Journal of Privacy and Confidentiality
Vol. 12 (3) 2022

TPDP 2021

Submitted Nov 2021
Published Nov 2022

THE SAMPLE COMPLEXITY OF DISTRIBUTION-FREE PARITY LEARNING IN THE ROBUST SHUFFLE MODEL

KOBBI NISSIM AND CHAO YAN

Dept. of Computer Science, Georgetown University
e-mail address: kobbi.nissim@georgetown.edu

Dept. of Computer Science, Georgetown University
e-mail address: cy399@georgetown.edu

Abstract. We provide a lower bound on the sample complexity of distribution-free parity learning in the realizable case in the shuffle model of differential privacy. Namely, we show that the sample complexity of learning d-bit parity functions is $\Omega(2^{d/2})$. Our result extends a recent similar lower bound on the sample complexity of private *agnostic* learning of parity functions in the shuffle model by Cheu and Ullman (12). We also sketch a simple shuffle model protocol demonstrating that our results are tight up to poly(d) factors.

1. Introduction

The shuffle model of differential privacy (7; 11; 16) has received significant attention from researchers in the last few years. In this model, agents communicate with an untrusted analyzer via a trusted intermediary—a communication channel that shuffles all messages, hence potentially disassociating messages and their senders. Much of the recent interest in the shuffle model focuses on one-round differentially private protocols. This interest in the model is motivated, in part, by the potential to improve significantly over what is achievable in the local model of differential privacy (6; 8; 13; 21). Indeed, for functionalities such as bit addition, real addition, and histogram computation, shuffle model protocols provide accuracy comparable to that achievable with a trusted curator (1; 3; 4; 5; 11; 17; 18). See also Cheu's survey (10).

Recent works obtain lower bounds on the sample complexity of one-round robust shuffle model differentially private protocols by establishing a connection to pan-privacy (2; 12). Robust shuffle model protocols are those where differential privacy is guaranteed when a large enough fraction of agents participate honestly. In the pan-privacy model (15), individual information arrives in an online fashion to be processed by a curator. Privacy, however, is

Received by the editors October 28, 2022.
Key words and phrases: Differential privacy, private learning, parity learning.
Work K. N. was supported by NSF grant No. 1565387 TWC: Large: Collaborative: Computing Over Distributed Sensitive Data and by a gift to the McCourt School of Public Policy and Georgetown University.

www.journalprivacyconfidentiality.org
DOI:10.29012/jpc.805

required to be preserved in presence of a storage breach: as the input stream is processed by a curator, an attacker chooses a point in time in which it obtains access to observe the curator's internal state. Initiating this direction of research, Balcer, Cheu, Joseph, and Mao (2) provided reductions from pan-privacy to robust shuffle model in which a (robust) shuffle model protocol for a task is used as the main building block in the construction of a pan-private algorithm for the same or a related task. Doing so allowed them to apply lower bounds from pan-privacy to obtain lower bounds on (robust) shuffle model protocols for tasks such as histograms, uniformity testing, and counting distinct elements. A recent work of Cheu and Ullman (12) extended this proof paradigm by introducing a class of tasks that are hard for pan-privacy. This resulted in new lower bounds on the sample complexity of statistical estimation and learning tasks, including the learning of parity functions, where the latter is of specific interest because of the equivalence between the local model of differential privacy and the statistical queries model (21), as well as the impossibility of learning parity functions in the statistical queries model (22).[1]

Our results. Our main result is an exponential lower bound on the sample complexity of distribution-free parity learning in the shuffle model. Our proof has two main components. We first show how to construct a pan-private parity learner in the uniform distribution setting given a robust shuffle model distribution-free parity learner. Second, we show how to transform such a pan-private learner into a pan-private protocol for a distinguishing task requiring an exponential number of samples. We get:

Theorem 3 (informal). *For every realizable distribution-free parity learning algorithm in the shuffle model, the sample complexity is $n = \Omega(2^{d/2})$.*

This result is complemented by a robust shuffle model protocol for distribution-free parity learning with sample complexity $O(d2^{d/2})$.

Other related work. Also relevant to our work are the results of Chen, Ghazi, Kumar, and Manurangsi (9). They prove that the sample complexity of parity learning in the shuffle model is $\Omega(2^{d/(k+1)})$ for protocols with message complexity k. Comparing with our results, their lower bound depends on the message complexity of the protocol, whereas our bound holds regardless of the message complexity. On the other hand, our lower bound holds only for robust shuffle model protocols, whereas the result of Chen et al. does not require robustness.

2. Preliminaries

2.1. **Differential privacy, pan-privacy, and the shuffle model.**

Let X be a data domain. We say that two datasets $x, x' \in X^n$ are *neighboring* if they differ on exactly one entry, i.e., $|\{i : x_i \neq x'_i\}| = 1$.

Definition 1 (differential privacy (14)). *A randomized mechanism $M : X^n \to Y$ preserves (ε, δ)-differential privacy if for all neighboring $x, x' \in X^n$, and for all events $T \subseteq Y$,*

$$\Pr[M(x) \in T] \leq e^{\varepsilon} \cdot \Pr[M(x') \in T] + \delta,$$

[1] Considering the realizable setting with underlying uniform distribution on samples, the equivalence implies that no local model protocol exists for parity learning with polynomial round complexity and polynomial sample complexity.

where the probability is over the randomness of the mechanism M.

Definition 2 (pan-privacy (15))**.** *For an online mechanism $M : X^n \to Y$, let $S_{\leq t}(x)$ represent the internal state of $M(x)$ after receiving the t first inputs $x_1, \dots, x_t$. The pan-private protocol starts with an initial state $S_{\leq 0}(x)$. At each step $t+1$, the internal state $S_{\leq t+1}(x)$ is updated by aggregating $S_{\leq t}(x)$ and x_{t+1}. We say M is (ε, δ)-pan-private if for every two neighbouring datasets $x, x' \in X^n$, for every event $T \subseteq Y$, and for every $1 \leq t \leq n$,*

$$\Pr[(S_{\leq t}(x), M(x)) \in T] \leq e^{\varepsilon} \cdot \Pr[(S_{\leq t}(x'), M(x')) \in T] + \delta,$$

where the probability is over the randomness of the online mechanism M.

