Instance-Optimal Estimation with Multiple LLM Judges on a Budget

Junghyun Lee[*]
KAIST AI
jh_lee00@kaist.ac.kr

Sanghwa Kim[*]
KAIST AI
tkdghk9667@kaist.ac.kr

Yassir Jedra
ICL EEE
y.jedra@ic.ac.uk

Alexandre Proutière
KTH EECS, Digital Futures
alepro@kth.se

Se-Young Yun
KAIST AI
yunseyoung@kaist.ac.kr

Abstract

Evaluating large language models increasingly relies on LLM-as-a-judge protocols, but such evaluations remain costly: different judges have different prices and reliabilities, and the difficulty of each prompt-response pair can vary substantially. This raises a basic allocation question: under a fixed budget, how should one distribute evaluation queries across heterogeneous judges and instances to obtain the most accurate score estimates? We formalize this question as *budgeted heteroskedastic multi-judge estimation* under independent, unbiased score observations. Given K prompt-response pairs and J judges with known costs but unknown potentially heterogeneous query–judge variances, our goal is to estimate the underlying bounded score vector while minimizing an ℓ_p-error. Our first contribution is to analyze the inverse-variance weighted estimator (IVWE) and to derive the oracle allocation that minimizes its leading error bound. Since this allocation depends on the unknown variances, we then address the unknown-variance setting by proposing EST-IVWE, an adaptive algorithm that constructs and uses optimistic variance estimates to stabilize the empirical allocation. We prove that EST-IVWE matches the oracle IVWE rate up to lower-order terms in the budget. Our second and central theoretical contribution is a local minimax lower bound matching the leading allocation-dependent rate, establishing asymptotic instance-optimality up to logarithmic factors. An Assouad-type construction based on local, variance-preserving perturbations yields this bound both in expectation and with constant probability. A complementary Fano-type change-of-measure argument captures the logarithmic dependence on the confidence level with a different (suboptimal) allocation geometry. Finally, we numerically validate our approach on synthetic data and resampled HelpSteer2 evaluations, demonstrating improvements over naïve uniform allocation.

1 Introduction

Evaluating the quality of responses produced by large language models (LLMs) is a central component of modern model development, particularly for post-training (Bai et al., 2022; Guo et al., 2025; Ouyang et al., 2022; Rafailov et al., 2023) and model selection (Fernandez et al., 2026). Traditional evaluation methods, however, face important limitations. Human annotation remains the gold standard for capturing human preferences and judgments (Ouyang et al., 2022; Rafailov et al., 2023), but it is costly, time-consuming, and difficult to scale with the rapid pace of model development. Standardized benchmarks provide a more scalable alternative, but they are often too rigid for emerging applications

[*]Equal contributions

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

where suitable benchmarks do not yet exist. These limitations have motivated the growing use of the "LLM-as-a-judge" framework, in which strong frontier models are used to evaluate model responses across diverse tasks (Gu et al., 2026; Kim et al., 2024; Zheng et al., 2023).

While LLM-as-a-judge reduces reliance on costly human feedback, it still has computational and statistical concerns. Querying large frontier models at scale remains expensive, and different judges may have very different costs, expertise, and reliability (Kim et al., 2024; Li et al., 2024; Raju et al., 2024). A uniform allocation of evaluation queries is therefore typically inefficient: a small specialized judge may be more reliable than a costly general-purpose model on some domains, while using a frontier model for straightforward instances may waste budget (Fernandez et al., 2026). The difficulty is that, although the cost of each judge is usually known, the variability of its evaluations is not. This variability is *heteroskedastic*: it depends both on the judge and on the intrinsic difficulty of the prompt-response pair being evaluated. This raises the central question studied in this work:

Given a fixed evaluation budget, how should one allocate queries across prompt-response pairs and multiple LLM judges in order to obtain the most accurate estimates of response quality?

We formalize this question as a new statistical learning problem, referred to as *budgeted heteroskedastic multi-judge estimation*. This problem can be viewed as an adaptive, cost-aware generalization of Neyman allocation (Neyman, 1934). In classical Neyman allocation, samples are allocated across strata with known variances. Here, the learner must allocate evaluation queries across query–judge pairs with known costs but unknown, heteroskedastic variances, with the goal of estimating the full score vector as accurately as possible. Formally, given a total budget B, a set of queries $k \in [K] := \{1, \dots, K\}$, and a set of LLM judges $j \in [J]$, each judge j has a known query cost c_j, while each query–judge pair $(k, j) \in [K] \times [J]$ has an unknown variance $\sigma_{k,j}$. The learner's objective is to adaptively allocate queries to judges so as to estimate the true score vector $\boldsymbol{s} = (s_k)_{k \in [K]} \in \mathbb{R}^K$ while minimizing the ℓ_p-error, for a fixed $p \in [1, \infty]$ specified in advance. As we discuss later, this subsumes the recently studied budget allocation problem for a single LLM-as-a-judge in Saha et al. (2026) that only considered $J = 1$ and $p = \infty$. Furthermore, while our setting is motivated by the LLM-as-a-judge framework (Saha et al., 2026), it addresses a broader class of adaptive resource allocation problems across heterogeneous information sources under budget constraints.

Contributions. We provide a complete theoretical and algorithmic solution to our problem:

(a) Optimal allocation and an adaptive algorithm. We first analyze the inverse-variance weighted estimator (IVWE) and to derive the oracle allocation that minimizes its error rate (**Section 3**). Since this allocation depends on the unknown variances, we then address the practical unknown-variance setting by proposing EST-IVWE (Algorithm 1), an adaptive algorithm that constructs and leverages *optimistically biased* variance estimates to stabilize the empirical allocation. We prove that EST-IVWE matches the oracle IVWE rate up to lower-order terms in the budget (**Section 4**).

(b) Local minimax lower bounds. We derive *local* minimax lower bounds matching the leading allocation-dependent rate of EST-IVWE, hence establishing its asymptotic instance-optimality up to logarithmic factors (**Section 5**). A key technical insight is that our Fano-type change-of-measure and Assouad-type arguments capture different allocation geometries. The former characterizes the $\log \frac{1}{\delta}$ dependence, while the latter, based on local, variance-preserving perturbations, yields the sharp oracle allocation-dependent rate both in expectation and with constant probability.

(c) Empirical validation. We numerically validate EST-IVWE on both synthetic data and the real-world *HelpSteer2* dataset (Wang et al., 2024) processed by Saha et al. (2026). Our results demonstrate that EST-IVWE consistently and significantly outperforms naïve uniform allocation (**Section 7**).

Related Work. While we contextualize our contributions with specific literature throughout the text, we briefly situate our structural formulation within broader statistical frameworks. In classical statistics, estimating a shared mean from multiple heteroskedastic sources is known as the *common mean problem* (Cochran, 1937; Graybill and Deal, 1959; Norwood Jr and Hinkelmann, 1977). This line of work explores how to optimally aggregate samples, though traditionally without budget constraints. More recently, theoretical computer science has studied a variant featuring a single sample per population – referred to as *entangled mean estimation* (Chierichetti et al., 2014; Diakonikolas et al., 2025; Liang and Yuan, 2020)—largely motivated by crowdsourcing applications. Despite this rich literature on heteroskedastic aggregation, incorporating a fixed budget constraint across sources with varying costs has, to the best of our knowledge, remained unexplored prior to our work.

2 Problem Setting

There are K queries, e.g., prompt–answer pairs, to be evaluated by J LLM judges. Each query $k \in [K]$ has an unknown ground-truth score $s_k \in \mathbb{R}$. Whenever the learner assigns query k to judge j, she incurs a cost $c_j \in \mathbb{R}_{>0}$ and observes a noisy score $s_{k,j} \in [0, R]$, independently of past observations. We assume that $\mathbb{E}[s_{k,j}] = s_k$ and $\sigma^2_{k,j} := \mathrm{Var}(s_{k,j}) > 0$. The variances $\sigma^2_{k,j}$ are unknown, while the score upper bound R is known. The learner operates online: at each step, she selects a query–judge pair based on all observations collected so far. Given a total budget $B \in \mathbb{R}_{>0}$ and a known cost vector $\boldsymbol{c} = (c_j)_{j\in[J]} \in \mathbb{R}^J_{>0}$, her goal is to provide an estimator of the score vector $\boldsymbol{s} = (s_k)_{k\in[K]} \in \mathbb{R}^K$ as accurate as possible, i.e., minimizing the ℓ_p-error, for a fixed $p \in [1, \infty]$ specified in advance. With this, each problem instance is characterized by a triple $(\boldsymbol{s}, \boldsymbol{\sigma}, \boldsymbol{c}) \in \mathbb{R}^K \times \mathbb{R}^{K\times J}_{>0} \times \mathbb{R}^J_{>0}$, where $\boldsymbol{\sigma} := (\sigma_{k,j})_{k\in[K],j\in[J]}$ is the variance profile.

To assess and compare the performance, we introduce the following notion of learnability:

Definition 2.1 ((B, δ, ℓ_p)-Budget Efficient Algorithm)**.** Let $p \in [1, \infty]$, $B \in \mathbb{R}_{>0}$ be a total budget, and $\delta \in (0, 1)$ be a confidence level. An algorithm $\mathcal{A}$ with score estimate $\hat{\boldsymbol{s}} \in \mathbb{R}^K$ is said to be (B, δ, ℓ_p)**-budget efficient** at $(\boldsymbol{s}, \boldsymbol{\sigma}, \boldsymbol{c})$ with error rate $\varepsilon_p > 0$ if:

$$\mathbb{P}\left(\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_p \le \varepsilon_p\right) \ge 1 - \delta \quad \text{and} \quad \sum_{j\in[J]} c_j \sum_{k\in[K]} N_{k,j} \le B, \text{ a.s.}$$

where $N_{k,j}$ denotes the (random) number of times $\mathcal{A}$ queries pair $(k, j) \in [K] \times [J]$.

Discussion. We view this model as a first step toward a statistical foundation for budget-constrained estimation from multiple LLM judges. It captures two key characteristics: heterogeneous per-judge costs and query-dependent reliability through the variance profile. The K queries are fixed and given; the learner chooses which query–judge pairs to evaluate and must estimate $\boldsymbol{s}$.

The shared-mean assumption idealizes judges as unbiased estimators of a common target score, which may serve as an approximation after calibration. Likewise, the deterministic cost c_j can represent a conservative per-judge budget charge under fixed input and output token limits, even when the realized cost is smaller. These assumptions simplify practical evaluation pipelines, where residual bias, dependence, and query-dependent stochastic costs can be important. We next discuss the boundedness assumption and the role of heteroskedasticity in our model; further discussion of its limitations and possible extensions is provided in Appendix I.

The boundedness assumption $s_{k,j} \in [0, R]$ is quite realistic as the LLM-judges are typically required to assign scores in fixed rating scales (e.g., binary, $\{0, 1, 2, 3, 4\}$, or $[0, 100]$) based on scoring rubrics to ensure consistent evaluation (Gu et al., 2026). As we will see later, such a boundedness assumption necessitates fundamentally different analysis techniques and algorithmic design compared to Saha et al. (2026), which assumes Gaussianity for algorithm design. Additionally, we emphasize that nonnegativity is not necessary; all statements hold for any interval of length R.

We now comment on the heteroskedasticity of the problem setting. By Popoviciu's inequality on variances (Popoviciu, 1935), a trivial bound of $\sigma^2_{k,j} \le R^2/4$ holds, but we note that it can be significantly loose; for instance, if the pair (k, j) is "easy-to-evaluate" (the judge has sufficient expertise to "confidently" evaluate the query), $\sigma^2_{k,j} \ll R^2/4$. Accounting for such variation in the variances is a central challenge in the heteroskedastic learning literature (Chaudhuri et al., 2017; Fontaine et al., 2021; Kirschner and Krause, 2018).

3 Oracle Allocation for Inverse-Variance Weighted Estimation

In this section, we first provide high-probability guarantees for the ℓ_p-error of the celebrated Inverse-Variance Weighted Estimator (IVWE) (Cochran, 1937, 1954) under any fixed allocation, and then derive the allocation minimizing this error. This allocation is referred to as the *Oracle* allocation, because it requires the knowledge of the variances $\sigma^2_{k,j}$.

Consider a fixed allocation and let $N_{k,j}$ be the number sampled scores obtained for query k from judge j. Alternatively, the allocation can be represented by the cost-weighted proportions of the budget allocated to the (k,j)-pairs by $\boldsymbol{\omega} := (\omega_{k,j} \triangleq \frac{c_j N_{k,j}}{B})_{k,j}$. The budget constraint B is satisfied if and only if $\boldsymbol{\omega} \in \Lambda^{KJ} := \{\boldsymbol{\lambda} \geq 0 : \sum_{k,j} \lambda_{k,j} = 1\}$. Let $\{s_{k,j}^{(i)}\}_{i=1}^{N_{k,j}}$ denote sampled scores of query k from the judge j. The IVWE is defined by:

$$\hat{s}_k := \left(\sum_{j=1}^{J} \frac{N_{k,j}}{\sigma_{k,j}^2}\right)^{-1} \sum_{j=1}^{J} \frac{N_{k,j}\hat{s}_{k,j}}{\sigma_{k,j}^2}, \quad \hat{s}_{k,j} := \frac{1}{N_{k,j}} \sum_{i=1}^{N_{k,j}} s_{k,j}^{(i)}, \ \forall j \in [J],$$

where by convention, $\hat{s}_{k,j} = 0$ if $N_{k,j} = 0$. The next proposition provides error-rate guarantees for IVWE; its full statement and proof are given in Appendix A.1.

Proposition 3.1. *For any instance $(\boldsymbol{s}, \boldsymbol{\sigma}, \boldsymbol{c})$ and $p \in [1, \infty]$,* IVWE *is (B, δ, ℓ_p)-budget efficient with the following error rate ε_p:*

$$\varepsilon_p \approx_B \sqrt{\frac{2\mathcal{A}_p(\boldsymbol{\omega}; \boldsymbol{\sigma}, \boldsymbol{c}) \log \frac{2K}{\delta}}{B}}, \qquad \mathcal{A}_p(\boldsymbol{\omega}; \boldsymbol{\sigma}, \boldsymbol{c}) := \left(\sum\nolimits_{k\in[K]} \left(\sum\nolimits_{j\in[J]} \frac{\omega_{k,j}}{c_j \sigma_{k,j}^2}\right)^{-\frac{p}{2}}\right)^{\frac{2}{p}},$$

where $\approx_B$ omits lower-order terms in B.

In the above result, $\mathcal{A}_p(\cdot)$ is derived from the variance of IVWE. We refer to it as the *allocation objective*, since minimizing this quantity yields the allocation that minimizes the estimator's error rate. The following theorem, proved in Appendix A.2, characterizes this optimal allocation.

