Proceedings of Machine Learning Research vol 195:1–3, 2023 36th Annual Conference on Learning Theory

Open problem: $\log n$ factor in "Local Glivenko-Cantelli"

Doron Cohen DORONV@POST.BGU.AC.IL and **Aryeh Kontorovich** KARYEH@CS.BGU.AC.IL
Department of Computer Science
Ben-Gurion University of the Negev
Beer-Sheva, Israel

Editors: Gergely Neu and Lorenzo Rosasco

Abstract

Can the $\log(n)$ factor in the upper bound of Cohen and Kontorovich (COLT, 2023) be removed?

Introduction. Cohen and Kontorovich (2023) considered the following problem. Let Y_j, $j \in \mathbb{N}$ be a sequence of independent $\mathrm{Binomial}(n, p_j)$ random variables, where $n \in \mathbb{N}$ and $p_j \downarrow 0$ as $j \to \infty$. Since $\mathbb{E}\, Y_j = np_j$, let us consider the centered, normalized process $\bar{Y}_j := n^{-1} Y_j - p_j$. Finally, we define Δ_n to be the expected uniform absolute deviation:

$$\Delta_n \quad := \quad \mathbb{E} \sup_{j \in \mathbb{N}} |\bar{Y}_j|. \tag{1}$$

Cohen and Kontorovich gave an exact characterization of the $p \in [0, \frac{1}{2}]^{\mathbb{N}}_{\downarrow 0}$ (that is, $p \in [0, \frac{1}{2}]^{\mathbb{N}}$ with $p_j \downarrow 0$) for which $\Delta_n \to 0$ as $n \to \infty$. Namely, they defined the functional

$$T(p) \quad := \quad \sup_{j \in \mathbb{N}} \frac{\log(j+1)}{\log(1/p_j)}$$

and showed that $\Delta_n \to 0$ iff $T(p) < \infty$. They also gave the finite-sample bound

$$\Delta_n \quad \le \quad c \left(\sqrt{\frac{S(p)}{n}} + \frac{T(p) \log n}{n} \right), \qquad n \ge \mathrm{e}^3, \tag{2}$$

where $c > 0$ is an absolute constant and

$$S(p) \quad := \quad \sup_{j \in \mathbb{N}} p_j \log(j+1).$$

We conjecture that the $\log n$ factor in (2) is superfluous. This conjecture is motivated by the asymptotic lower bounds

$$\liminf_{n \to \infty} \sqrt{n} \Delta_n \quad \ge \quad c\sqrt{S(p)}, \tag{3}$$

$$\liminf_{n \to \infty} n \Delta_n \quad \ge \quad cT(p), \tag{4}$$

where $c > 0$ is a universal constant. This shows that beyond the conjecturally removable $\log n$ factor, (2) is tight up to constants.

Open problem. Is the $\log n$ factor in (2) necessary, or can it be removed?

Background and motivation. The open problem posed above is a concise version of the following natural mean estimation problem. Estimating the mean of a random variable $X \in \mathbb{R}^d$ from a sample of independent draws X_i is among the most basic problems of statistics. Much of the theory has focused on obtaining efficient estimators $\hat{m}_n$ of the true mean m and analyzing the decay of $\|\hat{m}_n - m\|_2$ as a function of sample size n, dimension d, and various moment assumptions on X (Devroye et al., 2016; Lugosi and Mendelson, 2019a,b; Cherapanamjeri et al., 2019, 2020; Lugosi and Mendelson, 2021). Inspired by Thomas (2018), Cohen and Kontorovich (2023) considered a distribution μ on $\{0,1\}^d$ with mean $p \in [0,1]^d$. Given n iid draws of $X_i \sim \mu$, they denote by $\hat{p}_n = n^{-1}\sum_{i=1}^n X_i$ the empirical mean. The central quantity of interest studied by Cohen and Kontorovich is the uniform absolute deviation

$$\Delta_n(\mu) := \mathbb{E}\,\|\hat{p}_n - p\|_\infty = \mathbb{E}\max_{j\in[d]} |\hat{p}_n(j) - p(j)|. \tag{5}$$

The ℓ_∞ norm in (5) is in some sense the most interesting of the ℓ_r norms; indeed, for $r < \infty$, $\Delta_n^{(r)} := \mathbb{E}\,\|\hat{p}_n - p\|_r^r$ decomposes into a sum of expectations and the condition $\Delta_n^{(r)} \to 0$ reduces to one of convergence of the appropriate series.

Cohen and Kontorovich (2023) obtained an almost complete understanding of the behavior of $\Delta_n(\mu)$ in the case of product measures (i.e., where the X_i are independent) — modulo the troublesome $\log n$ factor in the Open Problem. Indeed, the $\hat{p}_n - p$ in (5) is exactly the $\bar{Y}_j$ in (1).

Our restriction of $p \in [0,1]^{\mathbb{N}}$ to the range $[0, \frac{1}{2}]$ and requirement that $p(j) \downarrow 0$ incur no loss of generality. Indeed, the range restriction is justified since $|u - v| = |(1-u) - (1-v)|$. Furthermore, if $p(j) \not\to 0$ as $j \to \infty$ then certainly $T(p) = \infty$ and hence (4) implies that $\Delta_n \not\to 0$ as $n \to \infty$. Now whenever $p(j) \to 0$, there is a non-increasing permutation $p^\downarrow$, and since $\Delta_n(p) = \Delta_n(p^\downarrow)$, there is indeed no loss of generality in assuming that the decay is monotone.

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.

Yeshwanth Cherapanamjeri, Nilesh Tripuraneni, Peter L. Bartlett, and Michael I. Jordan. Optimal mean estimation without a variance. *CoRR*, abs/2011.12433, 2020. URL https://arxiv.org/abs/2011.12433.

Doron Cohen and Aryeh Kontorovich. Local Glivenko-Cantelli. In *Conference on Learning Theory*, Proceedings of Machine Learning Research, 2023.

Luc Devroye, Matthieu Lerasle, Gabor Lugosi, and Roberto I. Oliveira. Sub-Gaussian mean estimators. *The Annals of Statistics*, 44(6):2695 – 2725, 2016. doi: 10.1214/16-AOS1440. URL https://doi.org/10.1214/16-AOS1440.

Gábor Lugosi and Shahar Mendelson. Sub-Gaussian estimators of the mean of a random vector. *The Annals of Statistics*, 47(2):783 – 794, 2019a. doi: 10.1214/17-AOS1639. URL https://doi.org/10.1214/17-AOS1639.

Gábor Lugosi and Shahar Mendelson. Mean estimation and regression under heavy-tailed distributions: A survey. *Found. Comput. Math.*, 19(5):1145–1190, 2019b. doi: 10.1007/s10208-019-09427-x. URL https://doi.org/10.1007/s10208-019-09427-x.

Gábor Lugosi and Shahar Mendelson. Robust multivariate mean estimation: The optimality of trimmed mean. *The Annals of Statistics*, 49(1):393 – 410, 2021. doi: 10.1214/20-AOS1961. URL https://doi.org/10.1214/20-AOS1961.

Thomas. Is uniform convergence faster for low-entropy distributions? Theoretical Computer Science Stack Exchange, 2018. URL https://cstheory.stackexchange.com/q/42009. URL:https://cstheory.stackexchange.com/q/42009 (version: 2018-12-10).