

Calculs de précision dans Octopusse : théorie et application à l'enquête Logement 2013

Emmanuel Gros¹, Martin Chevalier² et Karim Moussallam³

¹ Insee, 18 boulevard Adolphe Pinard, 75675 Paris cedex 14, FRANCE – emmanuel.gros@insee.fr

² Insee, 18 boulevard Adolphe Pinard, 75675 Paris cedex 14, FRANCE – martin.chevalier@insee.fr

³ Dares, 39-43, quai André-Citroën, 75015 Paris, FRANCE – karim.moussallam@travail.gouv.fr

Thème : estimation de variance

Le système actuel d'échantillonnage des enquêtes ménages à l'Insee, baptisé Octopusse et présenté dans [1], s'articule autour de deux concepts majeurs : d'une part le principe de système de tirage à deux degrés des échantillons-maîtres classiques, et d'autre part le recensement rotatif de la population qui permet l'apport d'« informations fraîches » concernant les logements à échantillonner. Cette interaction entre ces deux concepts conduit à un système d'échantillonnage plus efficace mais également nettement plus complexe : aux aléas de sondage « classiques » des échantillons-maîtres s'ajoutent les aléas de sondage relatifs aux enquêtes annuelles de recensement. On passe ainsi du cadre théorique d'un sondage à deux degrés pour les anciens échantillons-maîtres à celui d'un sondage en trois phases pour le système Octopusse, ce qui rend particulièrement ardu les calculs de variance.

En 2011, Guillaume Chauvet s'est intéressé dans [2] au problème des calculs de précision dans le contexte d'Octopusse, et a proposé, dans un cadre légèrement simplifié¹, une méthode d'estimation de variance fondée sur l'utilisation d'estimateurs de variance de Yates-Grundy s'appuyant sur des probabilités d'inclusion double estimées par réplication à partir des propriétés de martingale de l'algorithme du Cube. En s'appuyant sur les programmes fournis par Guillaume Chauvet, l'Insee a prolongé ces travaux de façon à élaborer une formule de variance analytique « générique » valable dans le contexte réel d'Octopusse. Ces travaux complémentaires ont consisté :

- d'une part à calculer les probabilités d'inclusion double – des communes dans les groupes de rotation du recensement d'une part, des unités primaires dans l'échantillon-maître d'autre part – du système Octopusse, à analyser leur qualité ainsi que celle des estimateurs de Yates-Grundy les utilisant.
- d'autre part à prendre en compte les aspects du système Octopusse non traités dans le cadre simplifié des travaux de Guillaume Chauvet : sélection des logements au sein de la fraction des unités primaires recensée lors de la dernière enquête annuelle de recensement, prise en compte de la non-réponse, calage.

Ce cadre méthodologique a été appliqué à l'enquête Logement 2013, et a été adapté aux spécificités de son plan de sondage : tirage dans l'EMEX élargi pour la région Nord-Pas-de-Calais et directement dans l'Enquête annuelle de recensement 2011 pour la région Île-de-France, constitution d'un échantillon de logements neufs (construits entre mars 2011 et mars 2013) par tirage dans la base de permis de construire DGI-SITADEL, constitution d'un échantillon de logements situés en Zone urbaine sensible (ZUS). La mise en œuvre de cette méthodologie a permis de mesurer l'impact des différentes composantes du plan de sondage sur la précision des estimations, en particulier le tirage des logements dans la dernière Enquête annuelle de recensement et les extensions régionales.

Du point de vue pratique, cet investissement méthodologique s'est traduit par la mise à disposition des partenaires en charge de l'exploitation des données de l'enquête d'un programme de calcul de précision spécifique à l'enquête Logement 2013. Ce programme est structuré de façon à ce

¹ En particulier, le dernier degré de sondage correspondant à la sélection des logements au sein de la fraction des unités primaires recensée lors de la dernière enquête annuelle de recensement n'était pas pris en compte.

que les modules correspondant aux différentes composantes du calcul de variance soient aisément réutilisables dans le cadre d'autres enquêtes échantillonnées avec le système Octopusse.

Références :

[1] Marc Christine et Sébastien Faivre, « Octopusse : un système d'Échantillon-Maître pour le tirage des échantillons dans la dernière Enquête Annuelle de Recensement », *Actes des X^{èmes} Journées de Méthodologie Statistique*, 2009.

[2] Guillaume Chauvet, « On variance estimation for the French master sample », *Journal of Official Statistics*, Vol. 27, No. 4, 2011, pp. 651–668, 2011.