Mathematics and Statistics

David Haziza

University of Ottawa

**Monday
Feb. 12, 2024**

3:30 PM

**Classroom Bldg
CL 125
and Zoom**

Statistical inference in the presence of imputed survey data through regression trees and random forests

In recent years, machine learning procedures have attracted much attention in National Statistical Offices (e.g., Statistics Canada). In particular, random forests are currently being scrutinized as an alternative to traditional imputation procedures. Item nonresponse in surveys is usually handled through some form of single imputation. Random forests provide flexible tools for obtaining a set of imputed values. Belonging to the class of non-parametric methods, random forests have the ability to capture nonlinear trends in the data and tend to be robust to the non-inclusion of interactions or predictors accounting for curvature. In this presentation, we will discuss the properties of imputed estimators based on random forests. Also, to the best of our knowledge, how to estimate the variance while accounting for sampling and nonresponse, has not been addressed in the literature. We propose a novel variance estimator based on the so-called reverse approach for variance estimation. We will present the results from a simulation study to assess the proposed methods in terms of bias and efficiency. Finally, the choice of hyper-parameters will also be discussed. Co-authors: Mehdi Dagdoug (McGill University) and Camelia Goga (Université de Bourgogne Franche Comté).

Live streamed on Zoom. Register in advance for link:
https://uregina-ca.zoom.us/meeting/register/tJEodu6hqDkrH9d6E9z56xmnY9_mxaHZ8YwA

University of Regina

Pacific Institute *for the* Mathematical Sciences