

Introduction aux sondages de courbes

Camelia Goga

Université de Bourgogne Franche-Comté
Laboratoire de Mathématiques de Besançon
16 route de Gray, 25000, Besançon, FRANCE

17 septembre 2021

A l'ère des données massives, il n'est plus inhabituel d'avoir à gérer de très grandes bases de données de phénomènes temporels. Quand l'objectif est d'estimer des indicateurs simples tels que la trajectoire moyenne ou médiane ou bien encore les principaux modes de variation autour de la moyenne, capturés par l'intermédiaire d'une analyse en composantes principales, les techniques de sondage sont des approches intéressantes. Elles offrent en effet un bon compromis entre taille des données à traiter et précision de l'estimation.

Ce mini-cours a pour but de présenter une revue des approches de sondage qui ont été développées ces dernières années pour analyser de grandes bases de données fonctionnelles. Après une présentation générale du cadre statistique fonctionnel, l'accent sera mis sur la description de différentes stratégies d'échantillonnage utilisées pour estimer une trajectoire moyenne ou médiane, ou encore les éléments d'une analyse en composantes principales. Il s'agit en particulier de stratégies basées sur l'estimateur d'Horvitz-Thompson dans le cas des plans aléatoires simples sans remise, stratifiés ou proportionnels à la taille sans remise. L'estimateur assisté par un modèle linéaire fonctionnel sera également présenté et une attention spéciale sera dédiée à la construction de bandes de confiance pour la trajectoires moyenne.

Ces techniques sont illustrées sur un jeu de données de courbes de charge électriques mesurées chaque demi-heure pendant une semaine.

Notes biographiques :

Camelia Goga est professeur en statistique à l'Université de Franche-Comté, spécialité techniques de sondages et l'application de nouvelles méthodes de type machine learning en sondages.