Efficient Algorithms via Precision Sampling

Robert Krauthgamer, Weizmann Institute

Abstract

Randomized estimation algorithms are key to accurate yet efficient processing of massive data sets. In particular, it is common to have limited or costly access to n bounded reals a_1,...,a_n and seek an estimate for their sum. In a classical scenario, where reading a value a_i incurs a unit cost, a simple estimator based on sampling achieves a well-known tradeoff between the number of a_i's that are read and the estimator's error.

Motivated by applications where this tradeoff is not good enough, we introduce Precision Sampling, an estimation technique that takes advantage of a more refined cost model, where the cost of reading a_i within (any desired) precision p_i is proportional to p_i.

This technique leads to new sublinear algorithms, including simpler low-space algorithms in the data-stream model, and a query-efficient algorithm for approximating edit distance.

Joint work with Alexandr Andoni and Krzysztof Onak.