

Open Problem: Is Interaction Necessary for Order-Optimal 1-bit Mean Estimation?

Ivan Lau

National University of Singapore

IVAN.LAU@U.NUS.EDU

Jonathan Scarlett

National University of Singapore

SCARLETT@COMP.NUS.EDU.SG

Editors: Steve Hanneke and Tor Lattimore

Abstract

We ask whether interaction is necessary for order-optimal 1-bit mean estimation over nonparametric finite-moment classes. Adaptive threshold-query protocols achieve the order-optimal 1-bit minimax rate, and the same rate is attainable with general 1-bit queries using only one adaptive transition (i.e., two stages of querying). In the non-adaptive setting, threshold and interval queries are known to be highly suboptimal, but the case of arbitrary non-adaptive quantizers remains unresolved. Can such quantizers match the adaptive rate, yielding an optimal one-shot protocol? Or is the known two-stage estimator stage-optimal, with a single adaptive transition being necessary and sufficient?

Keywords: Communication constraints; 1-bit estimation; adaptivity; non-interactive protocols

1. Motivation

Mean estimation is one of the most fundamental primitives in statistics, machine learning, and theoretical computer science. In many modern data-acquisition systems, samples are not observed directly by the learner. Instead, they are held by agents, such as low-power sensors, edge devices, or clients in a federated system, that may transmit only a few bits per observation. A fundamental limiting case is 1-bit communication: each sample is compressed to a single binary message before being transmitted to the learner.

At first glance, one bit per sample may appear too restrictive for nonparametric inference. However, recent work shows that this intuition is misleading. With interaction/adaptivity (i.e., 1-bit queries that may be designed based on the previous bits observed), 1-bit mean estimation remains nearly as sample-efficient as *unquantized* mean estimation: adaptive 1-bit estimators match the unquantized minimax rate up to at most one (unavoidable) logarithmic factor. In fact, simple adaptive threshold queries already suffice (Lau and Scarlett, 2026b), and the resulting rates are order-optimal in the 1-bit model. However, the 1-bit constraint itself is not the whole story, and a more refined question is: *Is interaction/adaptivity truly necessary to achieve this order-optimal rate?*

We study this question in a nonparametric setting. The learner receives one bit per sample and must estimate the mean of an unknown distribution whose mean lies in a known bounded interval and whose k -th central moment is bounded. The quantizers may either be chosen *adaptively*, using previously received bits, or fixed *non-adaptively* before any messages are observed. The fully non-adaptive model is operationally appealing: devices can be pre-programmed and communicate without feedback or multiple adaptive rounds.

The current landscape is sharp but incomplete. Fully non-adaptive threshold and interval queries are known to be highly suboptimal (Lau and Scarlett, 2026a), i.e., these query families cannot match the adaptive rate. However, this does not rule out non-adaptive protocols using general 1-bit quantizers, which are potentially much more expressive. If just one adaptive transition is allowed, then the order-optimal rate is known to be achievable using general 1-bit queries (Lau and Scarlett, 2026b). The unresolved case is therefore precisely the zero-adaptivity, general-query regime:

Can fully non-adaptive arbitrary 1-bit quantizers match the adaptive 1-bit minimax rate for nonparametric mean estimation, or is interaction necessary?

Either answer would be significant. A positive answer would give an order-optimal non-adaptive protocol for nonparametric 1-bit mean estimation. A negative answer would establish a genuine interaction/adaptivity gap, and certify that the known two-stage estimator has the fewest stages possible: one adaptive transition suffices, but zero adaptive transitions do not.

2. Problem Setup and Open Problem

Distributional assumption. We consider one-dimensional mean estimation over the nonparametric distribution family

$$\mathcal{D}(k, \lambda, \sigma) = \left\{ D : \mu(D) := \mathbb{E}_{X \sim D}[X] \in [-\lambda, \lambda], \quad \mathbb{E}_{X \sim D} |X - \mu(D)|^k \leq \sigma^k \right\},$$

where $k > 1$ and $\lambda \geq \sigma > 0$ are known to the learner. Apart from the tail/moment condition, the class is unrestricted: the distribution may be discrete, asymmetric, or supported on an unbounded set. The parameter λ is a search radius for the mean, while σ represents the intrinsic scale (e.g., the standard deviation when $k = 2$).

