
Comment estimer l'effet causal en fédéré ?

Rémi Khellaf*¹

¹Inria Premedical – Institut National de Recherche en Informatique et en Automatique, Institut National de la Santé et de la Recherche Médicale - INSERM, Université de Montpellier et MBS – France

Résumé

L'inférence causale en fédéré est une approche émergente pour estimer l'effet d'un traitement à partir de données décentralisées provenant d'études ou centres différents. Cette approche décentralisée est particulièrement pertinente pour l'estimation d'effet traitement dans un cadre biomédical, où les propriétaires des données sont réticents à les partager (pour des raisons légales, de confidentialité ou de compétitivité) tout en souhaitant pouvoir collaborer avec d'autres acteurs afin d'améliorer la puissance et généralisabilité de leur analyse causale. Nous étudions trois classes d'estimateurs de l'Effet de Traitement Moyen (ATE) dérivés de l'estimateur G-Formula, allant de simples agrégations d'estimations locales ("méta-analyse") à des méthodes plus sophistiquées d'apprentissage fédéré (One-Shot et Multi-Shot) qui exploitent l'ensemble des données pour apprendre les paramètres du modèle de la variable d'intérêt (nécessitant toutefois plus de communications). Nous comparons ces estimateurs dans le cadre d'Essais Contrôlés Randomisés (RCTs) en calculant explicitement leur variance asymptotique dans le contexte des modèles linéaires. De plus, nous offrons des conseils pratiques pour sélectionner l'estimateur approprié dans divers scénarios, incluant un déséquilibre dans les tailles d'échantillons des centres, des distributions de covariables et des schémas d'assignation du traitement différents entre les centres, ainsi que la présence d'effets-centre. Enfin, nous validons nos résultats numériquement à travers une étude de simulation pour illustrer la performance des estimateurs sous ces différentes conditions.

*Intervenant