

Testing and correcting distributions over big domains

Ronitt Rubinfeld, Massachusetts Institute of Technology

We survey a body of work on the complexity of testing global properties of discrete distributions, when given access to only a few samples from the distribution. Such properties might include testing if two distributions have small statistical distance, testing various independence properties, testing whether a distribution has a specific shape (such as monotone decreasing, k-modal, k-histogram, monotone hazard rate,...), and approximating the entropy. We describe bounds for such testing problems whose sample complexities are sublinear in the size of the support. We then describe a model for "correcting" sample data that is obtained from a noisy or imperfect source in an "on-the-fly" manner.