1-bit protocols. A 1-bit communication protocol observes independent samples $X_1, \dots, X_n \sim D$ only through binary messages $Y_t = \mathbf{1}\{X_t \in A_t\}$, where $A_t \subseteq \mathbb{R}$ is measurable. In an adaptive protocol, A_t may be chosen based on the entire past transcript $(A_1, Y_1, \dots, A_{t-1}, Y_{t-1})$ as well as any internal randomness. In a fully non-adaptive protocol, all sets A_1, \dots, A_n are fixed before any messages are observed, possibly using public or private randomness. Threshold and interval queries correspond to A_t being a half-line or an interval, respectively. General quantizers/queries can be highly nonlocal: for example, coding-based queries may have preimages that are unions of many intervals. We say that a protocol is (ϵ, δ) -accurate over $\mathcal{D}(k, \lambda, \sigma)$ if its output $\hat{\mu}$ satisfies

$$\sup_{D \in \mathcal{D}(k, \lambda, \sigma)} \mathbb{P} \{ |\hat{\mu} - \mu(D)| > \epsilon \} \leq \delta,$$

where the probability is over the samples and any internal randomness of the protocol.

Order-optimal rate. For ϵ below a sufficiently small constant multiple of σ and $\delta \in (0, 1/2)$, the adaptive 1-bit minimax sample complexity is known, up to constants depending only on k , to be

$$r_k(\lambda, \sigma, \epsilon, \delta) := \underbrace{\log \frac{\lambda}{\sigma}}_{\text{localization}} + \underbrace{\begin{cases} \frac{\sigma^2}{\epsilon^2} \cdot \log \frac{1}{\delta}, & k > 2, \\ \frac{\sigma^2}{\epsilon^2} \cdot \log \frac{\sigma}{\epsilon} \cdot \log \frac{1}{\delta}, & k = 2, \\ \left(\frac{\sigma}{\epsilon}\right)^{k/(k-1)} \cdot \log \frac{1}{\delta}, & 1 < k < 2. \end{cases}}_{\text{refinement}} \quad (1)$$

This rate is attainable by adaptive 1-bit protocols and is minimax optimal among 1-bit protocols (Lau and Scarlett, 2026b). The additive term $\log(\lambda/\sigma)$ corresponds to the cost of *localizing* the mean from $[-\lambda, \lambda]$ to an interval of length $O(\sigma)$. The remaining term is the *refinement* cost once such an interval has been identified. For $k \neq 2$, this matches the usual unquantized minimax rate up to the localization term. For $k = 2$, the extra $\log(\sigma/\epsilon)$ factor is an unavoidable 1-bit penalty. The open problem is whether fully non-adaptive arbitrary 1-bit quantizers can achieve the same rate.

Open Problem 1 (Order-optimal non-adaptive 1-bit mean estimation) Fix $k > 1$. Do there exist constants $c_k, C_k > 0$ such that, for all $\lambda \geq \sigma > 0$, all $0 < \epsilon \leq c_k \sigma$, and all $\delta \in (0, 1/2)$, there is a fully non-adaptive 1-bit protocol that is (ϵ, δ) -accurate over $\mathcal{D}(k, \lambda, \sigma)$ using at most $n \leq C_k r_k(\lambda, \sigma, \epsilon, \delta)$ samples?

