

Quelques développements nouveaux sur les approches de vraisemblance pour l'estimation des probabilités de participation à un échantillon non probabiliste

Jean-François Beaumont, Cynthia Bocci et Keven Bosa

Résumé

L'estimation des paramètres d'une population finie à l'aide de données d'échantillons non probabilistes est un sujet de recherche qui a gagné en popularité ces dernières années. Souvent, les méthodes qui ajustent les estimations d'échantillons non probabilistes utilisent des données auxiliaires collectées dans un échantillon probabiliste et nécessitent l'estimation de la probabilité de participation à l'échantillon non probabiliste.

Dans cette présentation, nous nous concentrons sur l'estimation des probabilités de participation quand elles sont modélisées de façon paramétrique, par exemple en utilisant un modèle logistique. Tout d'abord, nous proposons une amélioration de la méthode de pseudo-vraisemblance de Chen, Li et Wu (2020), basée sur la théorie de la meilleure estimation linéaire sans biais, qui exploite plus efficacement les données disponibles des échantillons probabiliste et non probabiliste. Nous développons également une approche de vraisemblance d'échantillon qui prend correctement en compte le chevauchement entre les deux échantillons lorsqu'il peut être identifié dans au moins un des deux échantillons. Nous utilisons la théorie de la meilleure prédiction linéaire sans biais pour traiter le scénario où le chevauchement est inconnu. Il est intéressant de noter que les deux approches proposées coïncident dans le cas d'un chevauchement inconnu. Nous montrons également que de nombreuses méthodes proposées dans la littérature des 15 dernières années peuvent être obtenues comme un cas particulier d'une fonction d'estimation générale sans biais. Enfin, nous présentons les résultats d'une étude par simulation qui compare plusieurs méthodes sous divers scénarios de tailles d'échantillon. Nous nous concentrons d'abord sur l'estimation de probabilités de participation et montrons ensuite quelques résultats pour l'estimation d'une moyenne de population.