Errata on the paper "Geometric Lower Bound for Distributed Parameter Estimation under Communication Constraints"

Yanjun Han, Ayfer Özgür, and Tsachy Weissman

Abstract

In this errata we point out an error in the paper [HÖW18] and its short version [HMÖW18] in handling sequential and blackboard protocols, and present a modifcation of the proof to fix this error.

This errata is devoted to handling one technical error in [HÖW18] when proving lower bounds for interactive (including sequential and blackboard) communication protocols. Specifically, the following upper bound of mutual information is shown in [HÖW18, Equation 6]:

$$I(U;Y) \le \mathbb{E}_U \sum_{i=1}^n \sum_{y\in\{0,1\}^{nk}} \left(\prod_{j\neq i} \mathbb{E}_{P_U}[p_{j,y}(X_j)]\right) \cdot \frac{(\mathbb{E}_{P_U}[p_{i,y}(X_i)] - \mathbb{E}_{P_0}[p_{i,y}(X_i)])^2}{\mathbb{E}_{P_0}[p_{i,y}(X_i)]}. \qquad (1)$$

Based on (1), the subsequent analysis made the error that [HÖW18, Lemma 12] reduced to the upper bound of the quantities

$$\mathbb{E}_U \frac{(\mathbb{E}_{P_U}[p_{i,y}(X_i)] - \mathbb{E}_{P_0}[p_{i,y}(X_i)])^2}{\mathbb{E}_{P_0}[p_{i,y}(X_i)]}$$

with weights $w_{i,y}$ summing into 2^k. However, this step is not feasible as the probability measure P_U in (1), marked in blue, actually depends on U. Note that if P_U is replaced by any probability measure independent of U, or the simultaneous communication protocol is assumed, this reduction becomes valid again. A similar error also occurs in [HMÖW18] (above Equation 6) for interactive protocols.

We present two approaches to fix the above error. The first approach is to consider a continuous prior on U and invoke the van-trees inequality (instead of a discrete prior in the Fano-type arguments in (1)); the details of this approach are referred to the subsequent work [BHÖ19]. The second approach is to use Assouad's lemma instead of Fano's inequality to carry out the lower bound arguments, which is detailed below.

We first recall the problem setup and define some notations. Let $\theta = (\theta_1, \cdots, \theta_d) \in \Theta$ be the d-dimensional unknown parameter, i.i.d. samples $X_1, \cdots, X_n \sim P_\theta$ be drawn from the statistical model P_θ where each node has one observation, and $Y \in \{0,1\}^{nk}$ be the transcript outputted by the blackboard communication protocol with k bits of communication constraint. To construct the hypotheses required for the lower bound, we choose an interior point $\theta_0 \in \Theta$, and some scalar accuracy parameter $\delta > 0$ to be specified later. Let $U \sim \mathsf{Unif}(\{\pm 1\}^d)$, and P_U be the shorthand of the distribution $P_{\theta_0+\delta U}$, Q_U be the distribution of the transcript Y given P_U. The above setup is the same as [HÖW18, Section 3].

We shall need an additional notation and assumption. For each binary vector $u \in \{\pm 1\}^d$ and $j \in [d]$, let $u^{\oplus j} \in \{\pm 1\}^d$ be the binary vector obtained by flipping the j-th coordinate of u. The following assumption assumes a bounded likelihood ratio between neighboring distributions.

Assumption 1. *For each $u \in \{\pm 1\}^d$ and $j \in [d]$, it holds that $dP_{u^{\oplus j}}/dP_u \ge 1/2$ almost surely.*

We remark that Assumption 1 is not very restrictive as the parameters $\theta_0 + \delta u$ and $\theta_0 + \delta u^{\oplus j}$ only differ in the j-th coordinate. Later we will see that this assumption holds in the discrete distribution estimation model, and a slight modification of P_u makes the assumption work in the Gaussian location model as well.

