

LA MISE EN ŒUVRE D'ALGORITHMES DE STRATIFICATION UNIVARIÉE

Sophie Baillargeon ¹ & Marcin Kozak ² & Louis-Paul Rivest ¹

¹ *Département de mathématiques et de statistique, Université Laval, Québec, Canada,*
Sophie.Baillargeon@mat.ulaval.ca, Louis-Paul.Rivest@mat.ulaval.ca

² *Department of Experimental Design and Bioinformatics, Warsaw University of Life Sciences, 02-776 Warsaw, Poland,* nyggus@gmail.com

Cet exposé considère la stratification d'une population à partir d'une mesure de taille unidimensionnelle X . La strate h sera composée de toutes les unités ayant des valeurs de X dans l'intervalle $[b_{h-1}, b_h)$. La détermination des bornes b_h est un problème classique en échantillonnage. Une première solution, la méthode cum-root-f, a été proposée par Dalenius et Hodge il y a plus de 60 ans. De nos jours, dans la foulée des travaux de Lavallée et Hidiroglou (1988), de nombreux algorithmes ont été mis de l'avant pour déterminer une stratification optimale. On cherche, par exemple, à minimiser le CV de l'estimateur de la moyenne de la variable d'intérêt pour une taille d'échantillon n fixe, ou à minimiser n pour un CV prédéterminé. Une méthode performante pour déterminer des bornes de strate optimales est l'algorithme de Kozak (2004). Il considère, de façon aléatoire, une modification à l'une des bornes courantes et accepte cette dernière si elle diminue le critère d'optimalité. L'algorithme se termine lorsque les bornes restent inchangées après un grand nombre d'essais de modification. Ce problème d'optimisation a une composante discrète ; il faut, considérant la méthode d'allocation retenue, déterminer des tailles d'échantillons entières pour les strates. Plusieurs algorithmes font l'optimisation en continu avec des tailles d'échantillons réelles et discrétisent les résultats à la fin du calcul. Nous comparerons cette stratégie de calcul avec une implémentation des algorithmes de stratification qui contraignent les tailles d'échantillon à être des valeurs entières. Le programme R `stratification`, qui met en œuvre les deux stratégies d'optimisation, sera utilisé pour les calculs.

Bibliographie

Baillargeon, S. & Rivest, L.-P. (2011). Élaboration de plans stratifiés en R à l'aide du programme `stratification`, *Techniques d'enquête*, 37, 59-72.

Gabler, S., Ganniger, M. & Münnich, R. (2012) Optimal Allocation of the Sample Size to the Strata under Box Constraints. *Metrika*, 71, 151-161

Kozak, M. (2004). Optimal Stratification Using Random Search Method in Agricultural Surveys. *Statistics in Transition*, 6, 797-806.

Lavallée, P. & Hidiroglou, M. (1988). Sur la stratification de populations asymétriques. *Techniques d'enquête*, 14, 35-45.