

Actually, All Learnable Classes are Privately Learnable (self-contained countable case)

1 Introduction

Let $\mathcal{X}$ be the domain of examples, and denote by $\Delta(\mathcal{X})$ the set of all probability distributions on $\mathcal{X}$. Any $\mathcal{Q} \subseteq \Delta(\mathcal{X})$ is referred to as a *distribution class*. Denote by TV the total variation distance.

Definition 1.1. $\mathcal{Q} \subseteq \Delta(\mathcal{X})$ is *learnable* if for every $\alpha, \beta > 0$, there exists $m \in \mathbb{N}$, and a learner $A : \mathcal{X}^m \rightarrow \Delta(\mathcal{X})$ such that for all $p \in \mathcal{Q}$,

$$\Pr_{S \sim p^m} [\text{TV}(A(S), p) \geq \alpha] \leq \beta.$$

The choice of $m(\alpha, \beta)$ is called the *sample complexity* of the learner A .

Ashtiani conjectured that learnable distribution classes can be learned under (ε, δ) -differential privacy [Ash22]. Bun, Kamath, Mouzakis, and Singhal [BKMS24] provide a partial refutation: they give a counterexample when considering learnability up to constant error, instead of the all-scale learnability of Definition 1.1. Afzali, Ashtiani, and Liaw remark that the conjecture is more likely to hold in the agnostic setting [AAL24, AAL25]. We result proved here is a positive answer in the original realizable setting. The approach is essentially identical to Theorem 4 of [KKMV23], except considering distribution learners instead of PAC learners.

Theorem 1.2. *Let $\mathcal{X}$ be a countable domain. Suppose $\mathcal{Q} \subseteq \Delta(\mathcal{X})$ is learnable. For any $\alpha, \beta, \varepsilon, \delta > 0$, there exists $n \in \mathbb{N}$ and an (ε, δ) -differentially private learner $A : \mathcal{X}^n \rightarrow \Delta(\mathcal{X})$, such that for all $p \in \mathcal{Q}$,*

$$\Pr_{S \sim p^n} [\text{TV}(A(S), p) \geq \alpha] \leq \beta.$$

In terms of the non-private learner's sample complexity $m(\alpha, \beta)$, the differentially private learner's sample complexity is $n = O\left(\frac{m(\frac{\beta}{64m(\alpha, \beta/4)}, \frac{1}{16}) \log(\frac{1}{\beta\delta})}{\varepsilon}\right)$. If $m = \text{poly}(\frac{1}{\alpha}, \log \frac{1}{\beta})$, $n = \text{poly}(\frac{1}{\alpha}, \frac{1}{\beta}, \frac{1}{\varepsilon}, \log \frac{1}{\delta})$.

The sample complexity grows with $\log(1/\delta)$ which corresponds to a non-vacuous ($\delta = o(1/n)$) privacy regime. Note the exponential blowup in the failure probability β (which is fixed in the full paper).

2 Non-private-to-private conversion for distribution learners

Algorithm 1 Private learner A

Input: Samples $S \sim p^n$ where $p \in \mathcal{Q}$ is unknown. The non-private learner L .

Output: A distribution q close to p in total variation.

- 1: Partition S into k partitions $S_1, \dots, S_k \in \mathcal{X}^b$.
 - 2: Compute $q_i := L(S_i) \in \Delta(\mathcal{X})$ for all $i \in [k]$.
 - 3: Draw shared randomness r independent of the input data.
 - 4: **for** $i \in [k]$ **do**
 - 5: Sample the synthetic dataset $Z_i := (z_i^{(1)}, \dots, z_i^{(\ell)}) \in \mathcal{X}^\ell$ from q_i^ℓ under shared randomness r .
 - 6: **end for**
 - 7: Compute $Z = \text{StabilityBasedMode}(Z_1, \dots, Z_k)$.
 - 8: **return** $L(Z)$
-

2.1 Approach

Algorithm 1 describes the private learner A in the statement of Theorem 1.2 (omitting the parameter settings of $\alpha, \beta, \epsilon, \delta$ used).

At a high-level, the reduction works by splitting the input sample S into partitions S_i . The non-private learner is applied to each partition, resulting in candidate solutions q_i . From each candidate solution q_i , we sample a synthetic dataset Z_i .

The key ingredient is coordinated sampling. The below Fact 2.1 allows us to sample all the Z_i from their respective marginals q_i while minimizing disagreement.