Now for any estimator $\widehat{\theta}(Y)$ of θ based on the final transcript Y, one version of the Assouad's lemma [Ass83] states that

$$\mathbb{E}_U \mathbb{E}_{Q_U} \|\widehat{\theta}(Y) - \theta_U\|_2^2 \ge \frac{d\delta^2}{2}\left(1 - \frac{1}{d}\sum_{j=1}^d \mathbb{E}_U \|Q_U - Q_{U^{\oplus j}}\|_{\mathrm{TV}}\right). \tag{2}$$

Furthermore, by Pinsker's inequality $\|Q_U - Q_{U^{\oplus j}}\|_{\mathrm{TV}} \le \sqrt{D_{\mathrm{KL}}(Q_U \| Q_{U^{\oplus j}})/2}$ and Jensen's inequality applied to the convex function $x \mapsto -\sqrt{x}$, the inequality (2) implies

$$\mathbb{E}_U \mathbb{E}_{Q_U} \|\widehat{\theta}(Y) - \theta_U\|_2^2 \ge \frac{d\delta^2}{2}\left(1 - \sqrt{\frac{1}{d}\sum_{j=1}^d \mathbb{E}_U [D_{\mathrm{KL}}(Q_U \| Q_{U^{\oplus j}})]}\right). \tag{3}$$

The usage of the inequality (3) is partially motivated by [ACL+20].

To proceed, we recall the following representation of the distribution Q_U in [HÖW18, Lemma 9]: for each transcript $y \in \{0,1\}^{nk}$, there exists non-negative functions $p_{i,y}(x)$ such that

$$Q_U(y) = \prod_{i=1}^n \mathbb{E}_{X_i \sim P_U}[p_{i,y}(X_i)]. \tag{4}$$

Consequently,

$$\begin{aligned}
D_{\mathrm{KL}}(Q_U \| Q_{U^{\oplus j}}) &= \sum_{i=1}^n \sum_{y \in \{0,1\}^{nk}} \left(\prod_{s=1}^n \mathbb{E}_{X_s \sim P_U}[p_{s,y}(X_s)]\right) \cdot \log \frac{\mathbb{E}_{X_i \sim P_U}[p_{i,y}(X_i)]}{\mathbb{E}_{X_i \sim P_{U^{\oplus j}}}[p_{i,y}(X_i)]} \\
&\overset{\text{(a)}}{\le} \sum_{i=1}^n \sum_{y \in \{0,1\}^{nk}} \left(\prod_{s=1}^n \mathbb{E}_{X_s \sim P_U}[p_{s,y}(X_s)]\right) \cdot \left(\frac{\mathbb{E}_{X_i \sim P_U}[p_{i,y}(X_i)]}{\mathbb{E}_{X_i \sim P_{U^{\oplus j}}}[p_{i,y}(X_i)]} - 1\right) \\
&\overset{\text{(b)}}{=} \sum_{i=1}^n \sum_{y \in \{0,1\}^{nk}} \left(\prod_{s \ne i} \mathbb{E}_{X_s \sim P_U}[p_{s,y}(X_s)]\right) \cdot \left(\frac{(\mathbb{E}_{X_i \sim P_U}[p_{i,y}(X_i)] - \mathbb{E}_{X_i \sim P_{U^{\oplus j}}}[p_{i,y}(X_i)])^2}{\mathbb{E}_{X_i \sim P_{U^{\oplus j}}}[p_{i,y}(X_i)]}\right) \\
&\overset{\text{(c)}}{\le} 2\sum_{i=1}^n \sum_{y \in \{0,1\}^{nk}} \left(\prod_{s \ne i} \mathbb{E}_{X_s \sim P_U}[p_{s,y}(X_s)]\right) \cdot \left(\frac{(\mathbb{E}_{X_i \sim P_U}[p_{i,y}(X_i)] - \mathbb{E}_{X_i \sim P_{U^{\oplus j}}}[p_{i,y}(X_i)])^2}{\mathbb{E}_{X_i \sim P_U}[p_{i,y}(X_i)]}\right) \\
&\overset{\text{(d)}}{=} 2\sum_{i=1}^n \sum_{y \in \{0,1\}^{nk}} \left(\prod_{s \ne i} \mathbb{E}_{X_s \sim P_U}[p_{s,y}(X_s)]\right) \cdot \frac{(\mathbb{E}_{X_i \sim P_U}[p_{i,y}(X_i)(1 - dP_{U^{\oplus j}}/dP_U(X_i))])^2}{\mathbb{E}_{X_i \sim P_U}[p_{i,y}(X_i)]}
\end{aligned}$$

where (a) is due to the inequality $\log x \le x - 1$, (b) follows from the identity

$$\sum_{y\in\{0,1\}^{nk}} \prod_{s=1}^{n} \mathbb{E}_{X_s\sim P_U}[p_{s,y}(X_s)] = \sum_{y\in\{0,1\}^{nk}} \mathbb{E}_{X_i\sim P_{U\oplus j}}[p_{i,y}(X_i)] \cdot \prod_{s\neq i} \mathbb{E}_{X_s\sim P_U}[p_{s,y}(X_s)] = 1$$