A one-round shuffle model mechanism $M : X^n \to Y$, as introduced in (11), consists of three types of algorithms: (i) local randomizers $(R_1, \dots, R_n)$ where each randomizer R_i maps an input $x_i \in X$ to a collection of messages from an arbitrary message domain; (ii) A shuffle S receives a collection of messages and outputs it in a random order; and (iii) an analyzer algorithm A maps a collection of messages to an outcome in Y. The *robust* shuffle model considers malicious users who may avoid sending their messages to the shuffle (2). We denote the local randomizers of such users by $\perp$. The output of $M = ((R_1, \dots, R_n), S, A)$ on input $x = (x_1, \dots, x_n)$ is

$$A(S(\hat{R}_1(x_1), \dots, \hat{R}_n(x_n))),$$

where $\hat{R}_i = R_i$ for honest users and $\hat{R}_i = \perp$ for malicious users.

Definition 3 (robust one-round shuffle model (2))**.** *A one-round shuffle model mechanism $M = ((R_1, \dots, R_n), S, A)$ is γ-robust and (ε, δ)-differentially private if when at least γn of the parties are honest for all neighboring $x, x' \in X^n$ and for all events $T \subseteq Y$,*

$$\Pr[S(\hat{R}_1(x_1), \dots, \hat{R}_n(x_n)) \in T] \leq e^{\varepsilon} \cdot \Pr[S(\hat{R}_1(x'_1), \dots, \hat{R}_n(x'_n)) \in T] + \delta,$$

where the probability is over the randomness of $(\hat{R}_1, \dots, \hat{R}_n)$ and the shuffle S.

2.2. **Private learning.** A concept class C is a collection of predicates over the data domain $c : X \to \{\pm 1\}$. Let $P \in \Delta(X)$ be a probability distribution over the data domain X and let $h : X \to \{\pm 1\}$. The generalization error of hypothesis h with respect to the concept c is $\mathrm{error}_P(c, h) = \Pr_{x \sim P}[h(x) \neq c(x)]$.

Definition 4 (PAC learning (24))**.** *A concept class C is (α, β, m)-PAC learnable if there exists an algorithm L such that for all distributions $P \in \Delta(X)$ and all concepts $c \in C$,*

$$\Pr\left[\{x_i\}_{i=1}^m \sim P; h \leftarrow L\Big(\{(x_i, c(x_i)\}_{i=1}^m\Big); error_P(c, h) \leq \alpha\right] \geq 1 - \beta,$$

where the probability is over the choice of $x_1, \dots, x_m$ i.i.d. from P and the randomness of L.

For an arbitrary distribution over labeled pairs $P \in \Delta(X \times \{0, 1\})$ the classification error a hypothesis h obtains is $err_P(h) = \Pr_{(x,y) \sim P}[h(x) \neq y]$.

Definition 5 (agnostic PAC Learning (20))**.** *A hypothesis class $\mathcal{H}$ is (α, β, m)-agnostic PAC learnable if there exists an algorithm L, such that for any distribution P over $(\mathcal{X} \times \{0, 1\})$,*

$$\Pr\left[\{(x_i, y_i)\}_{i=1}^m \sim P; h \leftarrow L\Big(\{(x_i, y_i\}_{i=1}^m\Big); err_P(h) \leq \min_{h \in \mathcal{H}} (err_P(h)) + \alpha\right] \geq 1 - \beta,$$

where the probability is over the choice of $(x_1, y_1), \dots, (x_m, y_m)$ i.i.d. from P and the randomness of L.

Note that Definition 4 is of an *improper* learner as the hypothesis h need not come from the concept class C.

Definition 6 (weight k parity)**.** *Let $PARITY_{d,k} = \{c_{r,b}\}_{r\subseteq[d],|r|\leq k,b\in\{\pm1\}}$ where $c_{r,b} : \{\pm1\}^d \to \{\pm1\}$ is defined as $c_{r,b}(x) = b \cdot \prod_{i\in r} x_i$. When $k = d$, we omit k and write $PARITY_d$.*

Definition 7. *A* distribution-free *parity learner is a PAC learning algorithm for $PARITY_{d,k}$. A* uniform distribution *parity learner is a PAC learning algorithm for $PARITY_{d,k}$ where the underlying distribution P is known to be uniform over $X = \{\pm1\}^d$.*

Definition 8 (private learning (21))**.** *A concept class C is private PAC learnable by algorithm L with parameters $\alpha, \beta, m, \varepsilon, \delta$, if L is (ε, δ)-differentially private and L is (α, β, m)-PAC learns concept class C.*

2.3. **Hard tasks for pan-private mechanisms.** Cheu and Ullman (12) provide a family of distributions $\{P_v\}$ for which the sample complexity of any pan-private mechanism distinguishing a randomly chosen distribution in $\{P_v\}$ from uniform is high. Let $X = \{\pm1\}^d$ be the data domain. Let U be the uniform distribution over X. For $0 < \alpha \leq 1/2$, a non-empty set $\ell \subseteq [d]$, and a bit $b \in \{\pm1\}$, define the distribution $P_{d,\ell,b,\alpha} \in \Delta(X)$ to be

$$P_{d,\ell,b,\alpha}(x) = \begin{cases} (1+2\alpha)2^{-d} & \text{if } \prod_{i\in\ell} x_i = b \\ (1-2\alpha)2^{-d} & \text{if } \prod_{i\in\ell} x_i = -b \end{cases}$$

Equivalently,

$$P_{d,\ell,b,\alpha}(x) = (1 + 2b\alpha \prod_{i\in\ell} x_i) \cdot 2^{-d}.$$

Note that for $\alpha = 1/2$ the support of $P_{d,\ell,b,\alpha}$ is exactly the set of strings $x \in \{\pm1\}^d$ satisfying $\prod_{i\in\ell} x_i = b$. Define the family of distributions

$$\mathcal{P}_{d,k,\alpha} = \{P_{d,\ell,b,\alpha}(x) : \ell \subseteq [d], 0 < |\ell| \leq k, b \in \{\pm1\}\}.$$

Let $P_{d,L,B,\alpha}$ denote the distribution which is chosen uniformly at random from the family of distributions $\mathcal{P}_{d,k,\alpha}$, i.e., L is a uniformly random non-empty subset of $[d]$ with cardinality at most k and $B \in_R \{\pm1\}$.

