

Variance estimation for survey estimators based on statistical learning models

Mehdi Dagdoug

McGill University

Predictive models are widely used in survey sampling. Common examples include the model-based and model-assisted frameworks, as well as the treatment of nonresponse with imputation and rereighting. The last two decades have witnessed an increasing attention of applied and theoretical statisticians towards statistical learning. Statistical learning provides new flexible tools for survey researchers, but also poses new challenges; variance estimation is one of them. In this talk, we will illustrate that traditional variance estimators often do not perform well when applied to survey estimators built from complex statistical learning procedures. We will investigate and explain the reason for these ill behaviors through connections to high-dimensional linear models. Alternative variance estimators will be suggested; their performances will be discussed with theoretical and empirical results.