Fact 2.1 (Coupling, Proposition 6 from [AS19]). *Let $\mathcal{X}$ be a countable set. Let $\mathcal{W}$ be a collection of random variables taking values in $\mathcal{X}$. There exists a coupling such that for all $X, Y \in \mathcal{W}$,*

$$\Pr[X \neq Y] \leq \frac{2TV(X, Y)}{1 + TV(X, Y)} \leq 2TV(X, Y).$$

If the returned candidate q_i 's are close enough to the unknown p , Fact 2.1 allows most partitions to return the same synthetic dataset that hypothetically would be sampled from p under the same shared randomness. Since enough partitions release the same synthetic dataset, we can apply the stability-based mode mechanism to release it privately.

Fact 2.2 (Stability-based mode, Proposition 3.4 from [Vad17]; [TS13]). *Let $\mathcal{X}$ denote any domain. For every $\epsilon, \delta > 0$ and $m \in \mathbb{N}$, there exists an (ϵ, δ) -DP mechanism $M : \mathcal{X}^m \rightarrow \mathcal{X}$ such that if the difference in counts between the mode of $\mathcal{X}^m$ and the second most frequently occurring element is $\geq 4 \log(1/\delta)/\epsilon$, $M(x)$ outputs the mode with probability $\geq 1 - \delta$.*

Finally, the non-private learner is applied to the differentially private synthetic data.

2.2 Proof of the main result

Proof of Theorem 1.2. Let L be the non-private learner, and denote its sample complexity as $m(\alpha, \beta)$. We define a few terms:

- We receive a sample S of size n from p , which we divide into k partitions of size b , so $n = bk$.
- Let $\ell := m(\alpha, \frac{\beta}{4})$ denote the target synthetic dataset size.
- Let $b := m(\eta = \frac{\beta}{64\ell}, \frac{1}{16})$ denote the size of each partition.
- Let k be the number of partitions, which can be set to $k = C \log(1/\beta\delta)/\epsilon$ for some constant C .

We start by using Fact 2.1 to create a coupling. Let $\mathcal{H}$ denote the range of the learner L when invoked on b samples and targeting a $(\eta, \frac{1}{16})$ error guarantee. Now, for every $q \in (\mathcal{H} \cup \mathcal{Q}) \subseteq \Delta(\mathcal{X})$, the product distribution q^ℓ outputs samples Z_q in $\mathcal{X}^\ell$ which is indeed a countable set. So there exists a source of shared randomness $\mathcal{R}$, such that for all $p, q \in \mathcal{H} \cup \mathcal{Q}$,

$$\Pr_{r \in \mathcal{R}}[Z_p \neq Z_q] \leq 2TV(p^\ell, q^\ell) \quad \text{with} \quad Z_p \sim p^\ell, Z_q \sim q^\ell.$$

Now, on an input sample S drawn from our unknown p , we invoke L on the partitions $S_1, \dots, S_k$ of size b producing $q_1, \dots, q_k$. For all i , $\Pr[TV(q_i, p) \geq \eta] \leq 1/16$. Since the $q_1, \dots, q_k$ are independent,

$$\Pr[|\{q_i : TV(q_i, p) \geq \eta\}| \geq k/8] \leq \exp(-k/128) \leq \beta/4 \tag{1}$$

by Hoeffding's and our choice of k .

Sample our shared randomness r independently of the input data, and use our a-priori constructed coupling to sample our synthetic datasets $Z_i \in \mathcal{X}^\ell$. Only for analysis, consider $Z_* \sim p^\ell$ sampled under the same shared randomness. Note that for all i ,

$$\Pr[Z_i \neq Z_*] \leq 2 \cdot TV(q_i^\ell, p^\ell) \leq 2\ell \cdot TV(q_i, p).$$

Consider the indicator random variable I_i corresponding to the event that $Z_i \neq Z_*$ with $TV(q_i, p) \leq \eta$. We have that

$$\mathbb{E}[\sum_{i=1}^k I_i] \leq 2\ell k \eta = \beta k / 32$$

An application of Markov's inequality gives that $\Pr[\sum_{i=1}^k I_i \geq k/8] \leq \beta/4$. Conditioned on our event that at least $7/8$'s of the q_i are within η of p , we have that the number of Z_i with $Z_i = Z_*$ is at least $7k/8 - k/8 = 3k/4$. Hence, the gap between the mode and the second most frequent item is at least $k/2$. Selecting k such that

