Proceedings of Machine Learning Research vol 178:1–2, 2022 35th Annual Conference on Learning Theory

Optimal Mean Estimation without a Variance

Yeshwanth Cherapanamjeri YESHWANTH@BERKELEY.EDU
Nilesh Tripuraneni NILESH_TRIPURANENI@BERKELEY.EDU
Peter L. Bartlett PETER@BERKELEY.EDU
Michael I. Jordan JORDAN@CS.BERKELEY.EDU
University of California, Berkeley

Editors: Po-Ling Loh and Maxim Raginsky

Abstract

We study the problem of heavy-tailed mean estimation in settings where the variance of the data-generating distribution does not exist. Concretely, given a sample $\boldsymbol{X} = \{X_i\}_{i=1}^n$ from a distribution $\mathcal{D}$ over $\mathbb{R}^d$ with mean μ which satisfies the following *weak-moment* assumption for some $\alpha \in [0, 1]$:

$$\forall \|v\| = 1 : \mathbb{E}_{X \sim \mathcal{D}}[|\langle X - \mu, v \rangle|^{1+\alpha}] \leqslant 1,$$

and given a target failure probability, δ, our goal is to design an estimator which attains the smallest possible confidence interval as a function of n, d, δ. For the specific case of $\alpha = 1$, foundational work of Lugosi and Mendelson exhibits an estimator achieving *optimal* subgaussian confidence intervals, and subsequent work has led to computationally efficient versions of this estimator. Here, we study the case of general α, and provide a precise characterization of the optimal achievable confidence interval by establishing the following information-theoretic lower bound:

$$\Omega\left(\sqrt{\frac{d}{n}} + \left(\frac{d}{n}\right)^{\frac{\alpha}{(1+\alpha)}} + \left(\frac{\log 1/\delta}{n}\right)^{\frac{\alpha}{(1+\alpha)}}\right).$$

and devising an estimator matching the aforementioned lower bound up to constants. Moreover, our estimator is computationally efficient.

The chief difficulty in establishing our lower bound is the lack of a natural family of distributions witnessing it. This is in stark contrast to the $\alpha = 1$ regime, i.e the finite variance setting, where a natural family of isotropic Gaussians suffice for all settings of n, d and δ. However, when $\alpha < 1$, such an approach and more generally, any approach based on a universal class of distributions provably fails. Therefore, we devise a carefully crafted set of high-dimensional discrete distributions obeying the weak-moment assumption specific to the precise settings of n, d and δ. A further complication induced by this choice is that due to the discrete nature of our family, the associated divergences between pairs of distributions from the family tend to be infinite precluding standard approaches to establishing such lower bounds such as those based on Fano's Inequality (see, for example, Wainwright (2019)). We overcome this difficulty through an *explicit* analysis of the posterior distribution in the standard framework for establishing Bayesian lower bounds.

Our upper bound bound is based on approaches from prior work (Hopkins, 2020; Cherapanamjeri et al., 2019) which in turn build on the inefficient variant in Lugosi and Mendelson (2019). In particular, we show that the framework based on non-convex gradient descent (Cherapanamjeri et al., 2019) can be adapted to the weak-moment setting with crucial modifications owing to the lack of a variance. The formal statements of our technical results and their proofs are provided in complete detail in the full version of our paper.[1]

1. Extended abstract. Full version appears as [https://arxiv.org/abs/2011.12433, v2].

References

Yeshwanth Cherapanamjeri, Nicolas Flammarion, and Peter L. Bartlett. Fast mean estimation with sub-gaussian rates. In Alina Beygelzimer and Daniel Hsu, editors, *Conference on Learning Theory, COLT 2019, 25-28 June 2019, Phoenix, AZ, USA*, volume 99 of *Proceedings of Machine Learning Research*, pages 786–806. PMLR, 2019. URL http://proceedings.mlr.press/v99/cherapanamjeri19b.html.

Samuel B. Hopkins. Mean estimation with sub-Gaussian rates in polynomial time. *Ann. Statist.*, 48(2):1193–1213, 2020. ISSN 0090-5364. doi: 10.1214/19-AOS1843. URL https://doi.org/10.1214/19-AOS1843.

G. Lugosi and S. Mendelson. Sub-Gaussian estimators of the mean of a random vector. *Ann. Statist.*, 47(2):783–794, 04 2019.

Martin J. Wainwright. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge University Press, Cambridge, 2019. doi: 10.1017/9781108627771. URL https://doi.org/10.1017/9781108627771.