Kasiviswanathan, S. P., Lee, H. K., Nissim, K., Raskhodnikova, S., and Smith, A. (2011), "What Can We Learn Privately?" *SIAM Journal on Computing*, 40, 793–826. [182]

King, G., Pan, J., and Roberts, M. (2013), "How Censorship in China Allows Government Criticism But Silences Collective Expression," *American Political Science Review*, 107, 1–18. [199]

——— (2014), "Reverse-Engineering Censorship in China: Randomized Experimentation and Participant Observation," *Science*, 345. doi:10.1126/science.1251722. [199]

Le Cam, L. (1973), "Convergence of Estimates Under Dimensionality Restrictions," *Annals of Statistics*, 1, 38–53. [185]

Lehmann, E. L. (1999), *Elements of Large Sample Theory*, New York: Springer. [194]

Levy, R., and Manning, C. D. (2003), "Is It Harder to Parse Chinese, or the Chinese Treebank?" in *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*, Association for Computational Linguistics, pp. 439–446. [199]

Loh, P.-L., and Wainwright, M. J. (2012), "High-Dimensional Regression With Noisy and Missing Data: Provable Guarantees With Nonconvexity," *Annals of Statistics*, 40, 1637–1664. [193]

Massey, J. (1990), "Causality, Feedback and Directed Information," in *Proceedings of the International Symposium on Information Theory and its Applications (ISITA)*, pp. 303–305. [200]

McGregor, A., Mironov, I., Pitassi, T., Reingold, O., Talwar, K., and Vadhan, S. (2010), "The Limits of Two-Party Differential Privacy," in *51st Annual Symposium on Foundations of Computer Science*, IEEE, pp. 81–90. [188]

Negahban, S., Ravikumar, P., Wainwright, M., and Yu, B. (2012), "A Unified Framework for High-Dimensional Analysis of *M*-Estimators With Decomposable Regularizers," *Statistical Science*, 27, 538–557. [190]

Nemirovski, A., Juditsky, A., Lan, G., and Shapiro, A. (2009), "Robust Stochastic Approximation Approach to Stochastic Programming," *SIAM Journal on Optimization*, 19, 1574–1609. [187,193]

Permuter, H. H., Kim, Y.-H., and Weissman, T. (2011), "Interpretations of Directed Information in Portfolio Theory, Data Compression, and Hypothesis Testing," *IEEE Transactions on Information Theory*, 57, 3248–3259. [200]

Polyak, B. T., and Juditsky, A. B. (1992), "Acceleration of Stochastic Approximation by Averaging," *SIAM Journal on Control and Optimization*, 30, 838–855. [184,193]

Rubinstein, B. I. P., Bartlett, P. L., Huang, L., and Taft, N. (2012), "Learning in a Large Function Space: Privacy-Preserving Mechanisms for SVM Learning," *Journal of Privacy and Confidentiality*, 4, 65–100. [182]

Scott, D. (1979), "On Optimal and Data-Based Histograms," *Biometrika*, 66, 605–610. [195]

Smith, A. (2011), "Privacy-Preserving Statistical Estimation With Optimal Convergence Rates," in *Proceedings of the Forty-Third Annual ACM Symposium on the Theory of Computing*, ACM, pp. 813–822. [182,183]

Substance Abuse, Mental Health Services Administration. (2013), "Drug Abuse Warning Network 2011: National Estimates of Drug-Related Emergency Department Visits," Technical Report SMA 13-4760, U.S. Department of Health and Human Services. Available at *http://www.samhsa.gov/data/emergency-department-data-dawn/reports?tab=47* [198]

Tsybakov, A. B. (2009), *Introduction to Nonparametric Estimation*, New York: Springer. [185,194,195]

U. of California., (2010), "Annual Report on Employee Compensation, 2014," Available at *http://compensation.universityofcalifornia.edu/payroll2014/*. [196]

Warner, S. (1965), "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias," *Journal of the American Statistical Association*, 60, 63–69. [183,185,187]

Wasserman, L., and Zhou, S. (2010), "A Statistical Framework for Differential Privacy," *Journal of the American Statistical Association*, 105, 375–389. [182,185,195]

Yang, Y., and Barron, A. (1999), "Information-Theoretic Determination of Minimax Rates of Convergence," *Annals of Statistics*, 27, 1564–1599. [194]

Yang, X., Fienberg, S. E., and Rinaldo, A. (2012) "Differential Privacy for Protecting Multi-dimensional Contingency Table Data: Extensions and Applications," *Journal of Privacy and Confidentiality*, 4, Article 5. [183]

Yu, B. (1997), "Assouad, Fano, and Le Cam," in *Festschrift for Lucien Le Cam*, eds. D. Pollard, G. Yang, E. Torgersøn, New York: Springer-Verlag, pp. 423–435. [183,185,188,192,194]

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2018, VOL. 113, NO. 521, Theory and Methods
https://doi.org/10.1080/01621459.2018.1442605

Comment

Anderson Y. Zhang and Harrison H. Zhou

Department of Statistics and Data Science, Yale University, New Haven, CT

We congratulate Professors Duchi, Jordan, and Wainwright on their path-breaking work in statistical decision theory and privacy. Their extension of classical information-theoretical 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 article 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 ig- nored 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 article, local

CONTACT Harrison H. Zhou *huibin.zhou@yale.edu* Department of Statistics and Data Science, Yale University, New Haven, CT 06511.