$$k/2 \geq 4 \log(4/\beta)/\varepsilon + 4 \log(1/\delta)/\varepsilon, \quad (2)$$

the stability-based mode mechanism of Fact 2.2 targeting $(\varepsilon, \min\{\delta, \beta/4\})$ -DP yields Z_* with probability $\geq 1 - \beta/4$. Taking a union bound, our algorithm outputs a sample that matches what p^ℓ would produced under the same shared randomness, with probability at least $1 - 3\beta/4$. Hence, our learner L targeting $(\alpha, \beta/4)$ -error succeeds with probability $\geq 1 - \beta$. It can be verified that our choice of k satisfies (1) and (2).

To check privacy, note that for any setting of the shared randomness r , changing a single entry in the input sample S only changes a single one of the q_i , and therefore only a single one of the Z_i . The output of stability-based mode always satisfies an (ε, δ) -DP guarantee if only a single one of its inputs changes. $\square$

3 Discussion

The approach is essentially identical to Theorem 4 of [KKMV23], adapted to distribution learning instead of PAC learning. It goes through the same step of converting TV closeness to differential privacy. The only differences are: (1) We get TV closeness for free, as its guaranteed by successful distribution learners; (2) we use synthetic datasets to coordinate closeness instead of sampled classifiers.

The same technique of employing coordinated sampling to improve the private-utility tradeoff of subsample-and-aggregate was used in HotPATE [CCWL⁺26] in the context of sampling from language models with differential privacy (also independently proposed by Steurer [Ste25] for the same problem). Ghazi, Kumar, and Manurangsi [GKM21] use the same technique to convert similar hypotheses to the same hypothesis in the setting of user-level PAC learning. The reductions approach of applying a non-private learner to partitions of samples and aggregating results privately was explored by Ashtiani and Liaw [AL22]. Rather than aggregating hypotheses, the present paper aggregates synthetic data produced by the hypotheses, which is more general but incurs a larger blow-up in sample complexity.

References

- [AAL24] Mohammad Afzali, Hassan Ashtiani, and Christopher Liaw. Mixtures of gaussians are privately learnable with a polynomial number of samples. In *Proceedings of The 35th International Conference on Algorithmic Learning Theory*, 2024.
- [AAL25] Mohammad Afzali, Hassan Ashtiani, and Christopher Liaw. Agnostic private density estimation for gmms via list global stability. In *Proceedings of The 36th International Conference on Algorithmic Learning Theory*, 2025.
- [AL22] Hassan Ashtiani and Christopher Liaw. Private and polynomial time algorithms for learning gaussians and beyond. In *Proceedings of Thirty Fifth Conference on Learning Theory*, 2022.
- [AS19] Omer Angel and Yinon Spinka. Pairwise optimal coupling of multiple random variables, 2019.
- [Ash22] Hassan Ashtiani. Private learning of gaussians and their mixtures. <https://www.youtube.com/watch?v=bmNjm01x50I>, July 2022.
- [BKMS24] Mark Bun, Gautam Kamath, Argyris Mouzakis, and Vikrant Singhal. Not all learnable distribution classes are privately learnable. In *Proceedings of The 35th International Conference on Algorithmic Learning Theory*, 2024.

- [CCWL⁺26] Edith Cohen, Benjamin Cohen-Wang, Xin Lyu, Jelani Nelson, Tamas Sarlos, and Uri Stemmer. Hot PATE: Private aggregation of distributions for diverse tasks. In *The Fourteenth International Conference on Learning Representations*, 2026.
- [GKM21] Badih Ghazi, Ravi Kumar, and Pasin Manurangsi. User-level differentially private learning via correlated sampling. *Advances in Neural Information Processing Systems*, 2021.
- [KKMV23] Alkis Kalavasis, Amin Karbasi, Shay Moran, and Grigoris Velegkas. Statistical indistinguishability of learning algorithms. In *Proceedings of the 40th International Conference on Machine Learning*, ICML '23, pages 15586–15622. JMLR, Inc., 2023.
- [Ste25] David Steurer, 2025. Personal communication.
- [TS13] Abhradeep Guha Thakurta and Adam Smith. Differentially private feature selection via stability arguments, and the robustness of the lasso. In *Proceedings of the 26th Annual Conference on Learning Theory*, 2013.
- [Vad17] Salil P. Vadhan. The complexity of differential privacy. In *Tutorials on the Foundations of Cryptography*, 2017.