RÉSUMÉS
« Inférence statistique causale et ses applications à la génétique »
25 juillet—19 août 2016

ABSTRACTS
"Statistical Causal Inference and its Applications to Genetics"
July 25—August 19, 2016

Causal inference and the data-fusion problem

Elias Bareinboim*

eb@purdue.edu

Causal inference is usually dichotomized into two categories, experimental (Fisher, Cox, Cochran) and observational (Neyman, Rubin, Robins, Dawid, Pearl) which, by and large, are studied separately. Reality is more demanding. Experimental and observational studies are but two extremes of a rich spectrum of research designs that generate the bulk of the data available in practical, large scale situations. In typical medical explorations, for example, data from multiple observations and experiments are collected, coming from distinct experimental setups, different sampling conditions, and heterogeneous populations. In this talk, I will discuss some of the latest results in the field of causal inference related to big data.

In particular, I will introduce the data-fusion problem, which is concerned with piecing together multiple datasets collected under heterogeneous conditions (to be defined) so as to obtain valid answers to queries of interest. The availability of multiple heterogeneous datasets presents new opportunities to big data analysts since the knowledge that can be acquired from combined data would not be possible from any individual source alone. However, the biases that emerge in heterogeneous environments require new analytical tools. Some of these biases, including confounding, sampling selection, and cross-population biases, have been addressed in isolation, largely in restricted parametric models. I will present my work on a general, non-parametric framework for handling these biases and, ultimately, a theoretical solution to the problem of data-fusion in causal inference tasks.

*Department of Computer Science, Purdue University, 305 N. University Street 2142L, West Lafayette, IN 47907-2107, USA