Fast, Sample-Efficient, Affine-Invariant Private Mean and Covariance Estimation for Subgaussian Distributions Extended Abstract

Gavin Brown[*] GRBROWN@BU.EDU
Samuel B. Hopkins[†] SAMHOP@MIT.EDU
Adam Smith[*] ADS22@BU.EDU

Editors: Gergely Neu and Lorenzo Rosasco

We present a fast, differentially private algorithm for high-dimensional *covariance-aware* mean estimation with nearly optimal sample complexity.[1] Only exponential-time estimators were previously known to achieve this guarantee. Given n samples from a (sub-)Gaussian distribution with unknown mean μ and covariance Σ, our (ε, δ)-differentially private estimator produces $\tilde{\mu}$ such that $\|\mu - \tilde{\mu}\|_\Sigma \leq \alpha$ with high probability as long as $n \gtrsim \frac{d}{\alpha^2} + \frac{d\sqrt{\log 1/\delta}}{\alpha\varepsilon} + \frac{d \log 1/\delta}{\varepsilon}$. The Mahalanobis error metric $\|\mu - \hat{\mu}\|_\Sigma$ measures the distance between $\hat{\mu}$ and μ relative to Σ; it characterizes the error of the sample mean. This sample complexity is close to optimal, nearly matching the known lower bound of $n \gtrsim \frac{d}{\alpha^2} + \frac{d}{\alpha\varepsilon} + \frac{\log 1/\delta}{\varepsilon}$. Our algorithm runs in time $\tilde{O}(nd^{\omega-1} + nd/\varepsilon)$, where $\omega < 2.38$ is the matrix multiplication exponent. For modest privacy parameters, the running time is dominated by the time to compute the covariance of the data.

Adapting an exponential-time approach of Brown, Gaboardi, Smith, Ullman, and Zakynthinou (2021), our work introduces a pair of efficient and stable subroutines for nonprivate mean and covariance estimation. Two key technical innovations underlie their analysis. First, we use new notions of "outlier-free subsets" which admit efficient greedy algorithms. Second, we introduce a technique that finds a family of outlier-free subsets across a range of outlier thresholds. Through an elementary but subtle argument, we prove strong relationships between the families of subsets found on any two adjacent data sets.

Our stable covariance estimator can be turned to private covariance estimation for unrestricted subgaussian distributions. With $n \gtrsim d^{3/2}$ samples, our estimate is accurate in spectral norm. This is the first such algorithm using $n = o(d^2)$ samples, answering an open question posed by Alabi et al. (2023). With $n \gtrsim d^2$ samples, our estimate is accurate in Frobenius norm. This leads to a fast, nearly optimal algorithm for private learning of unrestricted Gaussian distributions in TV distance

Duchi, Haque, and Kuditipudi (2023) obtained similar results independently and concurrently.

---

[*] Department of Computer Science, Boston University. Supported by NSF Awards CNS-2120667 and CCF-1763786 and an Apple Faculty Award.

[†] Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology. Supported in part by funds from the MLA@CSAIL initiative and by NSF Award CCF-2238080.

1. Extended abstract. Full version appears as [arxiv:2301.12250, v3]

References

Daniel Alabi, Pravesh K Kothari, Pranay Tankala, Prayaag Venkat, and Fred Zhang. Privately estimating a gaussian: Efficient, robust and optimal. In *Proceedings of the fifty-fifth annual ACM symposium on Theory of computing*, 2023.

Gavin Brown, Marco Gaboardi, Adam Smith, Jonathan Ullman, and Lydia Zakynthinou. Covariance-aware private mean estimation without private covariance estimation. *Advances in Neural Information Processing Systems*, 34:7950–7964, 2021.

John Duchi, Saminul Haque, and Rohith Kuditipudi. A fast algorithm for adaptive private mean estimation. In *Conference On Learning Theory*. PMLR, 2023.