Theorem 1 ((12), restated)**.** *Let M be an (ε, δ)-pan-private algorithm. If $d_{TV}(M(P^n_{d,L,B,\alpha}), M(U^n)) \geq T$, then*

$$n = \Omega\left(T \middle/ \sqrt{\frac{\varepsilon^2\alpha^2}{\binom{d}{\leq k}} + \delta \log \frac{\binom{d}{\leq k}}{\delta}}\right).$$

In particular, when $\delta \log\left(\binom{d}{\leq k}/\delta\right) = o\left(\varepsilon^2\alpha^2/\binom{d}{\leq k}\right)$ we get that

$$n = \Omega\left(\frac{T \cdot \sqrt{\binom{d}{\leq k}}}{\varepsilon\alpha}\right).$$

Remark. Cheu and Ullman (12) argue agnostic parity learner in the condition of $0 < \alpha < 1/2$, where all the parity functions have a positive error. In this work, by setting $\alpha = 1/2$, the hypothesis class can be equivalent to the concept class, i.e. there exists a parity function with error of 0. Then the parity learner can be a realistic learner.

2.4. **Tail inequalities.**

Theorem 2 (Chebyshev's inequality)**.** *Let X be a random variable with expected value μ and non-zero variance σ^2. Then for any positive number a,*

$$\Pr(|X-\mu| \geq a) \leq \frac{\sigma^2}{a^2}.$$

2.5. **Divisibility of discrete Laplace distribution.**

The discrete Laplace distribution can be divided into n differences of two Pólya distributions (3; 23; 19). If X_i and Y_i are independent random variables that follow Pólya$(1/n, \alpha)$, then the random variable $Z = \sum_{i=1}^{n} X_i - Y_i$ follows the discrete Laplace distribution, where $\Pr[Z = k] \propto \alpha^{|k|}$.

3. A lower bound on the sample complexity of parity learning in the shuffle model

3.1. **From robust shuffle model parity learner to a pan-private parity learner.**

Given a robust shuffle model distribution-free parity learner, we show how to construct a uniform distribution pan-private parity learner. Our reduction–Algorithm LearnParUnif–is described in Algorithm 1. We use a similar technique to the padding presented in (2; 12), with small modifications. To allow the shuffle model protocol use differing randomzers $R_1, \ldots, R_n$, the pan-private learner applies these randomizers in a random order (the random permutation π). The padding is done with samples of the form $(1^d, \hat{b})$, where $\hat{b}$ is selected uniformly at random from $\{\pm 1\}$. Finally, as in (12), the number of labeled samples which the pan-private algorithm considers from its input is binomially distributed, so that if (x_i, y_i) are such that x_i is uniform in X and $y_i = c_{r,b}(x_i) = b \cdot \prod_{i \in r} x_i$ then (after a random shuffle) the input distribution presented to the shuffle model protocol is statistically close to a mixture of the two following distributions: (i) a distribution where $\Pr[(x_i, y_i) = (1^d, \hat{b})] = 1$ and (ii) a distribution where x_i is uniformly selected in $\{\pm 1\}^d$ and $y_i = c_{r,b}(x_i)$.

Proposition 1. *Algorithm LearnParUnif is (ε, δ)-pan-private.*

Proof sketch, following (2; 12). Let x and x' be two neighboring data sets, and let j be the index where x and x' differ. Let $1 \leq t \leq n/3$ be the time an adversary probes into the algorithm's memory.

If $t \geq j$, then $S_{\leq t} = (S \circ (R^{\pi(1)}, \ldots, R^{\pi(n/3+t)}))((1^d, b)^{n/3}, w_1, \ldots, w_t)$ and, as M is a robust differentially private mechanism $S_{\leq t}$ preserves (ε, δ)-differential privacy. Because $A(s_{final})$ is post-processing of $S_{\leq t}$ the outcome of LearnParUnif is (ε, δ)-pan-private.

If $t < j$, then $S_{\leq t}(x)$ is identically distributed to $S_{\leq t}(x')$. Note that as M is a robust differentially private mechanism, we get that

$$\sigma = (S \circ (R^{\pi(n/3+t+1)}, \ldots, R^{\pi(n)}))(w_{t+1}, \ldots, w_{N'}, (1^d, b), \ldots, (1^d, b))$$

preserves (ε, δ)-differential privacy. To conclude the proof, note that $(S_{\leq t}(x), A(s_{final}))$ is the result of post-processing σ. □

Algorithm 1: LearnParUnif, a uniform distribution pan-private parity learner

Let $M = ((R_1, \dots, R_n), S, A)$ be a 1/3-robust differentially private distribution-free parity learner.

Input: $n/3$ labeled examples (x_i, y_i) where $x_i \in X$ and $y_i \in \{\pm 1\}$.