Theorem 3.2. *For each query $k \in [K]$, let $j^*(k) := \arg\min_{j\in[J]} c_j \sigma_{k,j}^2$ be the optimal judge. For $p \in [1, \infty]$, consider the following optimal allocation problem: $\mathcal{A}_p^*(\boldsymbol{\sigma}, \boldsymbol{c}) := \min_{\boldsymbol{\omega}\in\Lambda^{KJ}} \mathcal{A}_p(\boldsymbol{\omega}; \boldsymbol{\sigma}, \boldsymbol{c})$. The value of this problem and the corresponding optimal allocation $\boldsymbol{\omega}^* = (\omega_{k,j}^*)_{k,j}$ are given by*

$$\mathcal{A}_p^*(\boldsymbol{\sigma}, \boldsymbol{c}) = \left(\sum_{k\in[K]} \left(c_{j^*(k)} \sigma_{k,j^*(k)}^2\right)^{\frac{p}{p+2}}\right)^{\frac{p+2}{p}}, \ \omega_{k,j}^* = \frac{\mathbb{1}[j = j^*(k)] \left(c_{j^*(k)} \sigma_{k,j^*(k)}^2\right)^{\frac{p}{p+2}}}{\sum_{k'\in[K]} \left(c_{j^*(k')} \sigma_{k',j^*(k')}^2\right)^{\frac{p}{p+2}}}.$$

A few remarks are in order. First, note that when $J = 1$ (a single judge), $c_1 = 1$ (unit cost), and $p = \infty$, the above result reduces exactly to Saha et al. (2026, Theorem 3). Our setting therefore generalizes the framework studied therein. An interesting consequence of the theorem is that, regardless of p, an optimal allocation can be chosen to be *sparse*: for each query k, the budget is assigned to a single optimal judge $j^*(k)$. This follows from the direct comparison in Appendix A.2: for a fixed budget assigned to query k, its contribution to the allocation objective is minimized by assigning that budget to a judge minimizing $c_j \sigma_{k,j}^2$. Intuitively, once the budget allocated to a query is fixed, the most accurate estimate $\hat{s}_k$ is obtained by assigning all of that budget to the judge with the best cost–variance trade-off. Allocating any positive budget to a judge j with $c_j \sigma_{k,j}^2 > c_{j^*(k)} \sigma_{k,j^*(k)}^2$ can only worsen, or at best leave unchanged, the resulting estimate. Thus, while the choice of p changes how budget is distributed across queries, a single optimal judge per query always suffices. If several judges attain the same minimum, splitting the query's budget among them is also optimal.

At this stage, the optimality statement concerns the minimization of our upper bound. In Section 5, we show that this bound is in fact minimax optimal.

4 EST-IVWE: An Adaptive Algorithm and its Error Rates

In this section, we consider unknown variances, and propose an adaptive algorithm EST-IVWE. It proceeds in two phases: the first phase estimates the variances, while the second allocates the remaining budget according to the empirical optimal allocation induced by these estimates. The pseudo code of EST-IVWE is given in Algorithm 1. Before providing error-rate guarantees for EST-IVWE, we first motivate its design and discuss the main challenge in its analysis.

Algorithm 1: Estimate-then-IVWE (Est-IVWE)

Input: Total budget $B > 0$, Cost vector $\boldsymbol{c} \in \mathbb{R}_+^J$, Error norm $p \in [1, \infty]$, Confidence level $\delta \in (0, 1)$, Number of forced exploration per pair $N_0 \geq 2$

`// Phase I: Forced Exploration`

for *each* $(k, j) \in [K] \times [J]$ **do**

  Pull arm (k, j) exactly N_0 times and observe noisy evaluation scores $\left\{s_{k,j}^{(n)}\right\}_{n \in [N_0]}$;

  Compute the empirical variance estimator

$$\hat{\sigma}_{k,j}^2 \quad := \quad \frac{1}{2N_0(N_0-1)} \sum_{1 \leq n \neq n' \leq N_0} \left(s_{k,j}^{(n)} - s_{k,j}^{(n')}\right)^2; \tag{1}$$

`// Phase II: IVWE`

Compute the optimal allocation $\left(\widehat{\omega}_{k,j}^*\right)_{k,j}$ as in Theorem 3.2 using the remaining budget $B' \triangleq B - N_0 K \sum_{j \in [J]} c_j - K \max_{j \in [J]} c_j$, based on $\left\{(\hat{\sigma}_{k,j} + \tau)^2\right\}_{k,j}$ with τ as in Eqn. (2);

Set $\widehat{N}_{k,j}^* = \left\lceil B' \widehat{\omega}_{k,j}^* / c_j \right\rceil$, and query each pair (k, j) exactly $\widehat{N}_{k,j}^*$ times;

return IVWE with $\left(\widehat{N}_{k,j}^*\right)_{k,j}$ as allocation and $(\tau + \hat{\sigma}_{k,j})^2$ as variance proxies;

Optimistically biased variance estimates. A key challenge in analyzing the performance of Algorithm 1 is quantifying *how accurately the variances are estimated from Phase I*. Ideally, we seek a multiplicative guarantee of the form $(1-\alpha)\sigma_{k,j}^2 \leq \hat{\sigma}_{k,j}^2 \leq (1+\alpha)\sigma_{k,j}^2$ for some $\alpha \in (0, 1)$. The common factors $(1 \pm \alpha)$ then give a direct comparison between empirical and oracle allocations (Theorem 3.2). Under Gaussian scores, Saha et al. (2026, Lemma 6) use such multiplicative bounds, obtained from χ^2 concentration (Laurent and Massart, 2000, Eqn. (4.4) & (4.5)). These bounds do not apply to general bounded scores. Instead, we utilize the empirical Bernstein inequality (Maurer and Pontil, 2009, Theorem 10), which provides an additive guarantee of $|\sigma_{k,j} - \hat{\sigma}_{k,j}| \lesssim \sqrt{\log(1/\delta)/N_0}$ (N_0 is the length of the first phase). Now, if we were to use $\hat{\sigma}_{k,j}$ directly and require it to satisfy analogous multiplicative guarantee, we would require $N_0 \gtrsim \frac{1}{\sigma_{k,j}^2} \log \frac{1}{\delta}$. This is problematic because we require the knowledge of $\sigma_{k,j}$ to set N_0, and for *easy* pairs in which $\sigma_{k,j}$ is near-zero, the requirement on the number of exploratory samples N_0 becomes prohibitively large.

To overcome these limitations, we use an *optimistic variance estimator*,

$$\bar{\sigma}_{k,j} = \hat{\sigma}_{k,j} + \tau, \qquad \text{where} \quad \tau := R\sqrt{\frac{2}{N_0 - 1} \log \frac{4KJ}{\delta}}. \tag{2}$$

Here, the bias τ comes from the empirical Bernstein inequality so that $\sigma_{k,j} \leq \bar{\sigma}_{k,j} \leq \sigma_{k,j} + 2\tau$ with high probability; see Lemma B.2. This optimistic correction stabilizes the empirical allocation, especially for query–judge pairs with near-zero variance, whose estimates are otherwise highly unstable. It also ensures that such pairs receive a sufficient amount of budget for reliable estimation. By tuning τ and N_0 appropriately as functions of the total budget B, the extra cost induced by this stabilization remains lower order relative to the leading error term.

The following theorem formalizes the resulting error guarantees.

Theorem 4.1 (Error Rates of EST-IVWE). *Let $p \in [1, \infty]$, $B \in \mathbb{R}_{>0}$, and $\delta \in (0, 1)$. Let us set*

$$N_0 = \begin{cases} \max\left\{2, \left\lceil (2B)^{1/3} \left(R^2 \log \frac{4KJ}{\delta}\right)^{2/3} \right\rceil\right\} & \text{if } p \geq 2, \\ \max\left\{2, \left\lceil 2^{\frac{p}{2p+2}} R^{\frac{3p+2}{2p+2}} \left(\log \frac{4KJ}{\delta}\right)^{\frac{3p+2}{4p+4}} B^{\frac{p+2}{4p+4}} \right\rceil\right\} & \text{if } 1 \leq p < 2. \end{cases} \tag{3}$$

Further suppose that $B \geq 2\left(N_0 K \sum_{j\in[J]} c_j + K \max_{j\in[J]} c_j\right)$. *Then,* EST-IVWE *is* (B, δ, ℓ_p)*-budget efficient at any* $(\boldsymbol{s}, \boldsymbol{\sigma}, \boldsymbol{c})$ *with the following error rate* ε_p*:*

$$\varepsilon_p = \sqrt{\frac{2\mathcal{A}_p^*(\boldsymbol{\sigma}, \boldsymbol{c}) \log \frac{4K}{\delta}}{B}} + \begin{cases} \mathcal{O}\left(\frac{R^{\frac{1}{3}} \mathcal{A}_p^*(\boldsymbol{\sigma}, \boldsymbol{c}) \left(\log \frac{4KJ}{\delta}\right)^{\frac{2}{3}}}{B^{\frac{2}{3}}}\right) & \text{if } p \geq 2, \\ \mathcal{O}\left(\frac{R^{\frac{p}{2p+2}} \mathcal{A}_p^*(\boldsymbol{\sigma}, \boldsymbol{c}) \left(\log \frac{4KJ}{\delta}\right)^{\frac{3p+2}{4p+4}}}{B^{\frac{3p+2}{4p+4}}}\right) & \text{if } 1 \leq p < 2. \end{cases}$$

Proof sketch. We divide the (k, j)-pairs into two cases using the bias τ as an implicit threshold. For "easy-to-evaluate" pairs with $\sigma_{k,j} \leq \tau$, the bias stabilizes the ratio $(\hat{\sigma}_{k,j} + \tau)/(\sigma_{k,j} + \tau)$, avoiding the numerical instability of $\hat{\sigma}_{k,j}/\sigma_{k,j}$ when $\sigma_{k,j} \approx 0$. For "hard-to-evaluate" pairs with $\sigma_{k,j} > \tau$, the multiplicative guarantee between $\hat{\sigma}_{k,j} + \tau$ and $\sigma_{k,j} + \tau$ follows, as we choose τ to be the concentration radius of (Maurer and Pontil, 2009, Theorem 10). Since the selection rule $\hat{j}^*(k) = \arg\min_j c_j \overline{\sigma}^2_{k,j}$ in Phase II implies $c_{\hat{j}^*(k)} \overline{\sigma}^2_{k,\hat{j}^*(k)} \leq c_{j^*(k)} \overline{\sigma}^2_{k,j^*(k)}$, we can upper bound the error ε_p of the empirical allocation; and then we decouple the effect of additive τ using Minkowski's inequality. We then conclude by separating the impact of initial exploration N_0 from the total budget B, to recover the *same* leading term as the error rate of IVWE. The full proof is deferred to Appendix B. □

We conclude the section with two remarks on EST-IVWE and its performance.

(1) Oracle performance without known variances. Remarkably, Theorem 4.1 shows that, for every fixed instance and all $p \in [1, \infty]$, the adaptive algorithm EST-IVWE (Algorithm 1) matches the leading $B^{-1/2}$ term of the oracle strategy that knows the variances in advance (Theorem 3.2). Thus, in the sufficiently large budget B, there is no price for not knowing the variances: a simple two-stage procedure recovers the oracle rate.

(2) Relations to Graybill-Deal Estimator. Estimating s_k from multiple sources is an instance of the classical *common mean problem* (Graybill and Deal, 1959; Pal et al., 2007). Our two-stage estimator is closely related to the *Graybill–Deal* estimator, which uses all available samples to estimate both means and variances. Reusing samples induces correlations between these estimates, complicating sharp performance analysis. Even for Gaussian populations, tight variance characterizations are technically involved (Nair, 1980; Voinov, 1984), and admissibility is known only for certain classes of problems (Ghosh and Sinha, 1981; Pal and Lim, 1997; Sinha and Mouqadem, 1982). We avoid this dependence through *sample splitting*: Phase-I samples estimate the variances $\hat{\sigma}^2_{k,j}$, while fresh Phase-II samples yield the final IVWE estimates $\hat{s}_k$. This separates the randomness used in the two phases and ensures that, conditioned on the estimated variances, IVWE remains unbiased, $\mathbb{E}[\hat{s}_k \mid \{\hat{\sigma}_{k,j}\}_{k,j}] = s_k$.

5 Local Minimax Lower Bounds

We now establish fundamental local minimax lower bounds for our budgeted heteroskedastic multi-judge estimation problem. The term *local* means that the lower bound holds in a small neighborhood of any fixed problem instance, in the same spirit as local minimax results for logistic bandits (Abeille et al., 2021, Theorem 2), online LQR (Simchowitz and Foster, 2020, Theorem 1), and generalized trace regression (Lee et al., 2026, Theorem 4.1). These results show that the instance-dependent allocation quantity $\mathcal{A}_p^*(\boldsymbol{\sigma}, \boldsymbol{c})$, initially derived from upper bounds for IVWE, is in fact the correct information-theoretic complexity measure. A key technical insight in our lower-bound analysis is that our two arguments capture complementary aspects of the problem: a Fano-type change-of-measure argument characterizes the $\log \frac{1}{\delta}$ dependence, while an Assouad-type construction recovers the oracle allocation-dependent complexity.

5.1 A Fano-Type High-Probability Lower Bound

We begin by stating a high-probability minimax lower bound that characterizes the $\log \frac{1}{\delta}$ dependence through a Fano-type change-of-measure argument:

Theorem 5.1 (High Probability Lower Bound). *Let $p \in [1,\infty]$, $B > 0$, $\delta \in (0, 1/16]$, and $(\boldsymbol{s}, \boldsymbol{\sigma}, \boldsymbol{c})$ be a given problem instance. Let $R_k := \min\{s_k, R - s_k\} > 0$, $R_{\min} = \min_k R_k$, and suppose that $\max_j \sigma_{k,j}^2 \le \frac{R_k^2}{2}$. Then, any algorithm that is (B, δ, ℓ_p)-budget efficient with an error rate of $\varepsilon_p \in \left(0, \frac{R_{\min}}{8}\right]$ at any $(\boldsymbol{s}', \boldsymbol{\sigma}, \boldsymbol{c})$ with $\|\boldsymbol{s}' - \boldsymbol{s}\|_p \le 2\varepsilon_p$[a] must satisfy the following: for an absolute constant $C_1 > 0$,*

$$\varepsilon_p \ge \sqrt{\frac{\log\frac{1}{\delta}}{C_1 B}\mathcal{A}^*_{\frac{2p}{2-p}}(\boldsymbol{\sigma}, \boldsymbol{c})} \quad \text{for } 1 \le p < 2, \qquad \varepsilon_p \ge \sqrt{\frac{\log\frac{1}{\delta}}{C_1 B}\mathcal{A}^*_{\infty}(\boldsymbol{\sigma}, \boldsymbol{c})} \quad \text{for } p \ge 2.$$

[a] This in the same spirit as locally stable algorithms used for proving local & high-probability lower bounds (Jedra and Proutière, 2023; Yun and Proutière, 2019).

Proof sketch. We suppose that the learner has knowledge of the true $\boldsymbol{\sigma}$. The proof uses a Fano-type change-of-measure argument, following the hypothesis testing framework (Jedra and Proutière, 2023) based on the binary KL data processing inequality (Garivier et al., 2019, Lemma 1). The main technical novelty is the construction of the prior over $\boldsymbol{s}$ to retain the instance-specific nature. Specifically, when constructing alternate models, one needs to ensure that (i) its mean stays close to $\boldsymbol{s}$, (ii) its variance remains the *same* as $\boldsymbol{\sigma}$, and (iii) it is bounded in $[0, R]$. This is done by considering a family of *beta distributions* (Proposition C.1). The proof "bifurcates" at $p = 2$: the function $z \mapsto z^{2/p}$ is convex for $p \le 2$ and concave for $p \ge 2$, and is linear at $p = 2$. Consequently, minimizing the quadratic KL bound at a prescribed ℓ_p distance admits a sparse solution for $p \ge 2$ and a dense solution for $1 \le p < 2$ (Lemma C.2). We provide the full proof in Appendix C. □

The high-probability bound matches the upper bound only in the case $p = \infty$, up to a factor of $\sqrt{\log K}$. For $p < \infty$, however, it does not recover the allocation quantity $\mathcal{A}^*_p(\boldsymbol{\sigma}, \boldsymbol{c})$ appearing in the upper bound of IVWE. Indeed, the lower bound only attains $\mathcal{A}^*_{\frac{2p}{2-p}}(\boldsymbol{\sigma}, \boldsymbol{c})$ for $1 \le p < 2$ and $\mathcal{A}^*_\infty(\boldsymbol{\sigma}, \boldsymbol{c})$ for $p \ge 2$. This is *not* the oracle allocation dependence for finite p. Its explicit $\log\frac{1}{\delta}$ dependence nevertheless remains informative, as we discuss in Section 5.3.

