Communication Complexity in Locally Private Distribution Estimation and Heavy Hitters
ICML 2019, Long Beach

June 11th, 2019

Jayadev Acharya, Cornell University
Ziteng Sun, Cornell University
Distribution Learning

- \([k] = \{0, 1, 2, \ldots, k - 1\}\), a discrete set of size \(k\).
- \(p\) : an **unknown** distribution over \([k]\).
- \(n\) users, user \(i\) has an independent \(X_i \sim p\).
- Estimator \(\hat{p} : [k]^n \rightarrow\) a distribution over \([k]\).

**Goal:** For all \(p\), with probability at least \(2/3\)

\[
\ell_1(\hat{p}, p) = \sum_{x \in [k]} |\hat{p}(x) - p(x)| \leq \alpha.
\]

\(n = \Theta \left( \frac{k}{\alpha^2} \right)\).
Frequency/ Heavy Hitter Estimation

- \([k] = \{0, 1, 2, ..., k - 1\}\) is a discrete set of size \(k\).
- \(n\) users, user \(i\) has a data point \(X_i \in [k]\).
- No distribution assumption.
- \(\forall x \in [k], N_x = \sum_i 1\{X_i = x\}\).

**Goal:** For all \(X^n\), with probability at least \(2/3\)

\[
\ell_\infty(\hat{p}, p) = \max_{x \in [k]} \left| \hat{p}(x) - \frac{N_x}{n} \right| \leq \beta.
\]
Each user sends a message $Y_i = W_i(X_i) \in \mathcal{Y}$
Resources to Consider

- **Privacy.** Data may contain sensitive information.
- **Communication.** How many bits are communicated from each user?
- **Shared Randomness.** Is shared randomness available among users?
- **Symmetry.** Are the channels symmetric?
Local Differential Privacy (LDP)

[Warner, 1965, Dwork et al., 2006, Kasiviswanathan et al., 2011, Erlingsson et al., 2014]

$\mathcal{W}$ is $\varepsilon$-LDP if for all $x, x' \in \mathcal{X}$, and $y \in \mathcal{Y}$,

$$\sup_{y \in \mathcal{Y}} \frac{\mathcal{W}(y|x)}{\mathcal{W}(y|x')} \leq e^\varepsilon.$$

We will focus on the case of high privacy. ($\varepsilon = O(1)$)
Private and Shared Randomness

**Private-coin protocols:**

$U_1, U_2, \ldots, U_n$ independent

$W_i$ is decided by $U_i$.

**Public-coin protocols:**

$U$: random bits generated at $\mathcal{R}$, available to all players.

$W_i$: determined by $U$.

0.5 round of interaction.
Symmetric, Private-coin Schemes
Theorem

[Acharya et al., 2019] Hadamard Response, which is a symmetric scheme without shared randomness, achieves the following sample complexity with only $\log k$ bits of communication from each user:

$$ \Theta\left(\frac{k^2}{\alpha^2\varepsilon^2}\right) $$

\[ n = \Theta\left(\frac{\log k}{\alpha^2 \varepsilon^2}\right) \]

Require **interaction** or **shared randomness**.
Theorem

[Acharya and Sun, 2019] To estimate each of the frequencies up to $\ell_\infty$ accuracy $\alpha$, HR uses

$$n = O\left(\frac{\log k}{\alpha^2 \varepsilon^2}\right).$$

samples.
Theorem

[Acharya and Sun, 2019] Without shared randomness, any optimal symmetric schemes for distribution learning/frequency estimation must require at least $\log k$ bits of communication.
Theorem

[Acharya and Sun, 2019] Without shared randomness, any optimal symmetric schemes for distribution learning/ frequency estimation must require at least $\log k$ bits of communication.

**Question:** What if we allow asymmetric schemes, or schemes with shared randomness?
Theorem

[Bassily and Smith, 2015] In the regime where $\varepsilon = O(1)$, for any locally private algorithm, using shared-randomness, there exists a locally private scheme with only one-bit communication which has the same privacy guarantee and the same performance, up to constant factors.
Theorem

[Bassily and Smith, 2015] In the regime where \( \varepsilon = O(1) \), for any locally private algorithm, using \textbf{shared-randomness}, there exists a locally private scheme with only one-bit communication which has the same privacy guarantee and the same performance, up to constant factors.

\textbf{Question:} Is \textbf{shared-randomness} necessary to reduce communication from users?
For distribution learning,

**Theorem**

[Acharya and Sun, 2019] There exists a private-coin scheme with only one bit communication from each user that achieve optimal performance for distribution learning.
For heavy hitter estimation,

YES!

Theorem

[Acharya and Sun, 2019] Any optimal private-coin schemes for frequency estimation must require at least \( \min\{\log k, \log n\} \) bits of communication.
### Summary of Results

**Table 3.** Sample Complexity for distribution learning under different communication budget and available randomness.

<table>
<thead>
<tr>
<th>Randomness</th>
<th>Communication</th>
<th>$O(1)$ bits</th>
<th>$O(\log k)$ bits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symmetric, Private Randomness</td>
<td>$\infty$ (Acharya &amp; Sun, 2019)</td>
<td>$\Theta\left(\frac{k^2}{\alpha^2 \epsilon^2}\right)$ (Acharya et al., 2019)</td>
<td></td>
</tr>
<tr>
<td>Private Randomness</td>
<td>$\Theta\left(\frac{k^2}{\alpha^2 \epsilon^2}\right)$ (Acharya &amp; Sun, 2019)</td>
<td>$\Theta\left(\frac{k^2}{\alpha^2 \epsilon^2}\right)$</td>
<td></td>
</tr>
<tr>
<td>Public Randomness</td>
<td>$\Theta\left(\frac{k^2}{\alpha^2 \epsilon^2}\right)$</td>
<td>$\Theta\left(\frac{k^2}{\alpha^2 \epsilon^2}\right)$</td>
<td></td>
</tr>
</tbody>
</table>

**Table 4.** Sample Complexity for frequency estimation under different communication budget and available randomness.

<table>
<thead>
<tr>
<th>Randomness</th>
<th>Communication</th>
<th>$O(1)$ bits</th>
<th>$O(\log k)$ bits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symmetric, Private Randomness</td>
<td>$\infty$</td>
<td>$\Theta\left(\frac{\log k}{\alpha^2 \epsilon^2}\right)$ (Acharya &amp; Sun, 2019)</td>
<td></td>
</tr>
<tr>
<td>Private Randomness</td>
<td>$\infty$ (Acharya &amp; Sun, 2019)</td>
<td>$\Theta\left(\frac{\log k}{\alpha^2 \epsilon^2}\right)$</td>
<td></td>
</tr>
<tr>
<td>Public Randomness</td>
<td>$\Theta\left(\frac{\log k}{\alpha^2 \epsilon^2}\right)$ (Bassily &amp; Smith, 2015)</td>
<td>$\Theta\left(\frac{\log k}{\alpha^2 \epsilon^2}\right)$</td>
<td></td>
</tr>
</tbody>
</table>
Paper available on arXiv:


06:30 – 09:00 PM, Pacific Ballroom
#177


Practical locally private heavy hitters.

Local, private, efficient protocols for succinct histograms.
In STOC, pages 127–135. ACM.

Heavy hitters and the structure of local privacy.

Calibrating noise to sensitivity in private data analysis.