Randomly choose a permutation $\pi : [n] \to [n]$.
Randomly choose $\hat{b} \in_R \{\pm 1\}$.
Create initial state $s_0 \leftarrow S(R_{\pi(1)}(1^d, \hat{b}), \dots, R_{\pi(n/3)}(1^d, \hat{b}))$.
Sample $N' \sim \mathbf{Bin}(n, 2/9)$.
Set $N' \leftarrow min(N', n/3)$.
for $i \in [n/3]$ **do**
    if $i \in [N']$ **then**
        $w_i \leftarrow (x_i, y_i)$
    else
        $w_i \leftarrow (1^d, \hat{b})$
    end
    $s_i \leftarrow S(s_{i-1}, R_{\pi(n/3+i)}(w_i))$
end
$s_{final} \leftarrow S(s_{n/3}, R_{\pi(2n/3+1)}(1^d, \hat{b}), \dots, R_{\pi(n)}(1^d, \hat{b}))$
return $A(s_{final})$

Proposition 2 (learning)**.** *Let M be a (α, β, m)-distribution-free parity learner, where $\alpha, \beta < 1/4$ and $m = n/9$. Algorithm LearnParUnif is a uniform distribution parity learner that with probability at least $1/4$ correctly identifies the concept $c_{r,b}$.*

Proof sketch. Algorithm LearnParUnif correctly guesses the label b for 1^d with probability $1/2$. Assuming $\hat{b} = b$ the application of M uniquely identifies r, b with probability at least $1/2$. Thus, LearnParUnif recovers $c_{r,b}$ with probability at least $1/4$. □

3.2. **From pan-private parity learner to distinguishing hard distributions.** In this section, we use Theorem 1 to obtain a lower bound on the sample complexity of parity learning in the shuffle model. In Algorithm 2, we provide a reduction from identifying the hard distribution $P_{d,\ell,b,1/2}$ presented in section 2.3 to pan-private parity learning. Recall that ℓ is a set of indexes, such that $\Pi_{i \in \ell} x_i = b$ for any example $(x_1, \dots, x_d)$ from distribution $P_{d,\ell,b,1/2}$.

Observation 1. *The pan-privacy of Algorithm 2 follows from the pan-privacy of algorithm Π.*

Proposition 3. *Given a uniform distribution parity learner that with probability at least $1/4$ correctly identifies the concept $c_{r,b}$, algorithm 2 can correctly identify the distribution $P_{d,\ell,b,1/2}$ with probability at least $|\ell|/4d$.*

Proof. Note that with probability $|\ell|/d$ we get that $i^* \in \ell$, in which case the inputs $x_1, \dots, x_n$ provided to the learner Π in Step 7 are uniformly distributed in $\{\pm 1\}^{d-1}$ and $y_j = b \cdot \prod_{i \in \ell \setminus \{i^*\}} x_j[i]$, i.e., the inputs to Π are consistent with the concept $c_{\ell \setminus \{i^*\}, b}$. □

Algorithm 2: IdentifyHard, a pan-private algorithm for identifying the distribution $P_{d,\ell,b,1/2}$

Let Π be a pan-private uniform distribution parity learner.

Input: A sample of n examples $z = (z_1, z_2, \ldots, z_n)$, where each example is of the form $z_j = (z_j[1], z_j[2], \ldots, z_j[d]) \in \{\pm 1\}^d$

Randomly choose $i^* \in_R [d]$.

/* Apply the uniform distribution parity learner Π: */

for $j \in [n]$ **do**

  $y_j \leftarrow z_j[i^*]$

  $x_j = z_j$

  $x_j[i^*] = \perp$ /* i.e., x_j equals z_j with entry i^* erased */

  Provide (x_j, y_j) to Π.

end

$(r, b) \leftarrow \Pi((x_1, y_1), \ldots, (x_n, y_n))$

$\ell \leftarrow r \cup \{i^*\}$

return (ℓ, b)

On the uniform distribution, the generalization error of any parity function is 1/2. On $P_{d,\ell,b,1/2}$ Algorithm 2 succeeds with probability $|\ell|/4d$ to identify ℓ, b. Algorithm 3 evaluates the generalization error of the concept learned in Algorithm 2 towards exhibiting a large total variance distance on $P^n_{d,L,B,1/2}$ and U^n.

Algorithm 3: DistPU: Distinguisher for $P^{n+m}_{d,L,B,1/2}$ and U^{n+m}

Let $M = ((R_1, \ldots, R_n), S, A)$ be the pan-private algorithm described in Algorithm 2.

Input: A sample of $m + n$ examples $z = (z_1, z_2, \ldots, z_{n+m})$, where $m = \max\{512d/k, 64\sqrt{2d/k}/\varepsilon\}$ and each example is of the form $z_j = (z_j[1], z_j[2], \ldots, z_j[d]) \in \{\pm 1\}^d$.

Let (ℓ, b) be the outcome of executing M on the first n examples $z_1, \ldots, z_n$.

$c \leftarrow \mathbf{Lap}(1/\varepsilon)$ /*Adding the laplace noise to make the internal state differentially private*/

for $i \in [m]$ **do**

  if $\prod_{j\in\ell} z_{i+n}[j] = b$ **then** $c \leftarrow c + 1$

end

$c^* \leftarrow c + \mathbf{Lap}(1/\varepsilon)$

if $c^* \geq 3m/4$ **then** **return** *1* **else** **return** *0*

Observe that if $z \sim P^{n+m}_{d,L,B,1/2}$ then in every execution of Algorithm 3 there exist $\ell \subset [d]$ of cardinality at most k and $b \in \{\pm 1\}$ such that $z \sim P^{n+m}_{d,\ell,b,1/2}$. In Proposition 4, we compute that if $z \sim P^{n+m}_{d,\ell,b,1/2}$, the probability of $\mathsf{DistPU}(z) = 1$ is at least $|\ell|/8d$. In Proposition 5, we use the result of Proposition 4 to compute that if $z \sim P^{n+m}_{d,L,B,1/2}$, the probability of $\mathsf{DistPU}(z) = 1$ is at least $|\ell|/32d$. In Proposition 6, we evaluate the upper bound of $\Pr_{z\sim U^{n+m}}[\mathsf{DistPU}(z) = 1]$. Then we show that the total variance distance of $\mathsf{DistPU}(U^{n+m})$ and $\mathsf{DistPU}(P^{n+m}_{d,L,B,1/2})$ is at least a constant.