5.2 An Assouad-Type In-Expectation Lower Bound

We now recover the allocation dependence missing from the previous Fano-type construction. Using an Assouad-type argument, we obtain a local in-expectation lower bound that matches $\mathcal{A}^*_p(\boldsymbol{\sigma}, \boldsymbol{c})$ for all $p \in [1,\infty]$. The same construction also yields a constant-probability lower bound, as noted below.

Theorem 5.2 (In-Expectation Lower Bound). *Let $p \in [1,\infty]$, $B > 0$ and $(\boldsymbol{s}^\star, \boldsymbol{\sigma}, \boldsymbol{c})$ be any given problem instance with $\boldsymbol{s}^\star = (s_k^\star)_k$. Suppose that $R_k := \min\{s_k^\star, R - s_k^\star\} > 0$ and $0 < \sigma_{k,j}^2 \le \frac{R_k^2}{2}$ for all (k, j). Define the "radius" $\xi(B) := \sqrt{\frac{\mathcal{A}^*_p(\boldsymbol{\sigma}, \boldsymbol{c})}{16B}}$ and the "threshold"[a] as $B_0 := \left(\sum_{k\in[K]}\left(c_{j^*(k)}\sigma^2_{k,j^*(k)}\right)^{\frac{p}{p+2}}\right)\max_{k\in[K]}\frac{\left(c_{j^*(k)}\sigma^2_{k,j^*(k)}\right)^{\frac{2}{p+2}}}{R_k^2}$. Then, there exists a universal constant $C_3 > 0$ such that for $B \ge B_0$,*

$$\inf_{\hat{\boldsymbol{s}}} \sup_{\boldsymbol{s}:\|\boldsymbol{s}-\boldsymbol{s}^\star\|_p \le \xi(B)} \mathbb{E}^{\hat{\boldsymbol{s}}}_{\boldsymbol{s}}\left[\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_p\right] \ge C_3\sqrt{\frac{\mathcal{A}^*_p(\boldsymbol{\sigma}, \boldsymbol{c})}{B}},$$

and furthermore for $1 \le p < \infty$ and $B \ge pB_0$,

$$\inf_{\hat{\boldsymbol{s}}} \sup_{\boldsymbol{s}:\|\boldsymbol{s}-\boldsymbol{s}^\star\|_p \le \sqrt{p}\xi(B)} \mathbb{E}^{\hat{\boldsymbol{s}}}_{\boldsymbol{s}}\left[\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_p^p\right]^{\frac{1}{p}} \ge 4^{1-\frac{1}{p}} C_3\sqrt{\frac{p\mathcal{A}^*_p(\boldsymbol{\sigma}, \boldsymbol{c})}{B}}.$$

[a] For $p = \infty$, the expressions are interpreted by continuity, giving $B_0 = \mathcal{A}^*_\infty(\boldsymbol{\sigma}, \boldsymbol{c}) \max_k R_k^{-2}$.

Proof sketch. The full proof is deferred to Appendix D. We follow the standard nonparametric lower bound framework based on the seminal Assouad's Lemma (Assouad, 1983; Yu, 1997) combined with the Bretagnolle-Huber inequality (Bretagnolle and Huber, 1979). We first construct a hypercube of 2^K hard instances, and each vertex corresponding to the same beta distribution construction as in the proof of Theorem 5.1 (Proposition C.1).

For $p > 1$, $\mathbb{E}[\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_p]$ is *not* decomposable coordinate-wise, hindering the direct application of Assouad's lemma. For finite p, we bypass this using a secant bounding argument (leveraging the concavity of $x \mapsto x^{1/p}$), decoupling the expectation to establish a lower bound proportional to $\|\boldsymbol{\Delta}\|_p$. For $p = \infty$, we use equal perturbations and lower bound the probability of at least one incorrect sign. For finite p, the p-th moment is coordinate-wise decomposable, so we can directly apply Assouad's Lemma. Finally, by tuning the adversarial shifts $\boldsymbol{\Delta}$ under KL-divergence constraints and minimizing the bound over all valid algorithm budget allocations, we recover the optimal rate $\mathcal{A}_p^*(\boldsymbol{\sigma}, \boldsymbol{c})$. □

Beyond In-Expectation. For finite p, the same Assouad construction also gives a constant-probability lower bound via the reverse Markov inequality; see Ma et al. (2026) for a general conversion from local minimax risk to minimax quantiles. (All proofs here are deferred to Appendix D.) Under the assumptions of Theorem 5.2, with the same C_3, for every $\delta \in (0, 4C_3)$,

$$\inf_{\hat{\boldsymbol{s}}} \sup_{\boldsymbol{s}:\|\boldsymbol{s}-\boldsymbol{s}^\star\|_p \le \xi(B)} \mathbb{P}_{\boldsymbol{s}}^{\hat{\boldsymbol{s}}}\left(\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_p \ge \xi(B) \left(\frac{4C_3 - \delta}{1 - \delta} \right)^{1/p} \right) \ge \delta.$$

Taking $\delta = 2C_3$ recovers the allocation-dependent rate with constant probability, but the threshold stays bounded as $\delta \downarrow 0$: this conversion does not recover the $\log \frac{1}{\delta}$ dependence of Theorem 5.1. For $p = \infty$, the construction directly gives $\mathbb{P}_{\boldsymbol{s}}^{\hat{\boldsymbol{s}}}(\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_\infty \ge \xi(B)) \ge 4C_3$ at some vertex in the same local neighborhood. Both arguments are given in Appendix D.

We make two further remarks concerning our lower bounds:

(1) Geometry of Constrained Lower Bounds. Optimizing over algorithms subject to a budget constraint is reminiscent of techniques used in statistical estimation under communication constraints (Chen et al., 2021; Chen and Özgür, 2024; Zhu and Lafferty, 2014). In such settings, restricting the algorithm class naturally gives rise to an optimal allocation problem. Our analysis reveals a related phenomenon, with different allocation quantities governing the bounds obtained from our Fano/Assouad-type arguments. We discuss their complementary roles in the next subsection.

(2) In-Expectation Upper Bounds. The constant-probability consequence above permits comparison with the high-probability upper bounds in Theorems 3.2 and 4.1 at sufficiently small fixed δ. For known variances and fixed $1 \le p < \infty$, a standard moment argument based on Bernstein's inequality (Bernstein, 1927) yields an oracle expected-error bound with leading rate $\sqrt{\mathcal{A}_p^*(\boldsymbol{\sigma}, \boldsymbol{c})/B}$, without logarithmic factors and up to constants depending on p. In particular, at the oracle allocation, independence gives $\mathbb{E}[\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_2^2] = \mathcal{A}_2^*(\boldsymbol{\sigma}, \boldsymbol{c})/B$; Gaussian scores further permit exact finite-p moment calculations (Theorem F.7). For $p = \infty$, $\log K$ dependencies may remain.

5.3 Different Allocation Geometries for Assouad and Fano

Our two lower-bound arguments capture complementary aspects of the estimation problem. The Fano-type change-of-measure argument in Theorem 5.1 yields an explicit $\log \frac{1}{\delta}$ dependence, but its allocation quantity differs from $\mathcal{A}_p^*$. The Assouad-type construction in Theorem 5.2 recovers $\mathcal{A}_p^*$ both in expectation and with constant probability, but lacks the $\log \frac{1}{\delta}$ factor.

The difference comes from how the two constructions aggregate errors across queries. Our Fano-type change-of-measure argument searches for an alternative with small KL divergence at a prescribed ℓ_p distance. The resulting quadratic optimization admits a dense solution for $1 \le p < 2$ and a sparse solution for $p \ge 2$ (Lemma C.2), leading to $\mathcal{A}_{2p/(2-p)}^*$ and $\mathcal{A}_\infty^*$, respectively. The Assouad-type argument instead *first* lower bounds the weighted sign errors coordinate-wise, *then* aggregates them, and *finally* optimizes over the algorithm's budget allocation. Its hypercube of perturbations retains the simultaneous contribution of all queries. Both constructions preserve the same variance profile; the distinction concerns the aggregation of coordinate errors.

Then one may wonder if the correct lower bound should be of the form $\sqrt{\mathcal{A}_p^* \log(1/\delta)/B}$. We argue that this may not be the case. Consider the Gaussian setting with one unit-cost judge, variance σ^2, and $p = 2$. With B/K samples per query, assumed integer for simplicity, the sample mean satisfies $\mathbb{P}\left(\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_2 \le \frac{\sigma}{\sqrt{B}}\left(K + \sqrt{2K \log \frac{1}{\delta}}\right)\right) \ge 1 - \delta$; see Eqn. (8) in Appendix F. Here, $\mathcal{A}_2^* = \sigma^2 K^2$ and $\mathcal{A}_\infty^* = \sigma^2 K$: the first term reflects the error accumulated across all queries, while the second reflects the additional cost of demanding higher confidence. Thus, the two terms naturally involve different allocation quantities. In particular, a lower bound of order $\sqrt{\mathcal{A}_2^* \log(1/\delta)/B}$ would not hold uniformly in K and δ even in this simple Gaussian example.

A similar decomposition is established by Ma et al. (2026, Proposition 10) for Gaussian mean estimation: the minimax squared-error quantile has a trace term measuring the overall estimation difficulty and an operator-norm term multiplied by $\log \frac{1}{\delta}$. Their proof combines an Assouad risk bound converted to a probability statement with a separate two-point confidence argument. This supports viewing our two lower bounds as complementary, but we remark that a sharp characterization jointly in the variance profile, p, and δ remains open in our budgeted setting.

6 Beyond Boundedness: Gaussian Scores

We briefly describe how the results presented so far extend to *Gaussian scores*, as considered in (Saha et al., 2026), namely $s_{k,j} \sim \mathcal{N}(s_k, \sigma_{k,j}^2)$. In this setting, all results carry over, and in some cases take a sharper form. Technically, the Gaussian case is of independent interest because it allows the use of tools that exploit distribution-specific structure. For the interest of space, we only provide an overview, highlighting the key technical ingredients. Full details are deferred to Appendix F.

Upper Bounds. For known variances, we replace Bernstein's inequality with the standard Gaussian tail bound applied separately for each query k. This yields a cleaner upper bound, with no lower-order term from Bernstein's inequality. For unknown variances, we can exploit χ^2-concentration for the Gaussian sample variance (Laurent and Massart, 2000, Eq. (4.4)). This allows us to use the IVWE directly, without introducing the additional optimistic bias τ.

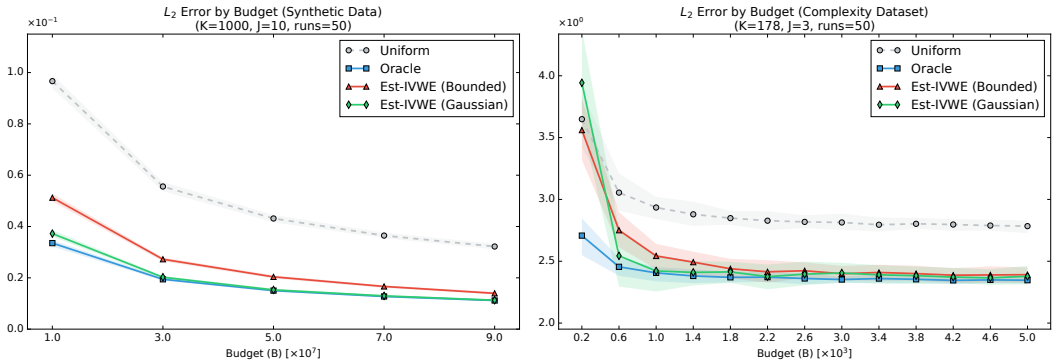
Lower Bounds. Here, we consider global rather than local lower bounds. The high-probability lower bound is essentially unchanged, except that we use Gaussian constructions instead of Beta distributions, which affects only constants. The in-expectation lower bounds, for both $\mathbb{E}[\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_p]$ and $\mathbb{E}[\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_p^p]^{1/p}$, are substantially different and technically more interesting, which we sketch here. We place a Gaussian prior over the instances and use a convexity argument to show that the posterior mean minimizes the expected ℓ_p-risk. Since the posterior distribution is Gaussian with inverse Fisher information as the covariance, rearranging yields the same optimal allocation in the lower bound. Remarkably, for $\mathbb{E}[\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_p^p]^{1/p}$, this lower bound is optimal up to universal constants.

7 Experiments

We now numerically evaluate the performance of our algorithms on synthetic and real-world datasets. The codes are available in our GitHub repository.

Settings. For the synthetic experiment, we set $(K, J) = (1000, 10)$ and consider the Beta distribution for each query–judge pair's score distribution ($R = 1$). The instance, characterized by $(\boldsymbol{s}, \boldsymbol{\sigma}, \boldsymbol{c})$, is randomly sampled from an appropriate prior distribution, detailed in Appendix H.1.

We conduct real-world experiments following Saha et al. (2026), in which queries from HelpSteer2 (Wang et al., 2024) are evaluated by three LLM judges — Llama-3-8B (Llama Team, 2024), GPT-4 (OpenAI Team, 2024), and Qwen (Qwen Team, 2024) — along four axes: complexity, correctness, helpfulness, and verbosity. For each axis, we utilize the pre-generated dataset from Saha et al. (2026), where each query is repeatedly evaluated with each judge. During the experiments, we simulate each judge by sampling with replacement from the generated dataset. We set the judge costs uniformly to 0.1 for simplicity, and we consider $p = 2$ for the error metric. Further details of the preprocessing procedure are provided in Appendix H.2.



(a) Synthetic data ($K = 1000$, $J = 10$) (b) "Complexity" dataset ($K = 178$, $J = 3$)

Figure 1: Experimental results on both synthetic and real-world datasets. All results are averaged over 50 independent runs, with the 10% and 90% quantiles (shaded regions)

Baselines. We compare EST-IVWE (Bounded) (Algorithm 1) and its practical Gaussian variant, EST-IVWE (Gaussian) (Algorithm 2), against two baselines: UNIFORM, which assigns each query–judge pair an equal number of times subject to the budget constraint, and ORACLE, which is IVWE with the optimal allocation $\boldsymbol{\omega}^*$ under known variances (Theorem 3.2). In implementing EST-IVWE (Algorithm 1) and EST-IVWE (Gaussian) (Algorithm 2), the exploration budget N_0 and the bias τ are set to the theoretical values as in Theorems 4.1 and F.4, rather than being heuristically tuned.

Results and Discussions. The results for synthetic and HelpSteer2 datasets are shown in Figure 1. First, EST-IVWE and ORACLE both outperform the UNIFORM baseline, confirming the efficacy of adaptive allocation under the heteroskedasticity of the query–judge pairs. Amongst the adaptive baselines, we note that EST-IVWE rapidly approaches the performance of the ORACLE as the budget increases, validating our theoretical results. Notably, in both settings, EST-IVWE (Gaussian) outperforms or performs on par with EST-IVWE (Bounded), despite the fact that the score distributions are strictly bounded. This shows that the Gaussian variant may be more effective in practice, although it lacks a rigorous theoretical guarantee. On HelpSteer2, the error curves appear to plateau at non-zero values, possibly reflecting differences between the mean judge scores and the consensus scores s_k. Additional results from different p and different datasets are provided in Appendices H.1 and H.2. Furthermore, in Appendix H.3, we provide an additional cost-sensitivity analysis on synthetic data to demonstrate that our algorithm remains robust to different choices of the cost vector.