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 (e.g., 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:

$$\theta \rightarrow X \overset{Q}{\rightarrow} Y \overset{\tilde{\theta}}{\rightarrow} \hat{\theta}.$$

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 article 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 article investigated this interplay under the *α-differentially local privacy*. New technical tools were developed in the article. 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 article 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 article 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 article 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 article.

References

Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I., and Naor, M. (2006), "Our Data, Ourselves: Privacy Via Distributed Noise Generation," in *Eurocrypt*, ed. S.Vaudenay, New York: Springer, vol. 4004, pp. 486–503. [202]

Dwork, C., and Roth, A. (2014), "The Algorithmic Foundations of Differential Privacy," *Foundations and Trends® in Theoretical Computer Science*, 9, 211–407. [202]

Warner, S. L. (1965), "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias," *Journal of the American Statistical Association*, 60, 63–69. [202]

Wasserman, L. (2014), "Rise of the Machines," *Past, Present, and Future of Statistical Science*, 1–12. [201]

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*, eds. C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, pp. 2328–2336. [202]

Zhu, Y., and Lafferty, J. (2017), "Quantized Minimax Estimation Over Sobolev Ellipsoids," *Information and Inference: A Journal of the IMA*. [202]

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2018, VOL. 113, NO. 521, Theory and Methods
https://doi.org/10.1080/01621459.2018.1442606

Comment

Alfred Hero

University of Michigan, Ann Arbor, MI

It is a privilege to be participating in the discussion of this interesting article, which offers a comprehensive study of fundamental tradeoffs between privacy and statistical estimation. Using the differential local privacy (DLP) measure introduced by Duchi, Jordan, and Wainwright (2013), the authors develop private versions of several classical bounds on estimator accuracy, quantifying the effect of different levels of DLP on minimax estimator performance. These bounds are illustrated for important estimation problems including mean and median estimation, estimation in generalized linear models, and density estimation. As these bounds are minimax, the authors are able to obtain optimal privacy mechanisms, that is, manipulations of the data that achieve the bounds. Minimax approaches are sometimes criticized in statistical estimation for being overly conservative and for leading to corner results that are not relevant to applications. For the privacy problem considered in this article, the minimax approach is natural, capturing the intrinsic conflict between the user's desire to maintain privacy versus the statistician's objective to maximize estimator accuracy. However, the severe privacy-induced degradation in estimator accuracy in high dimension d is troubling. One is tempted to take solace in the fact that the relevant structure of many high-dimensional datasets substantially lies in a space of much lower (intrinsic) dimension. However, this lower dimensional space is generally unknown and should be properly considered as part of the privatized estimation problem. Hence, as pointed out by the authors in the article's conclusion, this degradation is a strong motivation for pursuing weaker mechanisms of privacy than minimax optimality. I discuss a few additional points below.

The proposed DLP framework is deftly demonstrated to be amenable to analysis, specifying optimal mechanisms that are explicit and simple to implement. Several optimal mechanisms are obtained, each depending on the specific cost function that the statistician uses to compute risk. However, a single risk function may not apply to the full lifecycle of the data. Indeed, a dataset may be reused and repurposed for different objectives, each leading the statistician to use a different notion of risk. Thus, it would be interesting to consider a multiobjective extension of the proposed DLP approach, possibly leading to optimal mechanisms with respect to a plurality of cost functions. One can imagine several approaches to such an extension. For example, multiobjective scalarization by linear combining could be used to obtain optimal mechanisms relative to a given set of linear coefficients. Or, in cases where the convex hull of the multiobjective mechanisms is not sufficient, a Pareto-optimal theory of DLP risk analysis could be pursued. Alternatively, using the DLP framework one could explore sequential mechanism design for data that undergo a sequence of "data reuse" stages. For example, for a given set of cost functions, one could explore the construction of an iterative sequence of mechanism compositions achieving good estimation performance under arbitrary sequential reordering of the cost functions. Recent work on k-fold composition theory for differential privacy developed by Kairouz, Oh, and Viswanath (2017) could be relevant here.

The authors make an interesting conjecture on whether or not their DLP-constrained mutual information (MI) inequality (22) holds in the fully interactive setting. As the authors mention, the correctness of their conjecture depends on the extendability of this inequality to the case of channels with feedback. In feedback channels, as the authors point out, directed information (DI) introduced by Massey (1990) may provide a path to the desired extension using results relating information measures to estimation measures. For example, the classical result of Duncan relates the MI to the integrated mean-squared error (MSE) of an optimal non-causal predictor in a Gaussian channel without feedback. This has been generalized to causal

CONTACT Alfred Hero ✉ hero@eecs.umich.edu University of Michigan, Ann Arbor MI 48109, USA.