Proposition 4. $\Pr_{z\sim P^{n+m}_{d,\ell,b,1/2}}[\mathit{DistPU}(z)=1]\geq|\ell|/8d$.

Proof. For any $z\sim P^{n+m}_{d,\ell,b,1/2}$, we have $\prod_{i\in\ell} z_i=b$, so

$$\begin{aligned}\Pr_{z\sim P^{n+m}_{d,\ell,b,1/2}}[\mathsf{DistPU}(z)=1] &\geq \Pr[\mathsf{DistPU}\text{ correctly identifies }(\ell,b)]\cdot\Pr[c^*\geq 3m/4]\\ &\geq \frac{|\ell|}{4d}\cdot\Pr[\mathbf{Lap}(1/\varepsilon)+\mathbf{Lap}(1/\varepsilon)\geq -m/4]\\ &\geq \frac{|\ell|}{4d}\cdot\frac{1}{2} \qquad \text{(by symmetry of }\mathbf{Lap}\text{ around 0)}\\ &= \frac{|\ell|}{8d}.\end{aligned}$$

□

Proposition 5. $\Pr_{z\sim P^{n+m}_{d,L,B,1/2}}[\mathit{DistPU}(z)=1])\geq k/32d$.

Proof.

$$\begin{aligned}\Pr_{z\sim P^{n+m}_{d,L,B,1/2}}[\mathsf{DistPU}(z)=1] &= \sum_{\ell\in[d],|\ell|\leq k,b\in\{\pm1\}}\Pr_{z\sim P^{n+m}_{d,\ell,b,1/2}}[\mathsf{DistPU}(z)=1]\cdot\Pr[(L,B)=(\ell,b)]\\ &\geq \sum_{\ell\in[d],k/2\leq|\ell|\leq k,b\in\{\pm1\}}\Pr_{z\sim P^{n+m}_{d,\ell,b,1/2}}[\mathsf{DistPU}(z)=1]\cdot\Pr[(L,B)=(\ell,b)]\\ &\geq \sum_{\ell\in[d],k/2\leq|\ell|\leq k,b\in\{\pm1\}}\frac{k}{16d}\cdot\Pr[(L,B)=(\ell,b)]\\ &= \frac{k}{16d}\cdot\Pr[|L|\geq k/2]\\ &= \frac{k}{16d}\cdot\frac{\binom{d}{\leq k}-\binom{d}{\leq k/2}}{\binom{d}{\leq k}}\geq\frac{k}{32d}.\end{aligned}$$

The last inequality follows from $\frac{\binom{d}{\leq k}-\binom{d}{\leq k/2}}{\binom{d}{\leq k}}\geq 1/2$. [2] □

Proposition 6. $\Pr_{z\sim U^{n+m}}[\mathit{DistPU}(z)=1]\leq k/64d$.

Proof. For all (ℓ,b), we have that $\Pr_{z\sim U}[\prod_{j\in\ell} z[j]=b]=1/2$, so we have

$$\begin{aligned}\Pr_{z\sim U^{n+m}}[\mathsf{DistPU}(z)=1] &= \Pr[\mathbf{Bin}(m,1/2)+\mathbf{Lap}(1/\varepsilon)+\mathbf{Lap}(1/\varepsilon)\geq 3m/4]\\ &\leq \Pr[|\mathbf{Bin}(m,1/2)+\mathbf{Lap}(1/\varepsilon)+\mathbf{Lap}(1/\varepsilon)-m/2|\geq m/4]\\ &\leq \frac{m/4+2/\varepsilon^2+2/\varepsilon^2}{m^2/16} \qquad \text{(Theorem 2)}\\ &= 4/m+64/\varepsilon^2m^2\\ &\leq \frac{k}{128d}+\frac{k}{128d}=\frac{k}{64d}.\end{aligned}$$

[2]If $k=d$ then $\binom{d}{\leq k}\geq 2\binom{d}{\leq k/2}$. Otherwise ($k<d$) we get for $0\leq i\leq\lfloor k/2\rfloor$ that the difference between $\lfloor k/2\rfloor+1+i$ and $d/2$ is smaller than the difference between $\lfloor k/2\rfloor-i$ and $d/2$ hence $\binom{d}{\lfloor k/2\rfloor-i}<\binom{d}{\lfloor k/2\rfloor+1+i}$, thus $\binom{d}{\leq k}=\sum_{0\leq i\leq\lfloor k/2\rfloor}\binom{d}{i}+\sum_{\lfloor k/2\rfloor+1\leq i\leq k}\binom{d}{i}>2\sum_{0\leq i\leq\lfloor k/2\rfloor}\binom{d}{i}=2\binom{d}{\leq k/2}$.

□

Combining Propositions 6 and 5 we can now get a lower bound on the statistical distance between $\mathsf{DistPU}(U^{n+m})$ and $\mathsf{DistPU}(P^{n+m}_{d,L,B,1/2})$:

$$\begin{aligned} d_{TV}&(\mathsf{DistPU}(U^{n+m}), \mathsf{DistPU}(P^{n+m}_{d,L,B,1/2})) \\ &\geq \Pr_{z\sim P^{n+m}_{d,L,B,1/2}}[\mathsf{DistPU}(z)=1] - \Pr_{z\sim U^{n+m}}[\mathsf{DistPU}(z)=1] \\ &\geq \frac{k}{32d} - \frac{k}{64d} = \frac{k}{64d}. \end{aligned}$$

In particular, for all k we get that $d_{TV}(\mathsf{DistPU}(U^{n+m}), \mathsf{DistPU}(P^{n+m}_{d,L,B,1/2})) \geq k/64d$ and for $k = d$ we get $d_{TV}(\mathsf{DistPU}(U^{n+m}), \mathsf{DistPU}(P^{n+m}_{d,L,B,1/2})) \geq 1/64$. We can now conclude our main result:

Theorem 3. *For any $(\varepsilon, \delta, 1/3)$-robust private distribution-free parity learning algorithm in the shuffle model, where $\varepsilon = O(1)$, the sample complexity is*

$$n = \Omega\left(\frac{2^{d/2}}{\varepsilon}\right).$$

Proof. Let $k = d$, applying Theorem 1, DistPU has sample complexity

$$n + m = \Omega\left(\frac{2^{d/2}}{\varepsilon}\right).$$

Since $k \geq 1$, $\varepsilon = O(1)$, then $m = O(d/\varepsilon)$. By the reduction of Algorithm DistPU from a $(\varepsilon, \delta, 1/3)$-robust private parity learning algorithm, any $(\varepsilon, \delta, 1/3)$-robust private parity learning algorithm has sample complexity

$$n = \Omega\left(\frac{2^{d/2}}{\varepsilon}\right).$$

□

3.3. **Tightness of the lower bound.** We now observe that Theorem 3 is tight as there exists a 1/3-robust agnostic parity learner in the shuffle model with an almost matching sample complexity. For every possible hypothesis (ℓ, b) (there are 2^{d+1} hypotheses), the learner estimates the number of samples that are consistent with the hypothesis, $con_{\ell,b} = |\{i : b \cdot \prod_{j\in\ell} x_i[j] = y_i\}|$. Let N be the number of labeled examples.

One possibility for counting the number of consistent samples is to use the protocol by Balle et al. (3), which is an (ε, δ)-differentially private one-round shuffle model protocol for estimating $\sum a_i$ where $a_i \in [0, 1]$. The outcome of this protocol is statistically close to $\sum a_i + \mathbf{DLap}(1/\varepsilon)$ and the statistical distance δ can be made arbitrarily small by increasing the number of messages sent by each agent. (We use the notation $\mathbf{DLap}(1/\varepsilon)$ for the discrete Laplace distribution, where the probability of selecting $i \in \mathbb{Z}$ is proportional to $e^{-\varepsilon|i|}$.) The protocol uses the divisibility of discrete Laplace distribution, generating discrete Laplace noise ν as the sum of differences of Pólya random variables:

$$\nu = \sum_{i=1}^{n} \left(\mathbf{Pólya}(1/n, e^{-\varepsilon}) - \mathbf{Pólya}(1/n, e^{-\varepsilon})\right).$$

To make the protocol γ-robust, we slightly change the noise generation to guarantee (ε, δ) differential privacy in the case where only $n/3$ parties participate in the protocol. This can be done by changing the first parameter of the Pólya random variables to $3/n$, resulting in

$$\nu = \sum_{i=1}^{n} \left(\mathbf{Pólya}(3/n, e^{-\varepsilon}) - \mathbf{Pólya}(3/n, e^{-\varepsilon})\right).$$

Observe that ν is distributed as the sum of three independent $\mathbf{DLap}(1/\varepsilon)$ random variables. Using this protocol, it is possible for the analyzer to compute a noisy estimate of the number of samples consistent with each hypothesis, $\widetilde{con}_{\ell,b} = con_{\ell,b} + \nu$, and then output $(\ell, b) = \mathrm{argmax}_{\ell,b}(\widetilde{con}_{\ell,b})$. The sample complexity of this learner is $O_{\alpha,\beta,\varepsilon,\delta}(d2^{d/2})$.

Algorithm 4: LearnParity: an agnostic parity learning algorithm

Let $\varepsilon' = \frac{\varepsilon}{4\sqrt{2^d \ln(1/\delta^*)}}$. Let ShuffleCount be an $(\varepsilon', \delta', \gamma)$-robust shuffle protocol that compute the sum of $\{0, 1\}$ bits.

Input: $N \geq \max\left\{\frac{36((d+2)\ln 2 - \ln\beta)}{\alpha^2}, \frac{48(\ln 3 + (d+2)\ln 2)\sqrt{2^d \ln 1/\delta^*}}{\alpha\varepsilon}\right\}$ labeled examples (x_i, y_i), where $x_i \in \{\pm 1\}^d$ and $y_i \in \{\pm 1\}$.

```
for ℓ ⊆ [d], b ∈ {±1} do
    Apply ShuffleCount to obtain a noisy count con~_{ℓ,b} of samples for which
      b · ∏_{j∈ℓ} x_i[j] = y_i.
end
(ℓ̂, b̂) ← argmax_{ℓ,b}({con~_{ℓ,b}}_{ℓ⊆[d], b∈{±1}})
return (ℓ̂, b̂)
```

Proposition 7 (privacy)**.** *For $\varepsilon < 1$, LearnParity is $(\varepsilon, \delta, \gamma)$-robust private, where $\delta = k \cdot \delta' + \delta^*$.*

Proof. LearnParity performes k counting computations applying ShuffleCount and then selects the largest one. By the corollary of advanced composition, setting $\varepsilon' = \varepsilon/2\sqrt{2k \ln 1/\delta^*}$ can make LearnParity (ε, δ)-differentially private. Since ShuffleCount is γ-robust, LearnParity is γ-robust. □

To prove that LearnParity is an (α, β)-agnostic parity learner, we show that (i) the true number of samples that agree with the parity function is close to the expected number of samples that agree with the parity function (Proposition 8); (ii) the noisy estimate produced by ShuffleCount is close to the true number of samples that agree with the parity function (Proposition 9).

Let $p_{\ell,b}$ represent the probability that one example agrees with the parity function $Par_{\ell,b}$.