given by [HÖW18, Lemma 10], (c) is due to Assumption 1, and (d) follows from a simple change of measure. Consequently, for each realization of U we have

$$\frac{1}{d}\sum_{j=1}^{d} D_{\mathrm{KL}}(Q_U \| Q_{U\oplus j}) \le \frac{2}{d}\sum_{i=1}^{n}\sum_{y\in\{0,1\}^{nk}} \left(\prod_{s\neq i} \mathbb{E}_{X_s\sim P_U}[p_{s,y}(X_s)]\right) \cdot \frac{\|\mathbb{E}_{X_i\sim P_U}[p_{i,y}(X_i)s_U(X_i)]\|_2^2}{\mathbb{E}_{X_i\sim P_U}[p_{i,y}(X_i)]}, \tag{5}$$

where $s_U(x)$ is a d-dimensional vector of likelihood ratios:

$$s_U(x) \triangleq \left(1 - \frac{dP_{U\oplus 1}}{dP_U}(x), \cdots, 1 - \frac{dP_{U\oplus d}}{dP_U}(x)\right).$$

Note that in typical scenarios where each P_U is a product distribution, the random vector $s_U(X)$ with $X \sim P_U$ will have independent components, and we essentially reduce to the results involving score functions. Therefore, the final step is to apply the arguments of [HÖW18, Lemma 12] and the geometric inequalities to obtain an upper bound of (5) for each U, which are referred to the rest of the paper [HÖW18].

Finally we briefly discuss the validity of Assumption 1. In the discrete distribution estimation, we finally choose $\theta_0 = (1/d, \cdots, 1/d)$ and $\delta = O(1/\sqrt{n2^k})$. Since it is assumed that $n2^k \ge d^2$, we have $\delta = O(1/d)$, and Assumption 1 clearly holds upon choosing a small hidden constant. For the Gaussian location model, we finally choose $\theta_0 = (0, \cdots, 0)$ and $\delta = O(\sigma\sqrt{d/(nk)})$. Although the likelihood ratio between Gaussian distributions is unbounded, we could consider a modification $\widetilde{P}_u$ of P_u which is the restriction of P_u to the ℓ_∞ cube $\{\theta : \|\theta\|_\infty \le c\sigma\sqrt{\log d}\}$. Note that for a large constant $c > 0$, we have $\|P_u - \widetilde{P}_u\|_{\mathrm{TV}} = O(d^{-10})$ for each $u \in \{\pm 1\}^d$, therefore by the union bound, replacing P_u by $\widetilde{P}_u$ results in an additional estimation error at most $O(n/d^{10})$, which is negligible for moderate ranges of d. After the restriction, straightforward algebra shows that the likelihood ratio is at least $\exp(-O(\delta\sqrt{\log d}/\sigma + \delta^2/\sigma^2))$, which is at least $1/2$ by the choice of δ and the assumption $nk \ge d^2$.

Acknowledgement

The authors are grateful to Jayadev Acharya, Clément Canonne, Himanshu Tyagi, and Rishabh Dudeja for spotting this error and communicating with us.

References

[ACL+20] Jayadev Acharya, Clément L Canonne, Yuhan Liu, Ziteng Sun, and Himanshu Tyagi. Interactive inference under information constraints. *arXiv preprint arXiv:2007.10976*, 2020.

[Ass83] Patrice Assouad. Deux remarques sur l'estimation. *Comptes rendus des séances de l'Académie des sciences. Série 1, Mathématique*, 296(23):1021–1024, 1983.

[BHÖ19] Leighton Pate Barnes, Yanjun Han, and Ayfer Özgür. Lower bounds for learning distributions under communication constraints via fisher information. *arXiv preprint arXiv:1902.02890*, 2019.

[HMÖW18] Yanjun Han, Pritam Mukherjee, Ayfer Özgür, and Tsachy Weissman. Distributed statistical estimation of high-dimensional and nonparametric distributions. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 506–510. IEEE, 2018.

[HÖW18] Yanjun Han, Ayfer Özgür, and Tsachy Weissman. Geometric lower bounds for distributed parameter estimation under communication constraints. In *Conference On Learning Theory*, pages 3163–3188, 2018.