

ESTIMATION D'UNE FONCTION DE RÉPARTITION POUR DE PETITS DOMAINES.

Franck ARNAUD ()*

(*) INSEE, DSDS

L'estimation dans les domaines s'intéresse à la conception d'estimateurs, lorsque peu de données sont disponibles, une fois la collecte réalisée. Il existe principalement deux approches : l'une dite *synthétique*, qui revient à considérer que la population d'étude n'est pas si différente qu'une population plus vaste pour laquelle plus d'information est disponible. Les estimations de variance permettent d'illustrer les limites de l'approximation.

L'autre approche passe par un modèle dit de *superpopulation*, ce qui revient à quitter la sphère de la théorie classique des sondages, i.e. les statistiques avec des tailles de population finies (où l'aléatoire est sur l'échantillon), au profit de la statistique classique, avec de l'aléatoire sur les observations, en raisonnant conditionnellement à l'échantillon tiré. Les estimateurs obtenus, et notamment leurs performances, dépendent énormément de la validité du modèle considéré. C'est une des raisons pour laquelle on préfère éviter de les utiliser dans les INS.

Ces deux approches ont été utilisées, historiquement la première avant la seconde, pour des statistiques linéaires, i.e. des fonctions simples de totaux, comme la moyenne, ou la covariance, mais non pas la médiane. Néanmoins, nous pouvons nous intéresser à des statistiques d'un tout autre ordre, que nous appellerons non-linéaires. Elles reposent sur la fonction de répartition ; on comprend alors bien toute la difficulté : estimer une fonction à partir de peu de points. Nous proposons une modélisation particulière de la fonction de répartition, à l'aide d'une famille exponentielle généralisée. Nous dérivons alors les estimateurs, grâce aux propriétés des familles exponentielles. Nous nous intéressons aussi à l'estimateur de variance, qui ne se laisse pas traiter aisément par des techniques standards.