Proposition 8.

$$\Pr\left[|p_{\ell,b} \cdot N - con_{\ell,b}| \le \frac{\alpha N}{4}\right] \ge 1 - e^{-\frac{\alpha^2 \cdot N}{36}}$$

Proof. $con_{\ell,b}$ agrees with the distribution $\mathbf{Bin}(N, p_{\ell,b})$, by Chernoff bound,

$$\Pr[con_{\ell,b} > (p_{\ell,b} + \alpha/4) \cdot N] = \Pr[con_{\ell,b} > (1 + \alpha/4p_{\ell,b}) \cdot p_{\ell,b}N] \le e^{-\frac{\alpha^2 \cdot N}{32p_{\ell,b}+4\alpha}} \le e^{-\frac{\alpha^2 \cdot N}{36}}$$

$$\Pr[con_{\ell,b} < (p_{\ell,b} - \alpha/4) \cdot N] = \Pr[con_{\ell,b} < (1 - \alpha/4p_{\ell,b}) \cdot p_{\ell,b}N] \le e^{-\frac{\alpha^2 \cdot N}{32p_{\ell,b}}} \le e^{-\frac{\alpha^2 \cdot N}{36}}$$

□

Proposition 9.

$$\Pr\left[|\widetilde{con}_{\ell,b} - con_{\ell,b}| \le \frac{\alpha N}{4}\right] \ge 1 - 3 \cdot e^{-\frac{\alpha N \varepsilon'}{12}},$$

Proof. The noise added in ShuffleCount amounts to the sum of three $\mathbf{DLap}(e^{\varepsilon})$ variables . The probability that a $\mathbf{DLap}(e^{\varepsilon})$ variable exceeds $\alpha N/12$ is

$$\begin{aligned}
\Pr[|\mathbf{DLap}(e^{\varepsilon})| > \alpha N/12] &= 2 \cdot \frac{e^{\varepsilon'} - 1}{e^{\varepsilon'} + 1} \cdot ((e^{\varepsilon'})^{-\frac{\alpha N}{12} - 1} + (e^{\varepsilon'})^{-\frac{\alpha N}{12} - 2} + \dots) \\
&= 2 \cdot \frac{e^{\varepsilon'} - 1}{e^{\varepsilon'} + 1} \cdot \frac{e^{-\varepsilon' \cdot (\frac{\alpha N}{12} + 1)}}{1 - e^{-\varepsilon'}} \\
&= \frac{2 \cdot e^{-\frac{\alpha N \varepsilon'}{12}}}{e^{\varepsilon'} + 1} \\
&< e^{-\frac{\alpha N \varepsilon'}{12}}.
\end{aligned}$$

Hence, by union bound, the probaility the sum of three $\mathbf{DLap}(e^{\varepsilon})$ variables exceeds $\alpha N/4$ is at most $3 \cdot e^{-\frac{\alpha N \varepsilon'}{12}}$.

□

Let OPT be the lowest possible error of the hypothesis taken from all parity functions. If $\widetilde{con}_{\ell,b} - Np_{\ell,b} < \alpha N/2$ for all (ℓ, b), the error of hypothesis outputted by the algorithm is less than $OPT + \alpha$.

Proposition 10. *LearnParity is (α, β)-agnostic learning.*

Proof. By union bound,

$$\beta \le k \cdot e^{-\frac{\alpha^2 n}{36}} + k \cdot 3 \cdot e^{-\frac{\alpha n \varepsilon'}{12}} \le \beta/2 + \beta/2 = \beta.$$

□

References

[1] Victor Balcer and Albert Cheu. Separating Local and Shuffled Differential Privacy via Histograms. In Yael Tauman Kalai, Adam D. Smith, and Daniel Wichs, editors, *1st Conference on Information-Theoretic Cryptography (ITC 2020)*, volume 163 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 1:1–1:14, Dagstuhl, Germany, 2020. Schloss Dagstuhl–Leibniz-Zentrum für Informatik. URL: https://drops.dagstuhl.de/opus/volltexte/2020/12106, https://doi.org/10.4230/LIPIcs.ITC.2020.1.

[2] Victor Balcer, Albert Cheu, Matthew Joseph, and Jieming Mao. Connecting robust shuffle privacy and pan-privacy. In Dániel Marx, editor, *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, SODA 2021, Virtual Conference, January 10 - 13, 2021*, pages 2384–2403. SIAM, 2021. https://doi.org/10.1137/1.9781611976465.142.

[3] Borja Balle, James Bell, Adrià Gascón, and Kobbi Nissim. Differentially private summation with multi-message shuffling. *CoRR*, abs/1906.09116, 2019. URL: http://arxiv.org/abs/1906.09116, arXiv:1906.09116.

[4] Borja Balle, James Bell, Adrià Gascón, and Kobbi Nissim. The privacy blanket of the shuffle model. In Alexandra Boldyreva and Daniele Micciancio, editors, *Advances in Cryptology - CRYPTO 2019 - 39th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2019, Proceedings, Part II*, volume 11693 of *Lecture Notes in Computer Science*, pages 638–667. Springer, 2019. https://doi.org/10.1007/978-3-030-26951-7_22.

[5] Borja Balle, James Bell, Adrià Gascón, and Kobbi Nissim. Private summation in the multi-message shuffle model. In Jay Ligatti, Xinming Ou, Jonathan Katz, and Giovanni Vigna, editors, *CCS '20: 2020 ACM SIGSAC Conference on Computer and Communications Security, Virtual Event, USA, November 9-13, 2020*, pages 657–676. ACM, 2020. https://doi.org/10.1145/3372297.3417242.

[6] Amos Beimel, Kobbi Nissim, and Eran Omri. Distributed private data analysis: Simultaneously solving how and what. In David A. Wagner, editor, *Advances in Cryptology - CRYPTO 2008, 28th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 17-21, 2008. Proceedings*, volume 5157 of *Lecture Notes in Computer Science*, pages 451–468. Springer, 2008. https://doi.org/10.1007/978-3-540-85174-5_25.

[7] Andrea Bittau, Úlfar Erlingsson, Petros Maniatis, Ilya Mironov, Ananth Raghunathan, David Lie, Mitch Rudominer, Ushasree Kode, Julien Tinnés, and Bernhard Seefeld. Prochlo: Strong privacy for analytics in the crowd. In *Proceedings of the 26th Symposium on Operating Systems Principles, Shanghai, China, October 28-31, 2017*, pages 441–459. ACM, 2017. https://doi.org/10.1145/3132747.3132769.