8 Conclusion and Future Work

This paper introduces the problem of *budgeted heteroskedastic multi-judge estimation*, motivated by budget-constrained evaluation via multiple LLM judges of varying costs. Our main algorithmic template is the inverse-variance weighted estimator (IVWE), whose error rate we analyze to derive the optimal oracle allocation. Building on this, we develop Est-IVWE for unknown variances via a two-stage approach: Stage I, the key ingredient, computes *optimistic* variance estimates, and Stage II applies IVWE with those estimates. We establish the asymptotic *instance-wise* optimality of our algorithm up to logarithmic factors by proving a local minimax lower bound matching its leading allocation-dependent rate. Interestingly, our Fano-type change-of-measure and Assouad-type arguments capture different allocation geometries: the former captures the $\log \frac{1}{\delta}$ dependence, while the latter recovers the sharp allocation-dependent quantity; this distinction may be of independent interest. We numerically validate IVWE and Est-IVWE on synthetic and HelpSteer2 datasets, demonstrating improvements over naïve uniform allocation.

Our work's practical applicability is limited by the modeling assumptions introduced in Section 2. In Appendix I, we clarify their scope and practical justification and discuss concrete directions for future work. Two further directions are worth highlighting. On the theoretical side, as discussed in Section 5, sharply characterizing the joint dependence on the variance profile, p, and $\log \frac{1}{\delta}$ remains open. Empirically, extending to production-scale LLM-as-a-judge pipelines is an important next step.

Acknowledgments and Disclosure of Funding

We thank the authors of Saha et al. (2026) for providing their code and dataset for our experiments. We also thank B. Kveton for helpful discussions and comments. Last but not least, we thank the anonymous reviewers for helpful comments and suggestions that helped us improve the manuscript. Junghyun Lee, Sanghwa Kim and Se-Young Yun are partly supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (No. RS-2019-II190075 Artificial Intelligence Graduate School Support Program (KAIST), 5%; No. RS-2024-00457882, National AI Research Lab Project, 5%; No. RS-2022-II220311 Development of Goal-Oriented Reinforcement Learning Techniques for Contact-Rich Robotic Manipulation of Everyday Objects, 90%). Alexandre Proutière is supported by Vetenskaprådet, Digital Futures, and the Wallenberg AI, Autonomous Systems and Software program. The authors declare no competing interests.

References

Marc Abeille, Louis Faury, and Clément Calauzènes. Instance-Wise Minimax-Optimal Algorithms for Logistic Bandits. In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pages 3691–3699. PMLR, 13–15 Apr 2021. URL https://proceedings.mlr.press/v130/abeille21a.html.

Shun-ichi Amari. *Information Geometry and Its Applications*. Applied Mathematical Sciences. Springer Tokyo, 2016.

Theodore W. Anderson. The Integral of a Symmetric Unimodal Function over a Symmetric Convex Set and Some Probability Inequalities. *Proceedings of the American Mathematical Society*, 6(2): 170–176, 1955. doi:10.1090/S0002-9939-1955-0069229-1.

Patrice Assouad. Deux remarques sur l'estimation. *Comptes rendus des séances de l'Académie des sciences. Série 1, Mathématique*, 1983.

Yuntao Bai, Saurav Kadavath, Sandipan Kundu, Amanda Askell, Jackson Kernion, Andy Jones, Anna Chen, Anna Goldie, Azalia Mirhoseini, Cameron McKinnon, Carol Chen, Catherine Olsson, Christopher Olah, Danny Hernandez, Dawn Drain, Deep Ganguli, Dustin Li, Eli Tran-Johnson, Ethan Perez, Jamie Kerr, Jared Mueller, Jeffrey Ladish, Joshua Landau, Kamal Ndousse, Kamile Lukosuite, Liane Lovitt, Michael Sellitto, Nelson Elhage, Nicholas Schiefer, Noemi Mercado, Nova DasSarma, Robert Lasenby, Robin Larson, Sam Ringer, Scott Johnston, Shauna Kravec, Sheer El Showk, Stanislav Fort, Tamera Lanham, Timothy Telleen-Lawton, Tom Conerly, Tom Henighan, Tristan Hume, Samuel R. Bowman, Zac Hatfield-Dodds, Ben Mann, Dario Amodei, Nicholas Joseph, Sam McCandlish, Tom Brown, and Jared Kaplan. Constitutional AI: Harmlessness from AI Feedback. *arXiv preprint arXiv:2212.08073*, 2022. URL https://arxiv.org/abs/2212.08073.

Krishna Balasubramanian, Aleksandr Podkopaev, and Shiva Kasiviswanathan. Dependence-Aware Label Aggregation for LLM-as-a-Judge via Ising Models. In *Proceedings of the 43rd International Conference on Machine Learning*, Proceedings of Machine Learning Research. PMLR, 06–11 Jul 2026. URL https://openreview.net/forum?id=VCU3hGm0O6.

Santiago R. Balseiro, Rachitesh Kumar, Vahab Mirrokni, Balasubramanian Sivan, and Di Wang. Robust Budget Pacing with a Single Sample. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 1636–1659. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/balseiro23a.html.

Heinz Bauer. Minimalstellen von Funktionen und Extremalpunkte. *Archiv der Mathematik*, 9(4): 389–393, 1958. doi:10.1007/BF01898615.

Sergei N. Bernstein. Sur l'extension du théoréme limite du calcul des probabilités aux sommes de quantités dépendantes. *Mathematische Annalen*, 97(1):1–59, Dec 1927. doi:10.1007/BF01447859.

Aleksandr Alekseevich Borovkov. *Mathematical Statistics*. Gordon and Breach Science Publishers, Amsterdam, 1998. Translated from the Russian by A. Moullagaliev.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

J. Bretagnolle and C. Huber. Estimation des densités: risque minimax. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 47(2):119–137, 1979. doi:10.1007/BF00535278.

Kamalika Chaudhuri, Prateek Jain, and Nagarajan Natarajan. Active Heteroscedastic Regression. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 694–702. PMLR, 06–11 Aug 2017. URL https://proceedings.mlr.press/v70/chaudhuri17a.html.

Wei-Ning Chen, Peter Kairouz, and Ayfer Özgür. Pointwise Bounds for Distribution Estimation under Communication Constraints. In *Advances in Neural Information Processing Systems*, volume 34, pages 24593–24603. Curran Associates, Inc., 2021. URL https://openreview.net/forum?id=l41jc6kUfKr.

Wei-Ning Chen and Ayfer Özgür. Lq Lower Bounds on Distributed Estimation via Fisher Information. In *2024 IEEE International Symposium on Information Theory (ISIT)*, pages 91–96, 2024. doi:10.1109/ISIT57864.2024.10619530.

Flavio Chierichetti, Anirban Dasgupta, Ravi Kumar, and Silvio Lattanzi. Learning Entangled Single-Sample Gaussians. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '14, page 511–522, USA, 2014. Society for Industrial and Applied Mathematics. doi:10.1137/1.9781611973402.38.

B. S. Cirel'son, I. A. Ibragimov, and V. N. Sudakov. Norms of Gaussian sample functions. In Gisiro Maruyama and Jurii V. Prokhorov, editors, *Proceedings of the Third Japan — USSR Symposium on Probability Theory*, pages 20–41, Berlin, Heidelberg, 1976. Springer Berlin Heidelberg. doi:10.1007/BFb0077482.

Stephan Clémençon and Pierre Laforgue. Statistical learning from biased training samples. *Electronic Journal of Statistics*, 16(2):6086–6134, 2022. doi:10.1214/22-EJS2084.

Eugenio Clerico, Hamish Flynn, Wojciech Kotłowski, and Gergely Neu. Confidence Sequences for Generalized Linear Models via Regret Analysis. *arXiv preprint arXiv:2504.16555*, 2025. URL https://arxiv.org/abs/2504.16555.

William G. Cochran. Problems arising in the analysis of a series of similar experiments. *Supplement to the Journal of the Royal Statistical Society*, 4(1):102–118, 1 1937. doi:10.2307/2984123.

William G. Cochran. The combination of estimates from different experiments. *Biometrics*, 10(1): 101–129, 3 1954. doi:10.2307/3001666.

Ilias Diakonikolas, Daniel M. Kane, Sihan Liu, and Thanasis Pittas. Entangled Mean Estimation in High Dimensions. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, STOC '25, page 1680–1688, New York, NY, USA, 2025. Association for Computing Machinery. doi:10.1145/3717823.3718162.

Yann Dubois, Percy Liang, and Tatsunori Hashimoto. Length-Controlled AlpacaEval: A Simple Debiasing of Automatic Evaluators. In *First Conference on Language Modeling*, 2024. URL https://openreview.net/forum?id=CybBmzWBX0.

S. Yu. Efroimovich. Information Contained in a Sequence of Observations. *Problems of Information Transmission*, 15(3):178–189, 1979. Translated from *Problemy Peredachi Informatsii*, 15(3): 24–39.

Nigel Fernandez, Branislav Kveton, Ryan A. Rossi, Andrew Lan, and Zichao Wang. RADAR: Reasoning-Ability and Difficulty-Aware Routing for Reasoning LLMs. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=CB6Ds5T4ae.

Xavier Fontaine, Pierre Perrault, Michal Valko, and Vianney Perchet. Online A-Optimal Design and Active Linear Regression. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 3374–3383. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/fontaine21a.html.

Aurélien Garivier, Pierre Ménard, and Gilles Stoltz. Explore First, Exploit Next: The True Shape of Regret in Bandit Problems. *Mathematics of Operations Research*, 44(2):377–399, 2019. doi:10.1287/moor.2017.0928.

Sébastien Gerchinovitz, Pierre Ménard, and Gilles Stoltz. Fano's Inequality for Random Variables. *Statistical Science*, 35(2):178 – 201, 2020. doi:10.1214/19-STS716.

J. K. Ghosh and Bimal K. Sinha. A Necessary and Sufficient Condition for Second Order Admissibility with Applications to Berkson's Bioassay Problem. *The Annals of Statistics*, 9(6):1334 – 1338, 1981. doi:10.1214/aos/1176345650.

Richard D. Gill and Boris Y. Levit. Applications of the van Trees inequality: a Bayesian Cramér-Rao bound. *Bernoulli*, 1(1-2):59 – 79, 1995. doi:10.2307/3318681.

Franklin A. Graybill and R. B. Deal. Combining Unbiased Estimators. *Biometrics*, 15(4):543–550, 1959. doi:10.2307/2527652.

Jiawei Gu, Xuhui Jiang, Zhichao Shi, Hexiang Tan, Xuehao Zhai, Chengjin Xu, Wei Li, Yinghan Shen, Shengjie Ma, Honghao Liu, Saizhuo Wang, Kun Zhang, Zhouchi Lin, Bowen Zhang, Lionel Ni, Wen Gao, Yuanzhuo Wang, and Jian Guo. A survey on LLM-as-a-judge. *The Innovation*, page 101253, 2026. doi:10.1016/j.xinn.2025.101253.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Peiyi Wang, Qihao Zhu, Runxin Xu, Ruoyu Zhang, Shirong Ma, Xiao Bi, Xiaokang Zhang, Xingkai Yu, Yu Wu, Z. F. Wu, Zhibin Gou, Zhihong Shao, Zhuoshu Li, Ziyi Gao, Aixin Liu, Bing Xue, Bingxuan Wang, Bochao Wu, Bei Feng, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chong Ruan, Damai Dai, Deli Chen, Dongjie Ji, Erhang Li, Fangyun Lin, Fucong Dai, Fuli Luo, Guangbo Hao, Guanting Chen, Guowei Li, H. Zhang, Hanwei Xu, Honghui Ding, Huazuo Gao, Hui Qu, Hui Li, Jianzhong Guo, Jiashi Li, Jingchang Chen, Jingyang Yuan, Jinhao Tu, Junjie Qiu, Junlong Li, J. L. Cai, Jiaqi Ni, Jian Liang, Jin Chen, Kai Dong, Kai Hu, Kaichao You, Kaige Gao, Kang Guan, Kexin Huang, Kuai Yu, Lean Wang, Lecong Zhang, Liang Zhao, Litong Wang, Liyue Zhang, Lei Xu, Leyi Xia, Mingchuan Zhang, Minghua Zhang, Minghui Tang, Mingxu Zhou, Meng Li, Miaojun Wang, Mingming Li, Ning Tian, Panpan Huang, Peng Zhang, Qiancheng Wang, Qinyu Chen, Qiushi Du, Ruiqi Ge, Ruisong Zhang, Ruizhe Pan, Runji Wang, R. J. Chen, R. L. Jin, Ruyi Chen, Shanghao Lu, Shangyan Zhou, Shanhuang Chen, Shengfeng Ye, Shiyu Wang, Shuiping Yu, Shunfeng Zhou, Shuting Pan, S. S. Li, Shuang Zhou, Shaoqing Wu, Tao Yun, Tian Pei, Tianyu Sun, T. Wang, Wangding Zeng, Wen Liu, Wenfeng Liang, Wenjun Gao, Wenqin Yu, Wentao Zhang, W. L. Xiao, Wei An, Xiaodong Liu, Xiaohan Wang, Xiaokang Chen, Xiaotao Nie, Xin Cheng, Xin Liu, Xin Xie, Xingchao Liu, Xinyu Yang, Xinyuan Li, Xuecheng Su, Xuheng Lin, X. Q. Li, Xiangyue Jin, Xiaojin Shen, Xiaosha Chen, Xiaowen Sun, Xiaoxiang Wang, Xinnan Song, Xinyi Zhou, Xianzu Wang, Xinxia Shan, Y. K. Li, Y. Q. Wang, Y. X. Wei, Yang Zhang, Yanhong Xu, Yao Li, Yao Zhao, Yaofeng Sun, Yaohui Wang, Yi Yu, Yichao Zhang, Yifan Shi, Yiliang Xiong, Ying He, Yishi Piao, Yisong Wang, Yixuan Tan, Yiyang Ma, Yiyuan Liu, Yongqiang Guo, Yuan Ou, Yuduan Wang, Yue Gong, Yuheng Zou, Yujia He, Yunfan Xiong, Yuxiang Luo, Yuxiang You, Yuxuan Liu, Yuyang Zhou, Y. X. Zhu, Yanping Huang, Yaohui Li, Yi Zheng, Yuchen Zhu, Yunxian Ma, Ying Tang, Yukun Zha, Yuting Yan, Z. Z. Ren, Zehui Ren, Zhangli Sha, Zhe Fu, Zhean Xu, Zhenda Xie, Zhengyan Zhang, Zhewen Hao, Zhicheng Ma, Zhigang Yan, Zhiyu Wu, Zihui Gu, Zijia Zhu, Zijun Liu, Zilin Li, Ziwei Xie, Ziyang Song, Zizheng Pan, Zhen Huang, Zhipeng Xu, Zhongyu Zhang, and Zhen Zhang. DeepSeek-R1 incentivizes reasoning in LLMs through reinforcement learning. *Nature*, 645(8081):633–638, 2025. doi:10.1038/s41586-025-09422-z.