3. Prior Work and Known Results

Classical mean estimation. Without communication constraints, high-probability mean estimation under moment assumptions is well understood. For $k \geq 2$, the optimal sample complexity is of order $(\sigma^2/\epsilon^2) \cdot \log(1/\delta)$; for $1 < k < 2$, it is of order $(\sigma/\epsilon)^{\frac{k}{k-1}} \cdot \log(1/\delta)$ – see (Devroye et al., 2016; Lee and Valiant, 2022; Cherapanamjeri et al., 2022; Minsker, 2023; Dang et al., 2023) and the references therein. These unquantized rates are the baseline for asking how much additional sample complexity is caused by 1-bit communication and by the absence of interaction.

Communication constraints and interaction. Communication-constrained estimation has been studied extensively, with the goal of characterizing the minimax rate when each agent can transmit only a limited number of bits (Zhang et al., 2013; Shamir, 2014; Braverman et al., 2016; Han et al., 2018; Duchi and Rogers, 2019; Barnes et al., 2020). A recurring question is whether interaction changes the statistical complexity. The answer is problem-dependent: interaction can improve sample complexity in tasks such as structured high-dimensional estimation, large-margin learning, and locally private hypothesis selection (Dagan and Feldman, 2020; Gopi et al., 2020; Acharya et al., 2022, 2023; Pour et al., 2024), whereas in some distributed testing problems it gives no improvement beyond constants (Kazemi et al., 2025). These works motivate Problem 1, but they do not resolve it: the present setting is scalar rather than high-dimensional, nonparametric rather than a fixed parametric model, and constrained to exactly one bit per sample.

Parametric 1-bit mean estimation. Order-optimal non-adaptive 1-bit protocols are known for parametric families, including Gaussian, symmetric log-concave, and scale-location models (Kipnis and Duchi, 2022; Cai and Wei, 2024; Kumar and Vatedka, 2026). These results show that non-adaptive 1-bit estimation can be order-optimal when the distribution has exploitable structure. For example, in Gaussian mean estimation, Gray-code-type quantizers can encode location information across multiple spatial scales (Cai and Wei, 2024). However, these protocols rely on model-specific information such as likelihood shape, symmetry, parametric CDFs, or scale-location structure, and do not apply to the nonparametric class $\mathcal{D}(k, \lambda, \sigma)$.

Nonparametric 1-bit mean estimation. For $\mathcal{D}(k, \lambda, \sigma)$, the adaptive case is settled. Lau and Scarlett (2026b, Theorem 5) give an adaptive threshold-query estimator with order-optimal sample complexity (1). Although this estimator uses only (randomized) threshold queries, its noisy binary search based localization step is sequential, requiring $O(\log(\lambda/\sigma))$ adaptive localization steps. Allowing general 1-bit queries greatly reduces this adaptivity requirement: Lau and Scarlett (2026b,

Section 4.3) give an order-optimal two-stage protocol whose first stage non-adaptively localizes the mean to an $O(\sigma)$ interval using coding-based queries, and whose second stage, chosen *after* this interval is decoded, refines the estimate within it. Thus the two-stage estimator achieves order-optimality by separating localization from location-dependent refinement. Problem 1 asks whether this single adaptive transition is truly necessary.

