

Abstract

Dr. Elchanan Mossel proposes to study hierarchical generating models, which model such data-generation processes in mathematically precise terms.

Many types of data are generated via processes that are only partially observed. When we speak, we refine a high-level concept repeatedly, obtaining along the way sentences, words, syllables, and sounds that are finally communicated. When we draw a picture, we begin with a high-level concept that is repeatedly refined as we draw at a finer quality and with more localized attention. Biological data is often generated in a top-down fashion. The current-day genetic composition was obtained from previous generations by mathematically well-defined processes of recombination and mutation. Similarly, the evolution of social networks is influenced both by the observed parts of the network and by sources of influences that are unobserved.

Dr. Mossel aims to develop a mathematical theory of hierarchical models which has the following properties: (1) Realism. The models represent well the data. (2) Reconstruction. There are provably efficient algorithms for the reconstruction of models given the observed data and (3) Depth. There is a mathematical proof that the hidden layers of the process are necessary for inference. Our approach for studying properties of these models and their inference is based on the mathematical theory of the underlying models. This theory builds on modern tools from probability, combinatorics, analysis, and algorithms.