Discussion on "Minimax Optimal Procedures for Locally Private Estimation"

Anderson Y. Zhang & Harrison H. Zhou
Department of Statistics and Data Science, Yale University

November 27, 2017

We congratulate Professors Duchi, Jordan and Wainwright on their path-breaking work in statistical decision theory and privacy. Their extension of classical information-theoretic lower bounds of Le Cam, Fano, and Assouad to local differential privacy can potentially lead to a systematic study of various lower bounds under all kinds of privacy constraints. Their successful treatments of some interesting problems in the paper shed light on possibly a unified theory for a general statistical framework.

Computer Science and Statistics. The discipline of computer science has achieved remarkable progress recently and has exerted continuous and increasing influence on statistics. In *Rise of the Machines* [Wasserman, 2014], Professor Larry Wasserman writes,

> *"There are many statistical topics that are dominated by ML and mostly ignored by statistics. This is a shame because statistics has much to offer in all these areas. Examples include semi-supervised inference, computational topology, online learning, sequential game theory, hashing, active learning, deep learning, differential privacy, random projections and reproducing kernel Hilbert spaces."*

Some of the aforementioned topics have deep roots in statistics. They have been studied by statisticians for years and popularized in machine learning. The main topic of this paper, local differential privacy, will likely be among these topics. It was proposed in Warner [1965] for survey sampling, but it is becoming increasingly important in the big data era.

Decision Theory. The optimality study under a privacy constraint can be seen as a special case of constrained minimax analysis. The minimax theory lies at the heart of decision theory, which studies the difficulty and fundamental limits of various statistical tasks. The classical minimax analysis is often criticized for being both over-pessimistic and over-optimistic. It is pessimistic because it quantifies the performance of procedures by the least favorable case; on the other hand, it is optimistic because all procedures are considered, even those that are not feasible in practice. In spite of the existence of rich and abundant philosophical discussions and literature on the former pessimism of minimax theory, the latter optimism receives little attention and few investigations. But in practice procedures are often restricted for various reasons including privacy, computation, and communication.

For most statistical problems, we have observations X generated from some underlying model parameterized by θ from a parameter space Θ. The task is to estimate the unknown parameter θ from the data. The evaluation is carried out through some loss function $\ell(\cdot,\cdot)$, and the statistical hardness of the problem is measured by

$$\inf_{\hat{\theta}} \sup_{\theta\in\Theta} \mathbb{E}\ell(\hat{\theta}(X),\theta). \tag{1}$$

This is the standard minimax formulation. Due to constraints over $\hat{\theta}$, it is entirely possible that the minimax risk can never be attained in practice.

In the constrained minimax analysis, the estimator $\hat{\theta}$ is restricted to satisfy certain properties. It can be formulated in a way like Equation (1) as follows

$$\inf_{\hat{\theta}\in\mathcal{S}} \sup_{\theta\in\Theta} \mathbb{E}\ell(\hat{\theta}(X),\theta), \tag{2}$$

where the space $\mathcal{S}$ may include only algorithms under certain constraints such as: 1) privacy; 2) polynomial-time; 3) convex; or 4) computational resources (for example, storage constraint) [Zhu and Lafferty, 2017]. In some special cases, the space $\mathcal{S}$ may be restricted so that $\hat{\theta} = \tilde{\theta}\circ Q$ where Q is a mapping from X to Y and $\tilde{\theta}$ is an estimator on Y. In other words, it can be represented in the following diagram:

$$\theta \to X \xrightarrow{Q} Y \xrightarrow{\tilde{\theta}} \hat{\theta}.$$

Equation (2) then becomes

$$\inf_{Q\in\mathcal{Q}} \inf_{\tilde{\theta}} \sup_{\theta\in\Theta} \mathbb{E}\ell(\tilde{\theta}(Q(X)),\theta), \tag{3}$$

where the space $\mathcal{Q}$ may contain all mappings from X to Y that 1) preserve the privacy as considered in this paper or 2) meet certain communication requirements for distributed computation [Zhang et al., 2013].