The fully non-adaptive case is less understood. For $k = 2$, stochastic quantization, equivalently randomized threshold querying, gives some non-adaptive upper bounds. Along the lines of [Abdalla and Chen \(2026\)](#), truncating to a range of size $B = \Theta(\max\{\lambda, \sigma^2/\epsilon\})$ and then stochastically quantizing gives sample complexity of order $O((B^2/\epsilon^2) \cdot \log(1/\delta))$, which is at least quadratic in λ . On the lower bound side, [Lau and Scarlett \(2026a\)](#) show that non-adaptive threshold and interval queries are suboptimal: their sample complexity must scale linearly with λ/σ , whereas adaptive threshold queries incur only logarithmic dependence. This adaptivity gap, however, does not rule out arbitrary measurable 1-bit queries. For example, consider i.i.d. randomized Fourier-feature queries $Z_t = 2\mathbf{1}\{\cos(S_t X_t + \Theta_t) \geq U_t\} - 1$, where $S_t \sim \text{Unif}[-a, a]$, $\Theta_t \sim \text{Unif}[0, 2\pi]$, and $U_t \sim \text{Unif}[-1, 1]$. Let G_ϵ be an $O(\epsilon)$ -spaced grid of $[-\lambda, \lambda]$ and decode by $\hat{\mu} \in \arg \min_{\nu \in G_\epsilon} \sum_{t=1}^n (Z_t - \cos(S_t \nu + \Theta_t))^2$. The main population gap at distance ϵ is $\Omega(a^2 \epsilon^2)$, while a direct second-order bias bound gives $|\mathbb{E}[Z_t | S_t, \Theta_t] - \cos(S_t \mu + \Theta_t)| = O(S_t^2 \sigma^2)$. Thus this analysis requires $a \lesssim \epsilon/\sigma^2$, and Hoeffding’s inequality plus a union bound over $|G_\epsilon| = \Theta(\lambda/\epsilon)$ gives $O((\sigma/\epsilon)^8 (\log(\lambda/\epsilon) + \log(1/\delta)))$ samples. Hence the scheme has the desired logarithmic dependence on λ , but a highly suboptimal refinement rate compared to (1). Overall, known non-adaptive 1-bit estimators use i.i.d. queries that do not provide refinement queries tailored to the localized interval, which seems to be the cause of poor scaling in either λ or σ/ϵ .

4. Discussion and Technical Barriers

The known results leave a delicate gap. In the two-stage estimator of [Lau and Scarlett \(2026b\)](#), the only adaptive transition occurs when the refinement-query block is chosen after a coarse interval is decoded from the localization bits. A natural route to a positive answer to Problem 1 would therefore be a universal refinement scheme that is valid across all possible localized intervals. A fully non-adaptive protocol must choose all refinement queries in advance; only the final decoder may use the decoded interval to “reinterpret” the bits. For very light-tailed classes, one might hope to use periodic quantizers so that, after the coarse location is decoded, the same bits can be reinterpreted as local refinement information. For the considerably larger class $\mathcal{D}(k, \lambda, \sigma)$, however, the main difficulty is tail control. Under only moment assumptions, far-tail samples can alias into central regions of a fixed global query and corrupt small-scale estimates. The two-stage estimator avoids this by centering tail-sensitive refinement queries around the localized interval. A fully non-adaptive estimator would need comparable local variance control without knowing the interval in advance.

Conversely, a negative answer would require a lower bound ruling out such universal refinement schemes. Existing lower bounds for interval queries exploit locality: a non-adaptive protocol must spread informative tests across the $\Theta(\lambda/\sigma)$ possible mean locations. Arbitrary measurable quantizers do not obey this locality constraint; a single query can encode many locations at once, as in the coding-based localization stage of [Lau and Scarlett \(2026b\)](#). Hence standard packing arguments are insufficient: one would need to prove that simultaneous non-adaptive localization, refinement, and tail control necessarily incur an additional order-wise sample cost.