Yassir Jedra and Alexandre Proutière. Finite-Time Identification of Linear Systems: Fundamental Limits and Optimal Algorithms. *IEEE Transactions on Automatic Control*, 68(5):2805–2820, 2023. doi:10.1109/TAC.2022.3221705.

Julian Katz-Samuels, Lalit Jain, Zohar Karnin, and Kevin Jamieson. An Empirical Process Approach to the Union Bound: Practical Algorithms for Combinatorial and Linear Bandits. In *Advances in Neural Information Processing Systems*, volume 33, pages 10371–10382. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/hash/75800f73fa80f935216b8cfbedf77bfa-Abstract.html.

Seungone Kim, Juyoung Suk, Shayne Longpre, Bill Yuchen Lin, Jamin Shin, Sean Welleck, Graham Neubig, Moontae Lee, Kyungjae Lee, and Minjoon Seo. Prometheus 2: An Open Source Language Model Specialized in Evaluating Other Language Models. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 4334–4353, Miami, Florida, USA, November 2024. Association for Computational Linguistics. doi:10.18653/v1/2024.emnlp-main.248.

Johannes Kirschner and Andreas Krause. Information Directed Sampling and Bandits with Heteroscedastic Noise. In *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pages 358–384. PMLR, 06–09 Jul 2018. URL https://proceedings.mlr.press/v75/kirschner18a.html.

Tor Lattimore and Csaba Szepesvári. *Bandit Algorithms*. Cambridge University Press, 2020.

B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, 28(5):1302 – 1338, 2000. doi:10.1214/aos/1015957395.

Lucien Le Cam. Convergence of Estimates Under Dimensionality Restrictions. *The Annals of Statistics*, 1(1):38 – 53, 1973. doi:10.1214/aos/1193342380.

Junghyun Lee, Se-Young Yun, and Kwang-Sung Jun. Improved Regret Bounds of (Multinomial) Logistic Bandits via Regret-to-Confidence-Set Conversion. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pages 4474–4482. PMLR, 02–04 May 2024. URL https://proceedings.mlr.press/v238/lee24d.html.

Junghyun Lee, Kyoungseok Jang, Milan Vojnović, Kwang-Sung Jun, and Se-Young Yun. GL-LowPopArt: A Nearly Instance-Wise Minimax-Optimal Estimator for Generalized Low-Rank Trace Regression. In *Proceedings of The 29th International Conference on Artificial Intelligence and Statistics*, Proceedings of Machine Learning Research. PMLR, 02–05 May 2026. URL https://openreview.net/forum?id=mMxiIP3Gxu.

Haitao Li, Qian Dong, Junjie Chen, Huixue Su, Yujia Zhou, Qingyao Ai, Ziyi Ye, and Yiqun Liu. LLMs-as-Judges: A Comprehensive Survey on LLM-based Evaluation Methods. *arXiv preprint arXiv:2412.05579*, 2024. URL https://arxiv.org/abs/2412.05579.

Yingyu Liang and Hui Yuan. Learning Entangled Single-Sample Gaussians in the Subset-of-Signals Model. In *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pages 2712–2737. PMLR, 09–12 Jul 2020. URL https://proceedings.mlr.press/v125/liang20b.html.

Llama Team. The Llama 3 Herd of Models. *arXiv preprint arXiv:2407.21783*, 2024. URL https://arxiv.org/abs/2407.21783.

Tianyi Ma, Kabir A. Verchand, and Richard J. Samworth. High-Probability Minimax Lower Bounds. *Statistical Science*, 41(3):637 – 658, 2026. doi:10.1214/25-STS996.

Andreas Maurer and Massimiliano Pontil. Empirical Bernstein Bounds and Sample Variance Penalization. In *Proceedings of the 22nd Annual Conference on Learning Theory (COLT)*, 2009. URL https://arxiv.org/abs/0907.3740.

K. Aiyappan Nair. Variance and Distribution of the Graybill-Deal Estimator of the Common Mean of Two Normal Populations. *The Annals of Statistics*, 8(1):212 – 216, 1980. doi:10.1214/aos/1176344904.

Jerzy Neyman. On the Two Different Aspects of the Representative Method : The Method of Stratified Sampling and the Method of Purposive Selection. *Journal of the Royal Statistical Society*, 97(4): 558–606, 12 1934. doi:10.1111/j.2397-2335.1934.tb04184.x.

Thomas E. Norwood Jr and Klaus Hinkelmann. Estimating the Common Mean of Several Normal Populations. *The Annals of Statistics*, 5(5):1047 – 1050, 1977. doi:10.1214/aos/1176343959.

OpenAI Team. GPT-4 Technical Report. *arXiv preprint arXiv:2303.08774*, 2024. URL https://arxiv.org/abs/2303.08774.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul F Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems*, volume 35, pages 27730–27744. Curran Associates, Inc., 2022. URL https://openreview.net/forum?id=TG8KACxEON.

Nabendu Pal and Wooi K. Lim. A note on second-order admissibility of the Graybill-Deal estimator of a common mean of several normal populations. *Journal of Statistical Planning and Inference*, 63(1):71–78, 1997. doi:10.1016/S0378-3758(96)00202-9.

Nabendu Pal, Jyh-Jiuan Lin, Ching-Hui Chang, and Somesh Kumar. A revisit to the common mean problem: Comparing the maximum likelihood estimator with the Graybill–Deal estimator. *Computational Statistics & Data Analysis*, 51(12):5673–5681, 2007. doi:10.1016/j.csda.2007.04.004.

Benjamin Peherstorfer, Karen Willcox, and Max Gunzburger. Survey of Multifidelity Methods in Uncertainty Propagation, Inference, and Optimization. *SIAM Review*, 60(3):550–591, 2018. doi:10.1137/16M1082469.

Riccardo Poiani, Rémy Degenne, Emilie Kaufmann, Alberto Maria Metelli, and Marcello Restelli. Optimal Multi-Fidelity Best-Arm Identification. In *Advances in Neural Information Processing Systems*, volume 37, pages 121882–121927. Curran Associates, Inc., 2024. URL https://openreview.net/forum?id=gKMTM1i8Ew.

Tiberiu Popoviciu. Sur les équations algébriques ayant toutes leurs racines réelles. *Mathematica (Cluj)*, 9:129–145, 1935.

Qwen Team. Qwen2.5 Technical Report. *arXiv preprint arXiv:2412.15115*, 2024. URL https://arxiv.org/abs/2412.15115.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct Preference Optimization: Your Language Model is Secretly a Reward Model. In *Advances in Neural Information Processing Systems*, volume 36, pages 53728–53741. Curran Associates, Inc., 2023. URL https://openreview.net/forum?id=HPuSIXJaa9.

Ravi Shanker Raju, Swayambhoo Jain, Bo Li, Jonathan Lingjie Li, and Urmish Thakker. Constructing Domain-Specific Evaluation Sets for LLM-as-a-judge. In *Proceedings of the 1st Workshop on Customizable NLP: Progress and Challenges in Customizing NLP for a Domain, Application, Group, or Individual (CustomNLP4U)*, pages 167–181, Miami, Florida, USA, November 2024. Association for Computational Linguistics. doi:10.18653/v1/2024.customnlp4u-1.14.

Alexander Rakhlin and Karthik Sridharan. On equivalence of martingale tail bounds and deterministic regret inequalities. In *Proceedings of the 2017 Conference on Learning Theory*, volume 65 of *Proceedings of Machine Learning Research*, pages 1704–1722. PMLR, 07–10 Jul 2017. URL https://proceedings.mlr.press/v65/rakhlin17a.html.

Aadirupa Saha, Aniket Wagde, and Branislav Kveton. LLM-as-Judge on a Budget. In *Proceedings of The 29th International Conference on Artificial Intelligence and Statistics*, Proceedings of Machine Learning Research. PMLR, 02–05 May 2026. URL https://arxiv.org/abs/2602.15481.

Max Simchowitz and Dylan J. Foster. Naive Exploration is Optimal for Online LQR. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 8937–8948. PMLR, 13–18 Jul 2020. URL https://proceedings.mlr.press/v119/simchowitz20a.html.

Bimal Kumar Sinha and Omar Mouqadem. Estimation of the common mean of two univariate normal populations. *Communications in Statistics - Theory and Methods*, 11(14):1603–1614, 1982. doi:10.1080/03610928208828334.

Aleksandrs Slivkins, Xingyu Zhou, Karthik Abinav Sankararaman, and Dylan J. Foster. Contextual Bandits with Packing and Covering Constraints: A Modular Lagrangian Approach via Regression. *Journal of Machine Learning Research*, 25(394):1–37, 2024. URL http://jmlr.org/papers/v25/24-1220.html.

Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer New York, 2009.

Harry L. van Trees. *Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory*. John Wiley & Sons, New York, 1968.

V. G. Voinov. Variance and Its Unbiased Estimator for the Common Mean of Several Normal Populations. *Sankhyā: The Indian Journal of Statistics, Series B (1960-2002)*, 46(3):291–300, 1984. URL https://www.jstor.org/stable/25052354.

Martin J. Wainwright and Michael I. Jordan. Graphical Models, Exponential Families, and Variational Inference. *Foundations and Trends® in Machine Learning*, 1(1–2):1–305, 2008. doi:10.1561/2200000001.

Zhilin Wang, Yi Dong, Olivier Delalleau, Jiaqi Zeng, Gerald Shen, Daniel Egert, Jimmy J. Zhang, Makesh Narsimhan Sreedhar, and Oleksii Kuchaiev. HelpSteer 2: Open-source dataset for training top-performing reward models. In *Advances in Neural Information Processing Systems*, volume 37, pages 1474–1501. Curran Associates, Inc., 2024. URL https://openreview.net/forum?id=PvVKUFhaNy.

Andrew Chi-Chin Yao. Probabilistic computations: Toward a unified measure of complexity. In *18th Annual Symposium on Foundations of Computer Science (sfcs 1977)*, pages 222–227, 1977. doi:10.1109/SFCS.1977.24.

Bin Yu. Assouad, Fano, and Le Cam. In David Pollard, Erik Torgersen, and Grace L. Yang, editors, *Festschrift for Lucien Le Cam: Research Papers in Probability and Statistics*, pages 423–435. Springer New York, New York, NY, 1997. doi:10.1007/978-1-4612-1880-7_29.

Se-Young Yun and Alexandre Proutière. Optimal Sampling and Clustering in the Stochastic Block Model. In *Advances in Neural Information Processing Systems*, volume 32, pages 13444–13452. Curran Associates, Inc., 2019. URL https://papers.nips.cc/paper_files/paper/2019/hash/8fe04df45a22b63156ebabbb064fcd5e-Abstract.html.

Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric Xing, Hao Zhang, Joseph E. Gonzalez, and Ion Stoica. Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena. In *Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2023. URL https://openreview.net/forum?id=uccHPGDlao.

Yuancheng Zhu and John Lafferty. Quantized Estimation of Gaussian Sequence Models in Euclidean Balls. In *Advances in Neural Information Processing Systems*, volume 27, pages 3662–3670. Curran Associates, Inc., 2014. URL https://arxiv.org/abs/1409.6833.

Contents

1 Introduction **1**

2 Problem Setting **3**

3 Oracle Allocation for Inverse-Variance Weighted Estimation **3**

4 EST-IVWE: An Adaptive Algorithm and its Error Rates **4**

5 Local Minimax Lower Bounds **6**

5.1 A Fano-Type High-Probability Lower Bound 6

5.2 An Assouad-Type In-Expectation Lower Bound 7

5.3 Different Allocation Geometries for Assouad and Fano 8

6 Beyond Boundedness: Gaussian Scores **9**

7 Experiments **9**

8 Conclusion and Future Work **10**

A Proofs from Section 3 (Known Variances) **19**

A.1 Error Rate of **IVWE** 19

A.2 Proof of Theorem 3.2: Optimal Allocation 20

B Proof of Theorem 4.1: Error Rate of Est-IVWE **21**

B.1 Dealing with Variance Mismatch during Aggregation 21

B.2 Bounding (a) 22

B.3 Bounding (b) 23

B.4 Combining Everything and the Large-Budget Regime 24

C Proof of Theorem 5.1: Fano-Type High-Probability Lower Bound **27**

C.1 Proof of Lemma C.2: Optimal Hard Instance 30

D Proof of Theorem 5.2: Assouad-Type In-Expectation Lower Bound **32**

D.1 Lower Bound for Expected ℓ_p-Error 33

D.2 Lower Bound for p-th Moment of the Error 35

D.3 Proof of Lemma D.2: Adaptive Extension of Assouad's Lemma 35

E Proof of Proposition C.1: Construction of Beta Distributions **37**

E.1 Preliminaries and Construction 37

E.2 Proof of (i): Shifted Mean, Same Variance 37

E.3 Proof of (ii): KL of True(Null) and Perturbed Models 38

E.4 Proof of (iii): KL of Two "Adjacent" Perturbed Models 39

F Gaussian Scores **41**

F.1 Upper Bounds 41

F.2 Minimax Lower Bounds 43

G Proofs of Theorems in Appendix F (Gaussian Scores) **45**

G.1 Proof of Theorem F.4: Error Rate of Est-IVWE-Gaussian 45

G.2 Proof of Theorem F.6: Global In-Expectation Lower Bounds for Gaussian Scores . 46

G.3 Proof of Theorem F.8: Global In-Expectation Lower Bound for Gaussian Scores via van Trees Inequality 49

H Additional Experiments **51**

H.1 Details and Additional Results in Experiments on Synthetic Data 51

H.2 Details and Additional Results in Experiments on Real-data 51

H.3 Sensitivity Analysis on Cost-Quality Correlations 52

I Limitations and Future Directions **55**

A Proofs from Section 3 (Known Variances)

A.1 Error Rate of IVWE

Proposition A.1. *Let $p \in [1,\infty]$, $B \in \mathbb{R}_{>0}$ be a total budget, and $\delta \in (0,1)$ be a target confidence level. Also, let $(N_{k,j})_{k,j}$ be any given allocation. Then, **IVWE** is (B,δ,ℓ_p)-budget efficient at any $(\boldsymbol{s},\boldsymbol{\sigma},\boldsymbol{c})$ with the following error rate ε_p:*

$$\varepsilon_p = \sqrt{\frac{2\mathcal{A}_p(\boldsymbol{\omega};\boldsymbol{\sigma},\boldsymbol{c})\log\frac{2K}{\delta}}{B}} + \frac{R\mathcal{B}_p(\boldsymbol{\omega};\boldsymbol{\sigma},\boldsymbol{c})\log\frac{2K}{\delta}}{3B},$$

*where $\boldsymbol{\omega} \coloneqq \left(\omega_{k,j} \triangleq \frac{c_j N_{k,j}}{B}\right)_{k,j}$, the **allocation objective** $\mathcal{A}_p(\boldsymbol{\omega};\boldsymbol{\sigma},\boldsymbol{c})$ is defined as*[2]

$$\mathcal{A}_p(\boldsymbol{\omega};\boldsymbol{\sigma},\boldsymbol{c}) \coloneqq \left(\sum_{k\in[K]}\left(\sum_{j\in[J]}\frac{\omega_{k,j}}{c_j\sigma_{k,j}^2}\right)^{-\frac{p}{2}}\right)^{\frac{2}{p}},$$

and $\mathcal{B}_p(\boldsymbol{\omega};\boldsymbol{\sigma},\boldsymbol{c}) \coloneqq \left(\sum_{k\in[K]}\max_{j:\omega_{k,j}>0}\left(\sigma_{k,j}^2\sum_{j'\in[J]}\frac{\omega_{k,j'}}{c_{j'}\sigma_{k,j'}^2}\right)^{-p}\right)^{\frac{1}{p}}$.

