Robust estimation of discrete distributions under local differential privacy

Flore Sentenac*[1] and Julien Chhor*†

[1]HEC – Centre de Recherche en Économie et STatistique (CREST) – France

Résumé

In the realms of insurance and finance, minimizing the risk of model failure is of paramount importance. Robust learning methodologies have been extensively explored in these domains. More recently, spurred by regulatory requirements and societal considerations, significant emphasis has been placed on ensuring that algorithms prioritize user privacy.
Although robust learning and local differential privacy are both widely studied fields of research, combining the two settings is just starting to be explored. We consider the problem of estimating a discrete distribution in total variation from n contaminated data batches under a local differential privacy constraint.

A fraction $1-\alpha$ of the batches contain k i.i.d. samples drawn from a discrete distribution p over d elements. To protect the users' privacy, each of the samples is privatized using an ϵ-locally differentially private mechanism. The remaining αn batches are an adversarial contamination. The minimax rate of estimation under contamination alone, with no privacy, is known to be $\alpha/\sqrt{k} + \sqrt{d/kn}$. Under the privacy constraint alone, the minimax rate of estimation is $\sqrt{d^2/\epsilon^2 kn}$. We show, up to a $\sqrt{\log(1/\alpha)}$ factor, that combining the two constraints leads to a minimax estimation rate of $\alpha\sqrt{d/\epsilon^2 k} + \sqrt{d^2/\epsilon^2 kn}$, larger than the sum of the two separate rates.
We provide a polynomial-time algorithm achieving this bound, as well as a matching information theoretic lower bound.

Mots-Clés: Privacy, Robust

*Intervenant
†Auteur correspondant: julien.chhor@ensae.fr