References

- Pedro Abdalla and Junren Chen. Robust mean estimation under quantization. *arXiv preprint arXiv:2601.07074*, 2026.
- Jayadev Acharya, Clément L Canonne, Ziteng Sun, and Himanshu Tyagi. The role of interactivity in structured estimation. In *Conference on Learning Theory*, pages 1328–1355. PMLR, 2022.
- Jayadev Acharya, Clément L Canonne, Ziteng Sun, and Himanshu Tyagi. Unified lower bounds for interactive high-dimensional estimation under information constraints. *Advances in Neural Information Processing Systems*, 36:51133–51165, 2023.
- Leighton Pate Barnes, Yanjun Han, and Ayfer Özgür. Lower bounds for learning distributions under communication constraints via Fisher information. *Journal of Machine Learning Research*, 21:1–30, 2020. ISSN 1532-4435.
- Mark Braverman, Ankit Garg, Tengyu Ma, Huy L. Nguyen, and David P. Woodruff. Communication lower bounds for statistical estimation problems via a distributed data processing inequality. In *ACM Symposium on Theory of Computing*, pages 1011–1020, 2016.
- T Tony Cai and Hongji Wei. Distributed Gaussian mean estimation under communication constraints: Optimal rates and communication-efficient algorithms. *Journal of Machine Learning Research*, 25(37):1–63, 2024.
- Yeshwanth Cherapanamjeri, Nilesh Tripuraneni, Peter Bartlett, and Michael Jordan. Optimal mean estimation without a variance. In *Conference on Learning Theory*, pages 356–357. PMLR, 2022.
- Yuval Dagan and Vitaly Feldman. Interaction is necessary for distributed learning with privacy or communication constraints. In *ACM SIGACT Symposium on Theory of Computing*, pages 450–462, 2020.
- Trung Dang, Jasper Lee, Maoyuan’Raymond’ Song, and Paul Valiant. Optimality in mean estimation: Beyond worst-case, beyond sub-Gaussian, and beyond $1 + \alpha$ moments. In *Advances in Neural Information Processing Systems*, volume 36, pages 4150–4176, 2023.
- Luc Devroye, Matthieu Lerasle, Gabor Lugosi, and Roberto I. Oliveira. Sub-Gaussian mean estimators. *The Annals of Statistics*, 44(6):2695 – 2725, 2016.
- John Duchi and Ryan Rogers. Lower bounds for locally private estimation via communication complexity. In *Conference on Learning Theory*, pages 1161–1191. PMLR, 2019.
- Sivakanth Gopi, Gautam Kamath, Janardhan Kulkarni, Aleksandar Nikolov, Zhiwei Steven Wu, and Huanyu Zhang. Locally private hypothesis selection. In *Conference on Learning Theory*, pages 1785–1816. PMLR, 2020.
- Yanjun Han, Ayfer Özgür, and Tsachy Weissman. Geometric lower bounds for distributed parameter estimation under communication constraints. In *Conference on Learning Theory (COLT)*, volume 75, pages 3163–3188. PMLR, 2018.

- Hadi Kazemi, Ankit Pensia, and Jog Varun. The sample complexity of distributed simple binary hypothesis testing under information constraints. In *Conference on Learning Theory*, pages 3213–3214. PMLR, 2025.
- Alon Kipnis and John C Duchi. Mean estimation from one-bit measurements. *IEEE Transactions on Information Theory*, 68(9):6276–6296, 2022.
- Ritesh Kumar and Shashank Vatedka. One-bit distributed mean estimation with unknown variance. *Transactions on Machine Learning Research*, 2026.
- Ivan Lau and Jonathan Scarlett. Sequential 1-bit mean estimation with near-optimal sample complexity. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2026a.
- Ivan Lau and Jonathan Scarlett. Order-optimal sequential 1-bit mean estimation in general tail regimes. *arXiv preprint arXiv:2604.07796*, 2026b.
- Jasper CH Lee and Paul Valiant. Optimal sub-Gaussian mean estimation in \mathbb{R} . In *IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pages 672–683. IEEE, 2022.
- Stanislav Minsker. Efficient median of means estimator. In *Conference on Learning Theory (COLT)*, pages 5925–5933. PMLR, 2023.
- Alireza F Pour, Hassan Ashtiani, and Shahab Asoodeh. Sample-optimal locally private hypothesis selection and the provable benefits of interactivity. In *Conference on Learning Theory*, pages 4240–4275. PMLR, 2024.
- Ohad Shamir. Fundamental limits of online and distributed algorithms for statistical learning and estimation. In *Advances in Neural Information Processing Systems 27*, pages 163–171, 2014.
- Yuchen Zhang, John Duchi, Michael I. Jordan, and Martin J. Wainwright. Information-theoretic lower bounds for distributed statistical estimation with communication constraints. In *Advances in Neural Information Processing Systems 26*, pages 2328–2336, 2013.