Conférence spéciale

Plans de sondage à dispersion minimale

Jean-Claude DEVILLE et Mohamed El Hadj TIRARI[1]

On s'intéresse à l'échantillonnage de taille fixe n dans une population finie U de taille N. Un échantillon s a la probabilité $p(s)$ d'être tiré. On s'impose de respecter des probabilités d'inclusion $\pi_k = \sum_{s/k\in s} p(s)$, données pour chaque unité k de U. En l'absence d'autres informations, on cherche à n'accorder aucun privilège à quelque échantillon que ce soit. Quand toutes les π_k sont égales à n/N, on rend tous les échantillons équiprobables, donc on choisit le sondage aléatoire simple. Si les probabilités d'inclusion sont inégales, un critère de choix naturel consiste à minimiser la dispersion des nombres $p(s)$.

Dans la première partie de cet exposé, nous examinerons une formulation assez générale de la notion de dispersion et du problème de sa minimisation sous contraintes. Les cas typiques sont les critères d'entropie ou de variance. L'optimisation conduit pour le premier critère à une solution où tous les échantillons de taille n ont une probabilité strictement positive alors que c'est généralement le contraire pour le second. On caractérise les critères de dispersion selon cette propriété.

Nous nous intéresserons ensuite plus spécifiquement au critère de variance et nous établirons ses liens avec le schéma de MIDZUNO-LAHIRI en exhibant une méthode générale d'échantillonnage qui étend de façon surprenante ce schéma. On étend ensuite ces résultats au cas d'un critère général de dispersion.

Bien que le plan à entropie maximum possède bien des propriétés théoriques ou pratiques attirantes, les plans à dispersion minimale, variance en particulier, peuvent parfois avoir leur intérêt en éliminant ou en réduisant la probabilité d'échantillons peu désirables.

[1] Ensai/Crest, Insee - deville@ensai.fr et tirari@ensai.fr