel.

Nous proposons de chercher directement des plans proches de l'équilibre parmi des plans à supports minimaux. Dans Pea et coll. (2006), il est prouvé que les plans systématiques sont à support minimal. Pour un vecteur $\boldsymbol{\pi}$ choisi, les plans systématiques sont tous obtenus en changeant l'ordre des unités de la population. Nous avons programmé quelques procédures de recherche d'un ordre tel que le plan obtenu soit proche de l'équilibre et effectué des simulations. Sur des populations de quelques centaines d'unités, l'équilibrage obtenu est en général meilleur que celui obtenu avec la méthode du cube quelle que soit l'option d'atterrissage. Pour des populations plus grandes, et un nombre de variables d'équilibrage élevé, cette procédure peut être utilisée comme phase d'atterrissage et fonctionne aussi bien que l'option d'élimination successive des contraintes.

Bibliographie

Deville, J.-C., et Tillé, Y. (2004). Efficient balanced sampling : The cube method, *Biometrika*, 91, 893-912.

Pea, J., et Qualité, L., et Tillé, Y. (2007). Systematic sampling is a minimal support design, *Computational Statistics and Data Analysis*, 51, 5591-5602.