Mercredi 16 mars 2005 : 14h35-15h00 *(en parallèle avec la session 15)*

Eléments sur la non-réponse non-ignorable et les mécanismes de sélection dans les enquêtes.

Eric GAUTIER
Insee, Unité Méthodes Statistiques

Les mécanismes de sélection sont présents à plusieurs niveaux lors de la collecte d'une enquête : lors de l'établissement du plan de sondage, lorsqu'une partie des unités tirées ne répond pas du tout à l'enquête (non-réponse totale) et enfin en dernier lieu lorsque seulement certaines questions sont renseignées (non-réponse partielle). Dans les deux derniers modes de sélection le mécanisme est inconnu et l'on peut au mieux mener une estimation. Considérons désormais la sélection à un niveau donné. Supposons que les observations, les variables intervenant lors de la conception du plan de sondage et le mécanisme de sélection sont des réalisations aléatoires. Dans le cas le plus général qui soit la sélection, quel que soit le niveau, peut dépendre du premier jeu de données (les observations augmentées des variables intervenant dans le plan de sondage). Il se peut qu'en conditionnant par un sous-ensemble de variables parfaitement observées on obtienne de l'indépendance entre la sélection et des variables d'intérêt. Alors un modèle correctement spécifié permettra de fournir des estimations de sondage ou économétriques sans biais et l'on pourra ignorer la dépendance et n'inférer que sur le mécanisme de sélection ou que sur le comportement. Si par contre, quel que soit le conditionnement, la dépendance subsiste, l'ignorer, ce qui revient à ignorer la sélection en inférant sur un comportement ou à ignorer le comportement en inférant sur la sélection, entraînera des biais.

En pratique il est tout à fait crédible qu'au stade de la non-réponse totale la sélection dépende des variables d'intérêt et pas seulement des variables qui sont intervenues à la conception du plan de sondage, par exemple si les enquêtés reçoivent une lettre avis et ne souhaitent pas signaler qu'ils sont sensiblement différents des autres (revenu, patrimoine, pratiques sexuelles…). Il en est de même au moment de la sélection du fait de la non-réponse partielle. Même si nous disposons cette fois de plus de variables pour appréhender correctement la sélection et obtenir l'indépendance par conditionnement.

Nous présenterons différents aspects de la non-réponse et plus généralement de la sélection non-ignorable. Nous aborderons plusieurs approches proposées dans la littérature statistique. La plus ancienne est peut être celle basée sur le modèle de sélection d'Heckman. Elle permet, sous une hypothèse de loi du comportement en population générale non testable, un test de sélection ignorable et de développer une méthode d'imputation tenant compte de la sélection. Des familles paramétrées de lois plus générales peuvent être envisagées, de même qu'une généralisation à des données qualitatives ou de comptage. Enfin nous présenterons des méthodes semi-paramétriques.