Méthodes de couplage pour un tirage à plusieurs degrés

Guillaume Chauvet, ENSAI (CREST), Campus de Ker Lann, Bruz, chauvet@ensai.fr

L'échantillonnage à plusieurs degrés est une technique couramment utilisée pour les enquêtes auprès des ménages ou pour les enquêtes sur la santé, quand il n'existe pas de base de sondage ou quand la population d'intérêt est dispersée sur un grand territoire. Il n'est pas rare d'utiliser trois degrés de tirage, ou plus. Par exemple, la troisième Enquête sur la Santé et la Nutrition (NHANES III) réalisée aux Etats-Unis a nécessité quatre degrés d'échantillonnage, avec la sélection de comtés au premier degré, de segments au second degré, de ménages au troisième degré et d'individus au quatrième degré (voir Ezzati et al., 1992). Un traitement détaillé de l'échantillonnage à plusieurs degrés est donné dans Cochran (1977), Särndal et al. (1992) ou encore Fuller (2009).

L'échantillonnage à plusieurs degrés introduit une dépendance complexe dans la sélection des unités finales, ce qui rend les résultats asymptotiques difficiles à prouver. Dans ce travail, nous utilisons des méthodes de couplage (Hajek, 1961, 1964 ; Thorisson, 2000) pour relier l'échantillonnage à plusieurs degrés à des plans de sondage à la fois proches et plus simples, au sens où les unités primaires sont sélectionnées indépendamment. Le principe de la méthode est de générer un vecteur aléatoire (X_t, Z_t) avec des lois marginales appropriées, et de façon à ce que X_t et Z_t soient proches. Par exemple, la distribution de Z_t peut être celle de l'estimateur de Narain (1951)-Horvitz-Thompson (1952) pour un tirage à plusieurs degrés avec sondage aléatoire simple sans remise (SI) des unités primaires, et la distribution de X_t peut être celle de l'estimateur de Hansen-Hurwitz (1943) pour un tirage à plusieurs degrés avec sondage aléatoire simple avec remise (SIR) des unités primaires. Le couplage est réalisé de façon à ce que $E(X_t - Z_t)^2$ soit plus petit que le taux de convergence de X_t
.

Dans ce travail, nous utilisons les méthodes de couplage pour obtenir des résultats asymptotiques pour un tirage à plusieurs degrés avec tirage sans remise des unités primaires. Un algorithme de couplage proposé par Hajek (1961) peut être généralisé au contexte d'un tirage SI des unités primaires pour obtenir la normalité asymptotique de l'estimateur de Narain-Horvitz-Thompson. Plus spécifiquement, nous introduisons un nouvel algorithme de couplage pour un tirage SI/SIR d'unités primaires, et nous montrons que l'estimateur de Hansen-Hurwitz et l'estimateur de Narain-Horvitz-Thompson sont proches quand la fraction de sondage du premier degré devient négligeable. Cet algorithme de couplage est utilisé pour résoudre un vieux problème, i.e. que la méthode appelée "Bootstrap avec remise des unités primaires" (Rao and Wu, 1988) est consistante dans le cas d'un tirage SI d'unités primaires avec une faible fraction de sondage au premier degré, et conduit à des estimateurs de variance consistants pour des fonctions lisses de moyennes. La généralisation au cas d'un tirage à probabilités inégales des unités primaires est actuellement à l'étude. Les propriétés de la procédure de Bootstrap sont également évaluées à l'aide d'une courte étude par simulations.

Références

Cochran, W.G. (1977). Sampling Techniques. New-York, Wiley.

Ezzati, T.M., Hoffman, K., Judkins, D.R., Massey, J.T., and Moore, T.F. (1992). Sample design: Third National Health and Nutrition Examination Survey. Vital and Health Statistics, 2, 113, National Center for Health Statistics.

Fuller, W.A. (2009). Sampling Statistics. New-York, Wiley.

Hajek, J. (1960). Limiting distributions in simple random sampling from a finite population. Publications of the Mathematics Institute of the Hungarian Academy of Science, 5, 361-74.

Hajek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. The Annals of Mathematical Statistics, 35, 1491-1523.

Hansen, M.H., and Hurwitz, W.N. (1943). On the theory of sampling from finite populations. The Annals of Mathematical Statistics, 14, 333-362.

Horvitz, D.G., and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association, 47, 663-685.

Narain, R.D. (1951). On sampling without replacement with varying probabilities. Journal of the Indian Society of Agricultural Statistics, 3, 169-175.

Rao, J.N.K and Wu, C.F.J. (1988). Resampling inference with complex survey data. Journal of the american statistical association, 83, 231-241.

Särndal, C.-E., Swensson, B. and Wretman, J.H. (1992). Model Assisted Survey Sampling. New-York, Springer-Verlag.