

Résolution d'une des limites de l'allocation de Neyman

Malik KOUBI¹ et Sandrine MATHERN²

L'allocation de Neyman estime avec la plus grande précision possible la moyenne d'une variable d'intérêt sur l'ensemble de la population étudiée. Cette allocation peut en revanche souffrir d'imprécision lorsqu'on s'intéresse à des estimations sur des sous-groupes de la population. La méthode développée ici consiste à ajouter des contraintes à l'optimisation classique de Neyman, contraintes qui permettent d'assurer une précision fixée à l'avance dans ces sous-groupes appelés « regroupements de publication ». La seule condition requise est que chaque regroupement de publication soit une réunion de strates d'échantillonnage (on exclut les domaines plus complexes).

L'intérêt de la méthode est tout d'abord d'automatiser une pratique que le statisticien fait en général « à la main », qui consiste à revoir à la marge l'allocation de Neyman, en réallouant les unités des regroupements où la précision est la meilleure, vers ceux où elle est la moins bonne. Par ailleurs, cette méthode garantit une optimalité de l'allocation, car le passage de l'allocation de Neyman à l'allocation proposée se fait le long de la frontière d'efficacité entre précision globale (précision de l'estimateur au niveau de la population) et précision locale (définie comme la plus mauvaise précision des estimateurs au niveau des regroupements de publication). Cette frontière permet d'évaluer la perte en précision globale lorsque l'on tient à fixer une précision minimale dans des regroupements de publication.

Pour résoudre ce programme d'optimisation sous contrainte, on procède en deux étapes. Premièrement, on traduit les contraintes de précision locale en termes de nombre d'unités à tirer dans les strates d'échantillonnage. Deuxièmement, on détermine l'ordre de saturation des strates d'échantillonnage quand la taille globale n de l'échantillon augmente. Cet ordre permet ensuite de déterminer, pour une valeur de n , la liste des strates saturées. Il ne reste plus alors qu'à appliquer une allocation de Neyman aux strates non saturées, avec une taille d'échantillon globale diminuée des valeurs de saturation.

¹ DR Languedoc-Roussillon, Insee

² Drees