Proof. Let $W_k \coloneqq \sum_{j\in[J]}\frac{N_{k,j}}{\sigma_{k,j}^2}$, and an allocation $(N_{k,j})_{k,j}$ be given. Then, first note that

$$\hat{s}_k - s_k = W_k^{-1}\sum_{j\in[J]}\frac{1}{\sigma_{k,j}^2}\sum_{i=1}^{N_{k,j}}(s_{k,j}^{(i)} - s_k) = \sum_{j\in[J]}\sum_{i=1}^{N_{k,j}}\underbrace{\frac{s_{k,j}^{(i)} - s_k}{W_k\sigma_{k,j}^2}}_{\triangleq X_j^{(i)}}$$

Note that $\mathbb{E}[X_j^{(i)}] = 0$, $|X_j^{(i)}| \le \frac{R}{W_k \min_{j:\omega_{k,j}>0}\sigma_{k,j}^2}$, and $\mathrm{Var}[X_j^{(i)}] = \frac{\mathrm{Var}[s_{k,j}^{(i)} - s_k]}{W_k^2\sigma_{k,j}^4} = \frac{1}{W_k^2\sigma_{k,j}^2}$.

We now utilize the Bernstein's inequality for bounded random variables (Bernstein, 1927) to the aggregated $\hat{s}_k$: with probability at least $1-\delta$, the following holds for for any $k \in [K]$:

$$\begin{aligned}|\hat{s}_k - s_k| &\le \sqrt{2\left(\sum_{j,i}\frac{1}{W_k^2\sigma_{k,j}^2}\right)\log\frac{2K}{\delta}} + \frac{R\log\frac{2K}{\delta}}{3W_k\min_{j:\omega_{k,j}>0}\sigma_{k,j}^2} \\ &= \sqrt{2\log\frac{2K}{\delta}}W_k^{-\frac{1}{2}} + \frac{R\log\frac{2K}{\delta}}{3}\left(W_k\min_{j:\omega_{k,j}>0}\sigma_{k,j}^2\right)^{-1}.\end{aligned}$$

We then apply Minkowski's inequality for ℓ_p-norm as follows:

$$\begin{aligned}&\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_p \\ &\le \left(\sum_{k\in[K]}\left(\sqrt{2\log\frac{2K}{\delta}}W_k^{-\frac{1}{2}} + \frac{R\log\frac{2K}{\delta}}{3}\left(W_k\min_{j:\omega_{k,j}>0}\sigma_{k,j}^2\right)^{-1}\right)^p\right)^{\frac{1}{p}} \\ &\le \left(\sum_{k\in[K]}\left(\sqrt{2\log\frac{2K}{\delta}}W_k^{-\frac{1}{2}}\right)^p\right)^{\frac{1}{p}} + \left(\sum_{k\in[K]}\left(\frac{R\log\frac{2K}{\delta}}{3}\left(W_k\min_{j:\omega_{k,j}>0}\sigma_{k,j}^2\right)^{-1}\right)^p\right)^{\frac{1}{p}} \\ &\qquad\qquad\text{(Minkowski's inequality)} \\ &= \sqrt{2\log\frac{2K}{\delta}}\left(\sum_{k\in[K]}W_k^{-\frac{p}{2}}\right)^{\frac{1}{p}} + \frac{R\log\frac{2K}{\delta}}{3}\left(\sum_{k\in[K]}\left(W_k\min_{j:\omega_{k,j}>0}\sigma_{k,j}^2\right)^{-p}\right)^{\frac{1}{p}}\end{aligned}$$

[2] Note that the allocation objective's limit as $p\to\infty$ is well defined as $\max_{k\in[K]}\left(\sum_{j\in[J]}\frac{\omega_{k,j}}{c_j\sigma_{k,j}^2}\right)^{-1}$.

$$\leq \sqrt{\frac{2}{B}\log\frac{2K}{\delta}} \underbrace{\left(\sum_{k\in[K]}\left(\sum_{j=1}^{J}\frac{\omega_{k,j}}{c_j\sigma_{k,j}^2}\right)^{-\frac{p}{2}}\right)^{\frac{1}{p}}}_{\triangleq\sqrt{\mathcal{A}_p(\boldsymbol{\omega};\boldsymbol{\sigma},\boldsymbol{c})}} + \frac{R\log\frac{2K}{\delta}}{3B} \underbrace{\left(\sum_{k\in[K]}\max_{j:\omega_{k,j}>0}\left(\sigma_{k,j}^2\sum_{j'\in[J]}\frac{\omega_{k,j'}}{c_{j'}\sigma_{k,j'}^2}\right)^{-p}\right)^{\frac{1}{p}}}_{\triangleq\mathcal{B}_p(\boldsymbol{\omega};\boldsymbol{\sigma},\boldsymbol{c})},$$

where in the last inequality, we reparametrize as a ratio of budget allocated to (k,j)-pair, $\omega_{k,j} \coloneqq \frac{c_j N_{k,j}}{B}$, which is invariant to the amount of budget B. □

A.2 Proof of Theorem 3.2: Optimal Allocation

For notational simplicity, let

$$g_k \coloneqq c_{j^*(k)}\sigma_{k,j^*(k)}^2 = \min_{j\in[J]} c_j\sigma_{k,j}^2, \qquad b_k \coloneqq \sum_{j\in[J]}\omega_{k,j}.$$

Here, b_k is the fraction of the budget allocated to query k. Since $g_k > 0$, any allocation with finite objective must have $b_k > 0$ for every k. Increasing all allocations by a common factor decreases the objective, so the budget is fully used at an optimum: $\sum_k b_k = 1$.

For a fixed budget b_k assigned to query k, we have

$$\sum_{j\in[J]}\frac{\omega_{k,j}}{c_j\sigma_{k,j}^2} \leq \frac{b_k}{g_k}.$$

Equality holds when the budget is assigned only to judges minimizing $c_j\sigma_{k,j}^2$. In particular, we may assign all of b_k to a single optimal judge $j^*(k)$; if several judges attain the minimum, splitting the budget among them also gives equality. It remains to optimize the budget fractions across queries.[3]

Optimal Allocation minimizing $\mathcal{A}_\infty$ Using the preceding comparison and $\sum_k b_k = 1$, we obtain

$$\mathcal{A}_\infty(\boldsymbol{\omega};\boldsymbol{\sigma},\boldsymbol{c}) \geq \max_{k\in[K]}\frac{g_k}{b_k} \geq \sum_{k\in[K]} b_k\frac{g_k}{b_k} = \sum_{k\in[K]} g_k.$$

Equality is attained by setting $b_k = g_k/\sum_{k'} g_{k'}$ and assigning each query's budget to an optimal judge. Thus,

$$\omega_{k,j}^* = \mathbb{1}[j = j^*(k)]\frac{g_k}{\sum_{k'} g_{k'}}, \qquad \mathcal{A}_\infty^*(\boldsymbol{\sigma},\boldsymbol{c}) = \sum_k g_k.$$

Optimal Allocation minimizing $\mathcal{A}_p$ For $1 \leq p < \infty$, the same comparison gives

$$\mathcal{A}_p(\boldsymbol{\omega};\boldsymbol{\sigma},\boldsymbol{c}) \geq \left(\sum_k\left(\frac{g_k}{b_k}\right)^{p/2}\right)^{2/p}.$$

By Hölder's inequality with conjugate exponents $(p+2)/2$ and $(p+2)/p$,

$$\sum_k g_k^{\frac{p}{p+2}} = \sum_k\left[\left(\frac{g_k}{b_k}\right)^{p/2}\right]^{\frac{2}{p+2}} b_k^{\frac{p}{p+2}} \leq \left(\sum_k\left(\frac{g_k}{b_k}\right)^{p/2}\right)^{\frac{2}{p+2}}\left(\sum_k b_k\right)^{\frac{p}{p+2}} \leq \mathcal{A}_p(\boldsymbol{\omega};\boldsymbol{\sigma},\boldsymbol{c})^{\frac{p}{p+2}}.$$

Rearranging yields

$$\mathcal{A}_p(\boldsymbol{\omega};\boldsymbol{\sigma},\boldsymbol{c}) \geq \left(\sum_k g_k^{\frac{p}{p+2}}\right)^{\frac{p+2}{p}}.$$

Equality in Hölder's inequality holds when $b_k = g_k^{p/(p+2)}/\sum_{k'} g_{k'}^{p/(p+2)}$. Combining this with the optimal judge choice gives

$$\omega_{k,j}^* = \mathbb{1}[j = j^*(k)]\frac{g_k^{\frac{p}{p+2}}}{\sum_{k'} g_{k'}^{\frac{p}{p+2}}}, \qquad \mathcal{A}_p^*(\boldsymbol{\sigma},\boldsymbol{c}) = \left(\sum_k g_k^{\frac{p}{p+2}}\right)^{\frac{p+2}{p}}.$$

This recovers Theorem 3.2, with the convention $p/(p+2) = (p+2)/p = 1$ when $p = \infty$. □

[3] The KKT analysis yields the same results but is more complicated.

B Proof of Theorem 4.1: Error Rate of Est-IVWE

B.1 Dealing with Variance Mismatch during Aggregation

To show the effect of N_0 and τ, we will tune them as we go along. Let $\overline{\sigma}_{k,j} \coloneqq \tau + \hat{\sigma}_{k,j}$. Then the algorithm computes the allocation $\widehat{\omega}^*_{k,j}$ as

$$\widehat{\omega}^*_{k,j} = \frac{\mathbb{1}[j = \hat{j}^*(k)] \left(c_{\hat{j}^*(k)} \overline{\sigma}^2_{k,\hat{j}^*(k)}\right)^{\frac{p}{p+2}}}{\sum_{k\in[K]} \left(c_{\hat{j}^*(k)} \overline{\sigma}^2_{k,\hat{j}^*(k)}\right)^{\frac{p}{p+2}}},$$

where $\hat{j}^*(k) \coloneqq \arg\min_{j\in[J]} c_j \overline{\sigma}^2_{k,j}$, and allocates each query k to $\hat{j}^*(k)$ for $\widehat{N}_k \coloneqq \left\lceil \frac{B' \widehat{\omega}^*_{k,\hat{j}^*(k)}}{c_{\hat{j}^*(k)}} \right\rceil$ times. Here, $B' = B - CN_0 - Kc_{\max}$ with $C \coloneqq K\sum_j c_j$. Since the budget allocated to Phase-II is at most $B' + \sum_k c_{\hat{j}^*(k)} \le B - CN_0$, the total budget constraint is satisfied under this allocation. Conditional on Phase I, the allocation is fixed and Phase-II samples are independent and fresh. All expressions for $p = \infty$ are interpreted by taking limits.

We first follow the proof of Proposition A.1 while accounting for the variance mismatch: the empirical variance used in **Est-IVWE** does not match the true variance. Applying the Bernstein's inequality to $\hat{s}_k - s_k = \frac{1}{\widehat{N}_k} \sum_{i=1}^{\widehat{N}_k} \left(s^{(i)}_{k,\hat{j}^*(k)} - s_k\right)$, the following holds with probability at least $1-\delta$: for all $k \in [K]$,

$$|\hat{s}_k - s_k| \le \sqrt{\frac{2\sigma^2_{k,\hat{j}^*(k)} \log\frac{2K}{\delta}}{\widehat{N}_k}} + \frac{2R\log\frac{2K}{\delta}}{3\widehat{N}_k}.$$

By Minkowski's inequality, we then have

$$\begin{aligned}
\varepsilon_p &\le \sqrt{2\log\frac{2K}{\delta}} \left(\sum_{k\in[K]} \left(\frac{\widehat{N}_k}{\sigma^2_{k,\hat{j}^*(k)}}\right)^{-\frac{p}{2}}\right)^{\frac{1}{p}} + \frac{2R\log\frac{2K}{\delta}}{3} \left(\sum_{k\in[K]} \widehat{N}_k^{-p}\right)^{\frac{1}{p}} \\
&\le \sqrt{\frac{2\log\frac{2K}{\delta}}{B'}} \underbrace{\left(\sum_{k\in[K]} \left(\frac{\widehat{\omega}^*_{k,\hat{j}^*(k)}}{c_{\hat{j}^*(k)}\sigma^2_{k,\hat{j}^*(k)}}\right)^{-\frac{p}{2}}\right)^{\frac{1}{p}}}_{(a)} + \frac{2R\log\frac{2K}{\delta}}{3B'} \underbrace{\left(\sum_{k\in[K]} \left(\frac{\widehat{\omega}^*_{k,\hat{j}^*(k)}}{c_{\hat{j}^*(k)}}\right)^{-p}\right)^{\frac{1}{p}}}_{(b)}.
\end{aligned}$$

We bound each (a) and (b) separately. We utilize the useful confidence bound on variance estimates given by Maurer and Pontil (2009, Theorem 10). By applying a union bound across all query-judge pairs, we obtain the following result based on the empirical Bernstein inequality:

Lemma B.1 (Theorem 10 of Maurer and Pontil (2009))**.** *With $\hat{\sigma}^2_{k,j}$ computed in Phase I of Algorithm 1,*

$$\mathbb{P}\left(|\hat{\sigma}_{k,j} - \sigma_{k,j}| \le R\sqrt{\frac{2\log\frac{2KJ}{\delta}}{N_0 - 1}}, \quad \forall (k,j) \in [K]\times[J]\right) \ge 1-\delta.$$

We then derive the following lemma that sandwiches our optimistic biased estimates $\overline{\sigma}_{k,j}$'s with the true $\sigma_{k,j}$'s, up to τ: We then derive the following lemma that sandwiches our optimistic biased estimates $\overline{\sigma}_{k,j}$'s with the true $\sigma_{k,j}$'s, up to τ:

Lemma B.2. *Suppose that $\tau \ge R\sqrt{\frac{2}{N_0-1}\log\frac{2KJ}{\delta}}$. Then,*

$$\mathbb{P}\left(\sigma_{k,j} \le \overline{\sigma}_{k,j} \le \sigma_{k,j} + 2\tau, \ \forall (k,j) \in [K]\times[J]\right) \ge 1-\delta.$$

Proof. By Lemma B.1 and the bound on τ, with probability at least $1-\delta$ we have $|\hat{\sigma}_{k,j} - \sigma_{k,j}| \le \tau$ simultaneously for all pairs. On this event,

$$\sigma_{k,j} \le \hat{\sigma}_{k,j} + \tau = \overline{\sigma}_{k,j} \le \sigma_{k,j} + 2\tau,$$

which proves the claim. □

From the above lemma, we choose $\tau = R\sqrt{\frac{2}{N_0 - 1} \log \frac{2KJ}{\delta}}$ before rescaling δ to $\delta/2$ below.