Equations (2) and (3) are generalizations of the classical minimaxity. They provide statistically meaningful ways for studying constrained tasks. It would be very interesting, although possibly extremely challenging, to have a systematic study of constrained minimax theory, at least for some important spaces $\mathcal{S}$ and $\mathcal{Q}$.

Privacy. In this era of big data, privacy is becoming very important. Statisticians and data scientists ought to extract knowledge or insights from data, and hope that little personal identity or sensitive information is unveiled. There is a trade-off between statistical accuracy and privacy. The authors of this paper investigated this interplay under the *α-differentially local privacy*. New technical tools were developed in the paper. For example, the authors obtained the private versions of Le Cam's two-point hypothesis testing, Fano's lemma, and Assouad's method which are the cornerstones of establishing minimax lower bound. They also obtained sharp *α-private minimax rates* under various settings and proposed some mechanisms to attain them. Again we congratulate the authors on those exciting achievements, which open the door to many avenues of research ahead.

- *Centralized Privacy.* The private channel Q considered in this paper essentially operates on each data point of $X = (x_1, x_2, \ldots, x_n)$ individually. Since for each data point the mapping is α-differentially privacy-preserving, the channel Q satisfies α-differential privacy globally. It is more popular and less restrictive to quantify privacy globally. In most literature (e.g., Dwork and Roth [2014]), *α-differential privacy* is defined in a *centralized* sense,

$$\sup_A \frac{\mathbb{Q}(Q(X) \in A|X)}{\mathbb{Q}(Q(X') \in A|X')} \leq \exp(\alpha), \forall X, X' \text{ s.t. } H(X, X') = 1, \tag{4}$$

  where $H(\cdot, \cdot)$ measures how many data points differ in two sets. It will be interesting to see if the conclusions in this paper will be changed when the definition of privacy is shifted from *local* to its *centralized* counterpart. For example, the authors point out that the effect of local α-differential privacy is to reduce the effective sample size from n to $\alpha^2 n$ under several scenarios. But does the same reduction hold true if the centralized differential privacy is considered instead? Similar questions can be raised for other privacy constraints such as (α, δ)-differential privacy introduced in Dwork et al. [2006].

- *General Settings.* The authors obtained sharp private minimax rates for various statistical tasks, including mean/median estimation, logistic regression, nonparametric density estimation, etc. Though only the simplest cases were investigated, this paper successfully illustrates the effect of privacy constraint on the minimax rates and the potential difficulties in the theoretical analysis. It is of great value and interest to go beyond these basic cases to see how privacy-preserving minimax rates behave under more sophisticated and complex settings. For instance, the authors showed that for d-dimensional bounded mean estimation, the α-private minimax rate is proportional to the dimensionality d, which is different to the classical one. The same phenomenon was observed for high-dimensional parameter estimation with sparsity $s = 1$. A follow-up question is whether the existence of this extra d factor is universal. If so, it will be fascinating to have a unified theory depending only on complexity and dimensionality for a general class of statistical models including high-dimensional linear regression for arbitrary sparsity s. To achieve this, we will likely need a very sophisticated extension of private versions of lower bounds presented in this paper.

References

Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I. and Naor, M. [2006], Our data, ourselves: Privacy via distributed noise generation., *in* 'Eurocrypt', Vol. 4004, Springer, pp. 486–503.

Dwork, C. and Roth, A. [2014], 'The algorithmic foundations of differential privacy', *Foundations and Trends® in Theoretical Computer Science* **9**(3–4), 211–407.

Warner, S. L. [1965], 'Randomized response: A survey technique for eliminating evasive answer bias', *Journal of the American Statistical Association* **60**(309), 63–69.

Wasserman, L. [2014], 'Rise of the machines', *Past, present, and future of statistical science* pp. 1–12.

Zhang, Y., Duchi, J., Jordan, M. I. and Wainwright, M. J. [2013], Information-theoretic lower bounds for distributed statistical estimation with communication constraints, *in* 'Advances in Neural Information Processing Systems', pp. 2328–2336.

Zhu, Y. and Lafferty, J. [2017], 'Quantized minimax estimation over sobolev ellipsoids', *Information and Inference: A Journal of the IMA* .