



Department of Decision Sciences

Statistics Seminar

A Principled Experimental Design Approach to Big Data Analysis

Kerry Mengersen

Queensland University of Technology

Thursday, 9th May 2016

12:30pm Room 3-E4-SR03 Via Röntgen 1 Milano

Abstract

Big Datasets are endemic, but they are often notoriously difficult to analyse because of their size, complexity, history and quality. The purpose of this paper is to open a discourse on the use of modern experimental design methods to analyse Big Data in order to answer particular questions of interest. By appeal to a range of examples, it is suggested that this perspective on Big Data modelling and analysis has wide generality and advantageous inferential and computational properties. In particular, the principled experimental design approach is shown to provide a flexible framework for analysis that, for certain classes of objectives and utility functions, delivers equivalent answers compared with analyses of the full dataset. It can also provide a formalised method for iterative parameter estimation, model checking, identification of data gaps and evaluation of data quality. Finally it has the potential to add value to other Big Data sampling algorithms, in particular divide-and-conquer strategies, by determining efficient sub-samples.

with coauthors Christopher Drovandi, Christopher Holmes, James McGree, Elizabeth Ryan and Silvia Richardson