B.2 Bounding (a)

Substituting the definition of $\widehat{\omega}^*_{k,\hat{j}^*(k)}$ into the term (a), we get

$$\begin{aligned}
(a) &= \left(\sum_{k\in[K]} \left(\frac{1}{c_{\hat{j}^*(k)} \sigma^2_{k,\hat{j}^*(k)}} \cdot \frac{\left(c_{\hat{j}^*(k)} \overline{\sigma}^2_{k,\hat{j}^*(k)} \right)^{\frac{p}{p+2}}}{\sum_{k\in[K]} \left(c_{\hat{j}^*(k)} \overline{\sigma}^2_{k,\hat{j}^*(k)} \right)^{\frac{p}{p+2}}} \right)^{-\frac{p}{2}} \right)^{\frac{1}{p}} \\
&= \sqrt{\sum_{k\in[K]} \left(c_{\hat{j}^*(k)} \overline{\sigma}^2_{k,\hat{j}^*(k)} \right)^{\frac{p}{p+2}}} \left(\sum_{k\in[K]} \frac{\sigma^p_{k,\hat{j}^*(k)}}{\overline{\sigma}^p_{k,\hat{j}^*(k)}} \left(c_{\hat{j}^*(k)} \overline{\sigma}^2_{k,\hat{j}^*(k)} \right)^{\frac{p}{p+2}} \right)^{\frac{1}{p}} \\
&\le \sqrt{\sum_{k\in[K]} \left(c_{j^*(k)} \overline{\sigma}^2_{k,j^*(k)} \right)^{\frac{p}{p+2}}} \left(\sum_{k\in[K]} \frac{\sigma^p_{k,\hat{j}^*(k)}}{\overline{\sigma}^p_{k,\hat{j}^*(k)}} \left(c_{j^*(k)} \overline{\sigma}^2_{k,j^*(k)} \right)^{\frac{p}{p+2}} \right)^{\frac{1}{p}}, \\
&\qquad\qquad \text{(Minimality of } \hat{j}^*(k)\text{)}
\end{aligned}$$

where $j^*(k) \triangleq \arg\min_{j\in[J]} c_j \sigma^2_{k,j}$ is the true optimal judge per query; it holds that $c_{\hat{j}^*(k)} \overline{\sigma}^2_{k,\hat{j}^*(k)} \le c_{j^*(k)} \overline{\sigma}^2_{k,j^*(k)}$ by construction of Algorithm 1. With these, we have that with probability at least $1-2\delta$,

$$\begin{aligned}
(a) &\le \sqrt{\sum_{k\in[K]} \left(c_{j^*(k)} \overline{\sigma}^2_{k,j^*(k)} \right)^{\frac{p}{p+2}}} \left(\sum_{k\in[K]} \left(c_{j^*(k)} \overline{\sigma}^2_{k,j^*(k)} \right)^{\frac{p}{p+2}} \right)^{\frac{1}{p}} \\
&= \left(\sum_{k\in[K]} \left(c_{j^*(k)} \overline{\sigma}^2_{k,j^*(k)} \right)^{\frac{p}{p+2}} \right)^{\frac{p+2}{2p}} \qquad (4) \\
&\le \left(\sum_{k\in[K]} \left(\sqrt{c_{j^*(k)}} \sigma_{k,j^*(k)} + 2\tau \sqrt{c_{j^*(k)}} \right)^{\frac{2p}{p+2}} \right)^{\frac{p+2}{2p}}. \qquad \text{(Lemma B.2)}
\end{aligned}$$

Let us denote $q := \frac{2p}{p+2}$. When $p \ge 2$, we have that $q \ge 1$, and thus by Minkowski's inequality of L_q,

$$\begin{aligned}
&\left(\sum_{k\in[K]} \left(\sqrt{c_{j^*(k)}} \sigma_{k,j^*(k)} + 2\tau \sqrt{c_{j^*(k)}} \right)^q \right)^{\frac{1}{q}} \\
&\le \underbrace{\left(\sum_{k\in[K]} \left(\sqrt{c_{j^*(k)}} \sigma_{k,j^*(k)} \right)^q \right)^{\frac{1}{q}}}_{= \sqrt{\mathcal{A}^*_p(\boldsymbol{\sigma}, \boldsymbol{c})}} + \left(\sum_{k\in[K]} \left(2\tau \sqrt{c_{j^*(k)}} \right)^q \right)^{\frac{1}{q}} \\
&= \sqrt{\mathcal{A}^*_p(\boldsymbol{\sigma}, \boldsymbol{c})} + 2\tau \underbrace{\left(\sum_{k\in[K]} \left(\sqrt{c_{j^*(k)}} \right)^q \right)^{\frac{1}{q}}}_{\triangleq \left\| \sqrt{\boldsymbol{c}^*} \right\|_q}.
\end{aligned}$$

When $1 \le p < 2$, we have that $q \in [2/3, 1)$. Then,

$$
\begin{aligned}
&\left(\sum_{k\in[K]} \left(\sqrt{c_{j^*(k)}}\sigma_{k,j^*(k)} + 2\tau\sqrt{c_{j^*(k)}} \right)^q \right)^{\frac{1}{q}} \\
&\le \left(\sum_{k\in[K]} \left(\sqrt{c_{j^*(k)}}\sigma_{k,j^*(k)} \right)^q + \sum_{k\in[K]} \left(2\tau\sqrt{c_{j^*(k)}} \right)^q \right)^{\frac{1}{q}} \qquad \text{(subadditivity of } L_q \text{ when } q \in (0,1)) \\
&\le \underbrace{\left(\sum_{k\in[K]} \left(\sqrt{c_{j^*(k)}}\sigma_{k,j^*(k)} \right)^q \right)^{\frac{1}{q}}}_{=\sqrt{\mathcal{A}_p^*(\boldsymbol{\sigma},\boldsymbol{c})}} \\
&\quad + \frac{1}{q} \left(\sum_{k\in[K]} \left(\sqrt{c_{j^*(k)}}\sigma_{k,j^*(k)} \right)^q + \sum_{k\in[K]} \left(2\tau\sqrt{c_{j^*(k)}} \right)^q \right)^{\frac{1}{q}-1} \sum_{k\in[K]} \left(2\tau\sqrt{c_{j^*(k)}} \right)^q \\
&\qquad (x \mapsto x^{1/q} \text{ is convex for } q \in (0,1), \text{For convex } f,\ f(x+y) \le f(x) + f'(x+y)y) \\
&\le \sqrt{\mathcal{A}_p^*(\boldsymbol{\sigma},\boldsymbol{c})} + \frac{1}{q} \left\{ \left(\sum_{k\in[K]} \left(\sqrt{c_{j^*(k)}}\sigma_{k,j^*(k)} \right)^q \right)^{\frac{1}{q}-1} + (2\tau)^{1-q} \left\| \sqrt{\boldsymbol{c}^*} \right\|_q^{1-q} \right\} (2\tau)^q \left\| \sqrt{\boldsymbol{c}^*} \right\|_q^q \\
&\qquad (x \mapsto x^{\frac{1}{q}-1} \text{ is subadditive for } q \in [2/3, 1)) \\
&= \sqrt{\mathcal{A}_p^*(\boldsymbol{\sigma},\boldsymbol{c})} + \underbrace{\frac{\left\| \sqrt{\boldsymbol{h}^*} \right\|_q^{1-q} 2^q \left\| \sqrt{\boldsymbol{c}^*} \right\|_q^q}{q}}_{\triangleq C_{p,2}^*} \tau^q + \underbrace{\frac{2\left\| \sqrt{\boldsymbol{c}^*} \right\|_q}{q}}_{\triangleq C_{p,1}^*} \tau. \quad \left(\sqrt{\boldsymbol{h}^*} \coloneqq \left(\sqrt{c_{j^*(k)}}\sigma_{k,j^*(k)} \right)_{k\in[K]} \right)
\end{aligned}
$$

We also abuse the notation slightly and denote $C_{p,1}^* = 2\left\| \sqrt{\boldsymbol{c}^*} \right\|_q$ for $p \ge 2$.

Combining the two cases, we have that

$$
(a) \le \sqrt{\mathcal{A}_p^*(\boldsymbol{\sigma},\boldsymbol{c})} + C_{p,1}^* \tau + \mathbb{1}[1 \le p < 2] C_{p,2}^* \tau^{\frac{2p}{p+2}}.
$$

B.3 Bounding (b)

Again substituting in the definition of $\widehat{\omega}_{k,\hat{j}^*(k)}^*$, we go through a somewhat different (from when we proved (a)) series of algebraic manipulations as follows:

$$
\begin{aligned}
(b) &= \left(\sum_{k\in[K]} \left(\frac{1}{c_{\hat{j}^*(k)}} \frac{\left(c_{\hat{j}^*(k)} \overline{\sigma}_{k,\hat{j}^*(k)}^2 \right)^{\frac{p}{p+2}}}{\sum_{k\in[K]} \left(c_{\hat{j}^*(k)} \overline{\sigma}_{k,\hat{j}^*(k)}^2 \right)^{\frac{p}{p+2}}} \right)^{-p} \right)^{\frac{1}{p}} \\
&= \left(\sum_{k\in[K]} \left(c_{\hat{j}^*(k)} \overline{\sigma}_{k,\hat{j}^*(k)}^2 \right)^{\frac{p}{p+2}} \right) \left(\sum_{k\in[K]} \left(\frac{c_{\hat{j}^*(k)}}{\overline{\sigma}_{k,\hat{j}^*(k)}^p} \right)^{\frac{2p}{p+2}} \right)^{\frac{1}{p}} \\
&\le \left(\sum_{k\in[K]} \left(c_{j^*(k)} \overline{\sigma}_{k,j^*(k)}^2 \right)^{\frac{p}{p+2}} \right) \left(\sum_{k\in[K]} \left(\frac{c_{j^*(k)} \overline{\sigma}_{k,j^*(k)}^2}{\overline{\sigma}_{k,\hat{j}^*(k)}^{p+2}} \right)^{\frac{2p}{p+2}} \right)^{\frac{1}{p}} \qquad \text{(Minimality of } \hat{j}^*(k))
\end{aligned}
$$

$$\leq \tau^{-2}\left(\sum_{k\in[K]}\left(c_{j^*(k)}\overline{\sigma}^2_{k,j^*(k)}\right)^{\frac{p}{p+2}}\right)\left(\sum_{k\in[K]}\left(c_{j^*(k)}\overline{\sigma}^2_{k,j^*(k)}\right)^{\frac{2p}{p+2}}\right)^{\frac{1}{p}}$$

($\overline{\sigma}_{k,\hat{j}^*(k)} \geq \tau$ by Lemma B.2)

$$\leq \tau^{-2}\left(\sum_{k\in[K]}\left(c_{j^*(k)}\overline{\sigma}^2_{k,j^*(k)}\right)^{\frac{p}{p+2}}\right)\left(\left(\sum_{k\in[K]}\left(c_{j^*(k)}\overline{\sigma}^2_{k,j^*(k)}\right)^{\frac{p}{p+2}}\right)^2\right)^{\frac{1}{p}}$$

($\sum_k y_k^2 \leq (\sum_k y_k)^2$ for $y_k \geq 0$)

$$= \left\{\tau^{-1}\underbrace{\left(\sum_{k\in[K]}\left(c_{j^*(k)}\overline{\sigma}^2_{k,j^*(k)}\right)^{\frac{p}{p+2}}\right)^{\frac{p+2}{2p}}}_{\text{same as Eqn. (4)!}}\right\}^2.$$

Thus, borrowing our computations from bounding (a) and using $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$, we have that

$$(b) \leq 3\left(\mathcal{A}^*_p(\boldsymbol{\sigma}, \boldsymbol{c})\tau^{-2} + (C^*_{p,1})^2 + \mathbb{1}[1 \leq p < 2](C^*_{p,2})^2\tau^{\frac{2p-4}{p+2}}\right).$$

B.4 Combining Everything and the Large-Budget Regime

We now combine the preceding bounds and specify the choice of N_0. Recall that $C = K\sum_j c_j$ and

$$\tau = R\sqrt{\frac{2}{N_0 - 1}\log\frac{4KJ}{\delta}},$$

where we have rescaled δ to $\delta/2$. With this rescaling, the variance-concentration in Phase I fails with probability at most $\delta/2$, and, conditional on Phase I, the estimation concentration in Phase II fails with probability at most $\delta/2$. Thus, by a union bound, both events hold simultaneously with probability at least $1-\delta$.

Suppose that

$$B \geq 2\left(CN_0 + K\max_{j\in[J]} c_j\right), \tag{5}$$

then we have $B' \geq B/2$. Combining the bounds on (a) and (b) gives, with probability at least $1-\delta$,

$$\begin{aligned}\varepsilon_p \leq{}& \sqrt{\frac{2\log\frac{4K}{\delta}}{B'}}\left(\sqrt{\mathcal{A}^*_p(\boldsymbol{\sigma}, \boldsymbol{c})} + C^*_{p,1}\tau + \mathbb{1}[1 \leq p < 2]C^*_{p,2}\tau^q\right)\\ &+ \frac{2R\log\frac{4K}{\delta}}{B'}\left(\mathcal{A}^*_p(\boldsymbol{\sigma}, \boldsymbol{c})\tau^{-2} + (C^*_{p,1})^2 + \mathbb{1}[1 \leq p < 2](C^*_{p,2})^2\tau^{2q-2}\right),\end{aligned}$$

where $q = \frac{2p}{p+2}$, with $q = 2$ when $p = \infty$.

To separate the oracle term from the cost of learning the variances, note that by Lemma B.3 and (5),

$$\begin{aligned}\varepsilon_p \leq{}& \sqrt{\frac{2\mathcal{A}^*_p(\boldsymbol{\sigma}, \boldsymbol{c})\log\frac{4K}{\delta}}{B}}\\ &+ \frac{\left(CN_0 + K\max_{j\in[J]} c_j\right)\sqrt{2\mathcal{A}^*_p(\boldsymbol{\sigma}, \boldsymbol{c})\log\frac{4K}{\delta}}}{B^{3/2}}\\ &+ 2\sqrt{\frac{\log\frac{4K}{\delta}}{B}}\left(C^*_{p,1}\tau + \mathbb{1}[1 \leq p < 2]C^*_{p,2}\tau^q\right)\\ &+ \frac{4R\log\frac{4K}{\delta}}{B}\left(\mathcal{A}^*_p(\boldsymbol{\sigma}, \boldsymbol{c})\tau^{-2} + (C^*_{p,1})^2 + \mathbb{1}[1 \leq p < 2](C^*_{p,2})^2\tau^{2q-2}\right).\end{aligned} \tag{6}$$

The first term is exactly the leading oracle error rate. The remaining terms capture the loss from the initial exploration, the error in estimating the variances, and the lower-order Bernstein term. We now choose N_0 so that these terms decay strictly faster than $B^{-1/2}$.

