

*On Computational Thinking, Inferential Thinking and "Big Data"*

Michael I. Jordan

University of California, Berkeley

The rapid growth in the size and scope of datasets in science and technology has created a need for novel foundational perspectives on data analysis that blend the inferential and computational sciences. That classical perspectives from these fields are not adequate to address emerging problems in "Big Data" is apparent from their sharply divergent nature at an elementary level--- in computer science, the growth of the number of data points is a source of "complexity" that must be tamed via algorithms or hardware, whereas in statistics, the growth of the number of data points is a source of "simplicity" in that inferences are generally stronger and asymptotic results can be invoked. I aim to set the stage for the discussion of causality and Big Data by surveying recent progress at the computation/statistics interface, including fundamental tradeoffs between inferential quality, communication, runtime and privacy constraints, and mechanisms for implementing these tradeoffs, such as algorithmic weakening, subsampling and concurrency control.