

A Polynomial Time Algorithm for Lossy Population Recovery

Ankur Moitra, Massachusetts Institute of Technology

We give a polynomial time algorithm for the lossy population recovery problem. In this problem, the goal is to approximately learn an unknown distribution on binary strings of length n from lossy samples: for some parameter μ each coordinate of the sample is preserved with probability μ and otherwise is replaced by a '?'. The running time and number of samples needed for our algorithm is polynomial in n and $1/\epsilon$ for each fixed $\mu > 0$. This improves on algorithm of Wigderson and Yehudayoff that runs in quasi-polynomial time for any $\mu > 0$ and the polynomial time algorithm of Dvir et al which was shown to work for $\mu > 0.30$ by Batman et al.

The heart of the matter is that we are faced with a certain inverse problem $Ax = b$, where we are given a noisy approximation to b and the condition number of A is exponentially large. Yet, because we are interested in a single component of x , we can solve this problem despite the fact that it is not well conditioned. Hence, the condition number is not always a barrier to solving statistical inverse problems. Our approach is based on the notion of a robust local inverse (which was introduced in previous work) and we analyze our construction using tools from complex analysis.