

BOOTSTRAP POUR DONNEES D'ENQUETE PRINCIPES ET EXEMPLES D'APPLICATIONS

GUILLAUME CHAUVET¹, MARIA GUADARRAMA², JEAN RUBIN³

L'estimation de variance dans les enquêtes est un problème complexe. Les poids d'estimation sont le résultat de nombreux traitements statistiques (correction de non-réponse, calage), et leur impact sur la variance doit être pris en compte. Le bootstrap permet de produire des mesures d'incertitude par simulations, sans utiliser de formules analytiques de variance. Les méthodes de bootstrap pondérées sont particulièrement intéressantes, puisqu'elles permettent à des utilisateurs de produire des mesures d'incertitude sans information autre que les poids bootstrap. La méthode du « rescaled bootstrap » proposée par Rao et Wu (1988) est probablement la plus utilisée dans les enquêtes.

Le matin sera consacré à quelques rappels sur la théorie des sondages et sur l'approche bootstrap. Nous montrerons quelques utilisations possibles du bootstrap (estimation de variance, production d'intervalles de confiance, tests d'hypothèse). Nous présenterons des méthodes bootstrap spécifiquement proposées pour des enquêtes probabilistes. Nous discuterons également de solutions logicielles existantes.

L'après-midi sera consacré à des exemples d'applications du bootstrap dans des enquêtes réelles. Dans la mesure du temps disponible, nous traiterons dans l'ordre les exemples suivants :

- Comparaison d'approches pour l'estimation de variance – Application à l'enquête sur le racisme et la discrimination (Maria Guadarrama)
- Estimation de variance pour un plan de sondage produit (cross-classified sampling) – Application à l'Enquête Longitudinale Française depuis l'Enfance (Jean Rubin)
- Estimation de variance par bootstrap pour une enquête longitudinale – Application à l'enquête Histoire de Vie et Patrimoine (Guillaume Chauvet)
- Estimation de variance par bootstrap pour une enquête spatiale – Application à l'inventaire forestier français (Guillaume Chauvet)

Guillaume Chauvet est attaché principal de l'INSEE, enseignant-chercheur à l'ENSAI et membre de l'IRMAR (UMR CNRS 6625). Son travail de recherche porte sur la théorie des sondages, et en particulier les techniques d'échantillonnage et de redressement, l'estimation de variance, le traitement d'enquêtes longitudinales, et plus récemment les méthodes d'échantillonnage et d'estimation pour les inventaires forestiers.

Maria Guadarrama a un doctorat à l'Université Carlos III de Madrid dans l'estimation des petits domaines. Elle a rejoint le Luxembourg Institute of Socio-Economic Research (LISER) depuis septembre 2017 en tant que statisticienne. Elle est en charge de la méthodologie des enquêtes commandées par des Ministères ou autres entités. Elle continue aussi à faire ses travaux de recherche dans l'estimation des petits domaines ainsi que dans l'estimation de la variance sur la présence de non-réponse.

Jean Rubin est administrateur de l'INSEE, chargé d'études économiques à l'INSEE depuis septembre 2022 et doctorant à l'ENSAI depuis 2023. Ses études économiques portent sur le thème de la redistribution et des politiques sociales et ses travaux de recherches portent sur l'estimation de variance en théorie des sondages, notamment à l'aide de méthodes bootstrap.

¹Ecole Nationale de la Statistique et de l'Analyse de l'Information (ENSAI), France: chauvet@ensai.fr

²Luxemburg Institute of Socio-Economic Research (LISER), Luxembourg: maria.guarramaSanz@liser.lu

³Institut National de la Statistique et des Etudes Economiques (INSEE), France: jean.rubin@insee.fr