[8] T.-H. Hubert Chan, Elaine Shi, and Dawn Song. Optimal lower bound for differentially private multi-party aggregation. In Leah Epstein and Paolo Ferragina, editors, *Algorithms - ESA 2012 - 20th Annual European Symposium, Ljubljana, Slovenia, September 10-12, 2012. Proceedings*, volume 7501 of *Lecture Notes in Computer Science*, pages 277–288. Springer, 2012. https://doi.org/10.1007/978-3-642-33090-2_25.

[9] Lijie Chen, Badih Ghazi, Ravi Kumar, and Pasin Manurangsi. On distributed differential privacy and counting distinct elements. In James R. Lee, editor, *12th Innovations in Theoretical Computer Science Conference, ITCS 2021, January 6-8, 2021, Virtual Conference*, volume 185 of *LIPIcs*, pages 56:1–56:18. Schloss Dagstuhl - Leibniz-Zentrum

für Informatik, 2021. https://doi.org/10.4230/LIPIcs.ITCS.2021.56.

[10] Albert Cheu. Differential privacy in the shuffle model: A survey of separations. *CoRR*, abs/2107.11839, 2021. URL: https://arxiv.org/abs/2107.11839, arXiv:2107.11839.

[11] Albert Cheu, Adam D. Smith, Jonathan R. Ullman, David Zeber, and Maxim Zhilyaev. Distributed differential privacy via shuffling. In Yuval Ishai and Vincent Rijmen, editors, *Advances in Cryptology - EUROCRYPT 2019 - 38th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Darmstadt, Germany, May 19-23, 2019, Proceedings, Part I*, volume 11476 of *Lecture Notes in Computer Science*, pages 375–403. Springer, 2019. https://doi.org/10.1007/978-3-030-17653-2_13.

[12] Albert Cheu and Jonathan R. Ullman. The limits of pan privacy and shuffle privacy for learning and estimation. In Samir Khuller and Virginia Vassilevska Williams, editors, *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 1081–1094. ACM, 2021. https://doi.org/10.1145/3406325.3450995.

[13] John C. Duchi, Michael I. Jordan, and Martin J. Wainwright. Local privacy and statistical minimax rates. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA*, pages 429–438. IEEE Computer Society, 2013. https://doi.org/10.1109/FOCS.2013.53.

[14] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam D. Smith. Calibrating noise to sensitivity in private data analysis. In Shai Halevi and Tal Rabin, editors, *Theory of Cryptography, Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006, Proceedings*, volume 3876 of *Lecture Notes in Computer Science*, pages 265–284. Springer, 2006. https://doi.org/10.1007/11681878_14.

[15] Cynthia Dwork, Moni Naor, Toniann Pitassi, Guy N. Rothblum, and Sergey Yekhanin. Pan-private streaming algorithms. In Andrew Chi-Chih Yao, editor, *Innovations in Computer Science - ICS 2010, Tsinghua University, Beijing, China, January 5-7, 2010. Proceedings*, pages 66–80. Tsinghua University Press, 2010. URL: http://conference.iiis.tsinghua.edu.cn/ICS2010/content/papers/6.html.

[16] Úlfar Erlingsson, Vitaly Feldman, Ilya Mironov, Ananth Raghunathan, Kunal Talwar, and Abhradeep Thakurta. Amplification by shuffling: From local to central differential privacy via anonymity. In Timothy M. Chan, editor, *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019*, pages 2468–2479. SIAM, 2019. https://doi.org/10.1137/1.9781611975482.151.

[17] Badih Ghazi, Noah Golowich, Ravi Kumar, Rasmus Pagh, and Ameya Velingker. On the power of multiple anonymous messages: Frequency estimation and selection in the shuffle model of differential privacy. In Anne Canteaut and François-Xavier Standaert, editors, *Advances in Cryptology - EUROCRYPT 2021 - 40th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Zagreb, Croatia, October 17-21, 2021, Proceedings, Part III*, volume 12698 of *Lecture Notes in Computer Science*, pages 463–488. Springer, 2021. https://doi.org/10.1007/978-3-030-77883-5_16.

[18] Badih Ghazi, Pasin Manurangsi, Rasmus Pagh, and Ameya Velingker. Private aggregation from fewer anonymous messages. In Anne Canteaut and Yuval Ishai, editors, *Advances in Cryptology - EUROCRYPT 2020 - 39th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Zagreb, Croatia, May 10-14, 2020, Proceedings, Part II*, volume 12106 of *Lecture Notes in Computer Science*,

pages 798–827. Springer, 2020. https://doi.org/10.1007/978-3-030-45724-2_27.

[19] Slawomir Goryczka and Li Xiong. A comprehensive comparison of multiparty secure additions with differential privacy. *IEEE Trans. Dependable Secur. Comput.*, 14(5):463–477, 2017. https://doi.org/10.1109/TDSC.2015.2484326.

[20] David Haussler. Decision theoretic generalizations of the PAC model for neural net and other learning applications. *Inf. Comput.*, 100(1):78–150, 1992. https://doi.org/10.1016/0890-5401(92)90010-D.

[21] Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam D. Smith. What can we learn privately? *SIAM J. Comput.*, 40(3):793–826, 2011. https://doi.org/10.1137/090756090.

[22] Michael J. Kearns. Efficient noise-tolerant learning from statistical queries. *J. ACM*, 45(6):983–1006, 1998. https://doi.org/10.1145/293347.293351.

[23] Elaine Shi, T.-H. Hubert Chan, Eleanor Gilbert Rieffel, Richard Chow, and Dawn Song. Privacy-preserving aggregation of time-series data. In *Proceedings of the Network and Distributed System Security Symposium, NDSS 2011, San Diego, California, USA, 6th February - 9th February 2011*. The Internet Society, 2011. URL: https://www.ndss-symposium.org/ndss2011/privacy-preserving-aggregation-of-time-series-data.

[24] Leslie G. Valiant. A theory of the learnable. *Commun. ACM*, 27(11):1134–1142, 1984. https://doi.org/10.1145/1968.1972.