Case I: $p \geq 2$**.** In this case, the terms involving $C^*_{p,2}$ vanish. Among the lower-order terms in Eqn. (6), the two dominant terms depending on N_0 are

$$2C^*_{p,1}\sqrt{\frac{\log\frac{4K}{\delta}}{B}}\,\tau = O\left(\frac{R\log\frac{4KJ}{\delta}}{\sqrt{BN_0}}\right),$$

where we used $\log(4K/\delta) \leq \log(4KJ/\delta)$, and

$$\frac{4R\log\frac{4K}{\delta}}{B}\mathcal{A}^*_p(\boldsymbol{\sigma},\boldsymbol{c})\tau^{-2} = O\left(\frac{\mathcal{A}^*_p(\boldsymbol{\sigma},\boldsymbol{c})N_0}{RB}\right).$$

Since $N_0 = 2 \vee \left\lceil (2B)^{1/3}\left(R^2\log\frac{4KJ}{\delta}\right)^{2/3}\right\rceil$, we have $N_0 = \Theta(B^{1/3})$, and $\tau = \Theta(B^{-1/6})$ ignoring factors independent of B. Thus, the two terms above are both of order $B^{-2/3}$. The remaining lower-order terms decay faster as

$$\frac{CN_0 + K\max_j c_j}{B^{3/2}} = O(B^{-7/6})$$

up to fixed instance-dependent factors. Therefore, every term after the leading oracle term in Eqn. (6) is of order at most $B^{-2/3}$. More precisely, using $\log(4K/\delta) \leq \log(4KJ/\delta)$, we finally obtain

$$\varepsilon_p = \sqrt{\frac{2\mathcal{A}^*_p(\boldsymbol{\sigma},\boldsymbol{c})\log\frac{4K}{\delta}}{B}} + \mathcal{O}\left(\frac{R^{1/3}\mathcal{A}^*_p(\boldsymbol{\sigma},\boldsymbol{c})\left(\log\frac{4KJ}{\delta}\right)^{2/3}}{B^{2/3}}\right).$$

Case II: $1 \leq p < 2$**.** Recall that $q = \frac{2p}{p+2} < 1$. Among the lower-order terms in Eqn. (6), the two dominant terms depending on N_0 are

$$2C^*_{p,2}\sqrt{\frac{\log\frac{4K}{\delta}}{B}}\,\tau^q = O\left(\frac{R^q\left(\log\frac{4KJ}{\delta}\right)^{\frac{q+1}{2}}}{\sqrt{B}N_0^{q/2}}\right),$$

where we used $\log(4K/\delta) \leq \log(4KJ/\delta)$, and

$$\frac{4R\log\frac{4K}{\delta}}{B}\mathcal{A}^*_p(\boldsymbol{\sigma},\boldsymbol{c})\tau^{-2} = O\left(\frac{\mathcal{A}^*_p(\boldsymbol{\sigma},\boldsymbol{c})N_0}{RB}\right).$$

The choice of $N_0 = 2 \vee \left\lceil 2^{\frac{p}{2p+2}} R^{\frac{3p+2}{2p+2}}\left(\log\frac{4KJ}{\delta}\right)^{\frac{3p+2}{4p+2}} B^{\frac{p+2}{4p+4}}\right\rceil$ balances the orders of these two terms to the order of $B^{-\frac{3p+2}{4p+4}}$. The remaining lower-order terms in Eqn. (6) decay strictly faster:

$$B^{-1/2}\tau = O\left(B^{-\frac{5p+6}{8p+8}}\right), \qquad \frac{\tau^{2q-2}}{B} = O\left(B^{-\frac{5p+2}{4p+4}}\right),$$

and

$$\frac{CN_0 + K\max_j c_j}{B^{3/2}} = O\left(B^{-\frac{5p+4}{4p+4}}\right), \qquad \frac{1}{B} = O(B^{-1}).$$

Therefore, every term after the leading oracle term is of order at most $B^{-\frac{3p+2}{4p+4}}$. Finally, we obtain

$$\varepsilon_p = \sqrt{\frac{2\mathcal{A}^*_p(\boldsymbol{\sigma},\boldsymbol{c})\log\frac{4K}{\delta}}{B}} + \mathcal{O}\left(\frac{R^{\frac{p}{2p+2}}\mathcal{A}^*_p(\boldsymbol{\sigma},\boldsymbol{c})\left(\log\frac{4KJ}{\delta}\right)^{\frac{3p+2}{4p+4}}}{B^{\frac{3p+2}{4p+4}}}\right).$$

We choose

$$N_0 = \max\left\{2,\ \left\lceil 2^{\frac{p}{2p+2}} R^{\frac{3p+2}{2p+2}}\left(\log\frac{4KJ}{\delta}\right)^{\frac{3p+2}{4p+4}} B^{\frac{p+2}{4p+4}}\right\rceil\right\}.$$

Here the dominant variance-estimation term is $B^{-1/2}\tau^q$, which scales as $B^{-1/2}N_0^{-q/2}$. Balancing this term with the exploration loss N_0/B gives the above choice of N_0. It follows that

$$N_0 = \Theta\left(B^{\frac{p+2}{4p+4}}\right), \qquad \tau = \Theta\left(B^{-\frac{p+2}{8p+8}}\right),$$

and every remainder term in (6) is of order at most $B^{-\frac{3p+2}{4p+4}}$. Therefore,

$$\varepsilon_p = \sqrt{\frac{2\mathcal{A}_p^*(\boldsymbol{\sigma}, \boldsymbol{c}) \log \frac{4K}{\delta}}{B}} + \mathcal{O}\left(\frac{R^{\frac{p}{2p+2}} \mathcal{A}_p^*(\boldsymbol{\sigma}, \boldsymbol{c}) \left(\log \frac{4KJ}{\delta}\right)^{\frac{3p+2}{4p+4}}}{B^{\frac{3p+2}{4p+4}}}\right).$$

For either choice of N_0, we have $N_0 = o(B)$; hence the condition (5) is satisfied for all sufficiently large B. For $p \geq 2$, the remainder is $O(B^{-2/3})$, while for $1 \leq p < 2$ it is $O(B^{-(3p+2)/(4p+4)})$. Both decay strictly faster than the $B^{-1/2}$ oracle term. Therefore, for every fixed instance with $\mathcal{A}_p^*(\boldsymbol{\sigma}, \boldsymbol{c}) > 0$, the ratio of the remainder to the leading oracle term vanishes as $B \to \infty$. Consequently, there exists a finite instance-dependent budget B_0 beyond which the oracle term dominates the remainder.

Thus, beyond a finite instance-dependent budget B_0, the oracle term dominates the remainder, yielding the leading-order instance-optimality of Theorem 4.1. This threshold is neither needed by the learner nor used to set N_0, and need not be uniform over variance profiles: it may become large as $\mathcal{A}_p^*(\boldsymbol{\sigma}, \boldsymbol{c})$ approaches zero. □

We conclude with the proof of the simple algebraic inequality that we have utilized:

Lemma B.3. *For $x \in [0, 1/2]$, it holds that $(1-x)^{-1/2} \leq 1 + x$.*

Proof. Let's define a function $f(x) = 1 + x - (1-x)^{-1/2}$ on $(-\infty, 1)$, then $f(0) = 0$ and $f(0.5) > 0$. Its derivatives are

$$f'(x) = 1 - \frac{1}{2}(1-x)^{-3/2} \quad \text{and} \quad f''(x) = -\frac{3}{4}(1-x)^{-5/2},$$

which imply that f is concave, increasing at $x = 0$ ($f'(0) > 0$), and decreasing at $x = 0.5$ ($f'(0.5)$). Considering the shape of the function f, which has a single mode (maximum) in $(0, 1/2)$, $f(0) = 0$ and $f(0.5) > 0$ imply that $f > 0$ on $[0, 1/2]$. □

C Proof of Theorem 5.1: Fano-Type High-Probability Lower Bound

Our Fano-type change-of-measure argument relies on the binary KL data processing inequality (Garivier et al., 2019), as utilized in Jedra and Proutière (2023) to derive a $\log \frac{1}{\delta}$-dependent minimax lower bound.[4] The proof "bifurcates" at whether $1 \le p < 2$ or $p \ge 2$. For the former, we construct a pair of dense perturbations, while for the latter, we construct a pair of 1-sparse perturbations for each query and combine their information constraints using the budget.

Throughout, we assume that the learner knows the given cost vector $\boldsymbol{c}$ and the true variance profile $\boldsymbol{\sigma}$ with $\max_j \sigma_{k,j}^2 \le R_k^2/2$ for every $k \in [K]$.

Construction of Alternate Models. Let $\mathcal{A}$ be any (B, δ, ℓ_p)-budget efficient algorithm (Definition 2.1) with error rate ε_p, *potentially adaptive*. $\mathcal{A}$ interacts with the true environment $\boldsymbol{s}^\star$ sequentially[5] as follows at each timestep t, when the learner selects query k_t and judge j_t, they observe a score $\tilde{s}_t$. Let us denote $N_{k,j}(T) := \sum_t \mathbb{1}[k_t = k, j_t = j]$ as the number of times $\mathcal{A}$ queries the pair (k, j) throughout the interaction. We note that the knapsack constraint is in place as follows:

$$\sum_j c_j \sum_k N_{k,j}(T) \le B.$$

Let $(\boldsymbol{s}^\star, \boldsymbol{\sigma}, \boldsymbol{c})$ be given and fixed, as we aim for an instance-wise lower bound. The following proposition, whose proof is deferred to Appendix E, provides the desired construction of hypotheses while satisfying the necessary properties:

Proposition C.1. *There exists a family of probability measures $\mathbb{P}_{\boldsymbol{s}^\star+\boldsymbol{\Delta}} = (\mathbb{P}_{\boldsymbol{s}^\star+\boldsymbol{\Delta},k,j})_{k,j}$ indexed by perturbation $\boldsymbol{\Delta} = (\Delta_k)_{k\in[K]}$ with $|\Delta_k| \le \frac{R_k}{4}$ that satisfies the following: for each $(k, j) \in [K] \times [J]$,*

(i) $\mathbb{E}[X] = s_k^\star + \Delta_k$ *and* $\mathrm{Var}[X] = \sigma_{k,j}^2$ *for* $X \sim \mathbb{P}_{\boldsymbol{s}^\star+\boldsymbol{\Delta},k,j}$, *and* $\mathrm{supp}(\mathbb{P}_{\boldsymbol{s}^\star+\boldsymbol{\Delta},k,j}) \subseteq [0, R]$,

(ii) $D_{\mathrm{KL}}\left(\mathbb{P}_{\boldsymbol{s}^\star,k,j}, \mathbb{P}_{\boldsymbol{s}^\star+\boldsymbol{\Delta},k,j}\right) \vee D_{\mathrm{KL}}\left(\mathbb{P}_{\boldsymbol{s}^\star+\boldsymbol{\Delta},k,j}, \mathbb{P}_{\boldsymbol{s}^\star,k,j}\right) \le C_1 \frac{\Delta_k^2}{\sigma_{k,j}^2}$, *and*

(iii) $D_{\mathrm{KL}}\left(\mathbb{P}_{\boldsymbol{s}^\star+\boldsymbol{\Delta},k,j}, \mathbb{P}_{\boldsymbol{s}^\star-\boldsymbol{\Delta},k,j}\right) \vee D_{\mathrm{KL}}\left(\mathbb{P}_{\boldsymbol{s}^\star-\boldsymbol{\Delta},k,j}, \mathbb{P}_{\boldsymbol{s}^\star+\boldsymbol{\Delta},k,j}\right) \le C_2 \frac{\Delta_k^2}{\sigma_{k,j}^2}$.

where $C_1, C_2 > 0$ are absolute constants.

The construction is rather tedious as it is based on appropriate family of beta distributions (which ensures boundedness) while making sure that the variance remains invariant across perturbations, and we need to make sure that the KL-divergence is quadratically bounded as above. (Without the boundedness assumption, one can simply utilize location family of Gaussian distributions.) For this high-probability lower bound proof, we will primarily utilize (i) and (ii), and later for the in-expectation lower bound, we will utilize (ii) and (iii).

We consider the canonical bandit model (See Lattimore and Szepesvári (2020, Section 4.6)) over the observations, and let $\mathbb{P}_{\boldsymbol{s}^\star}$ and $\mathbb{P}_{\boldsymbol{s} \triangleq \boldsymbol{s}^\star+\boldsymbol{\Delta}}$ be the two probability measures induced by the true and alternate instance, respectively. The adversary's choice of perturbation $\boldsymbol{\Delta}$ will be optimized for later.

KL-Divergence Decomposition. By the Divergence Decomposition Lemma (Lattimore and Szepesvári, 2020, Lemma 15.1), the KL divergence between the probability measures $\mathbb{P}_{\boldsymbol{s}^\star}$ and $\mathbb{P}_{\boldsymbol{s}}$ is given by

$$\begin{aligned} D_{\mathrm{KL}}(\mathbb{P}_{\boldsymbol{s}^\star}, \mathbb{P}_{\boldsymbol{s}}) &= \sum_{k\in[K]} \sum_{j\in[J]} \mathbb{E}_{\boldsymbol{s}^\star}^{\mathcal{A}}[N_{k,j}(T)] D_{\mathrm{KL}}\left(\mathbb{P}_{\boldsymbol{s}^\star,k,j}, \mathbb{P}_{\boldsymbol{s},k,j}\right) \\ &\le C_1 \sum_{k\in[K]} \underbrace{\sum_{j\in[J]} \frac{\mathbb{E}_{\boldsymbol{s}^\star}^{\mathcal{A}}[N_{k,j}(T)]}{\sigma_{k,j}^2}}_{\triangleq w_k} (s_k - s_k^\star)^2. \qquad \text{(Proposition C.1 } (ii)\text{)} \end{aligned}$$

[4] This is somewhat reminiscent of Birgé's refinement of Fano's inequality (Gerchinovitz et al., 2020; Tsybakov, 2009).

[5] Even though the algorithm is batched, one could view it as sequentially observing i.i.d. samples. Note that this includes adaptive algorithms.

We first minimize the RHS over perturbations with $\|\boldsymbol{s}-\boldsymbol{s}^\star\|_p \geq 2\varepsilon_p$. The minimizers below lie on the boundary and satisfy $|s_k - s_k^\star| \leq 2\varepsilon_p \leq R_k/4$, so both signs of each perturbation are valid. This is characterized in the next lemma, whose proof is deferred to Appendix C.1:

Lemma C.2. *Let $w_k > 0$ and $\boldsymbol{s}^\star \in \mathbb{R}^K$ be fixed. Let $\boldsymbol{s} = (s_k)_{k\in[K]} \in \mathbb{R}^K$ be an optimal solution of the following constrained optimization:*

$$V^\star = \min_{\boldsymbol{s}\in\mathbb{R}^K} \sum_k w_k(s_k - s_k^\star)^2, \text{ subj. to } \|\boldsymbol{s}-\boldsymbol{s}^\star\|_p \geq 2\varepsilon_p.$$

Then,

1. ***Case I.*** *$1 \leq p < 2$:*

$$s_k = s_k^\star + 2\varepsilon_p \frac{w_k^{-\frac{1}{2-p}}}{\left(\sum_k w_k^{-\frac{p}{2-p}}\right)^{\frac{1}{p}}}, \quad V^\star = \frac{4\varepsilon_p^2}{\left(\sum_k w_k^{-\frac{p}{2-p}}\right)^{\frac{2-p}{p}}}.$$

2. ***Case II.*** *$p \geq 2$: denoting $k^\star \in \arg\min_k w_k$,*

$$\boldsymbol{s} = \boldsymbol{s}^\star + 2\varepsilon_p \boldsymbol{e}_{k^\star}, \quad V^\star = 4\varepsilon_p^2 \min_k w_k.$$

In both cases, note that as $\varepsilon_p \leq \frac{\min_k R_k}{8}$ (as given in the theorem statement), $\boldsymbol{s}$ satisfies $\boldsymbol{s} \in [0,R]^K$ and leads to a valid probability distribution.

Optimal Hard Instances. Now the idea is to construct a set of optimal hard instances and utilize the Bernoulli KL-variant of the data processing inequality (Garivier et al., 2019, Lemma 1), which we recall here:

Lemma C.3 (Data Processing Inequality; Lemma 1 of Garivier et al. (2019))**.** *Consider a measurable space $(\Gamma, \mathcal{G})$ equipped with two probability measure $\mathbb{P}_1$ and $\mathbb{P}_2$. Then, we have that*

$$D_{\mathrm{KL}}(\mathbb{P}_1, \mathbb{P}_2) \geq \sup_Z \mathrm{kl}(\mathbb{E}_1[Z], \mathbb{E}_2[Z]),$$

where $\sup_Z$ is over all possible $\mathcal{G}$-measurable random variable $Z : \Omega \to [0,1]$.

First, let $\boldsymbol{\Delta}$ satisfy $\|\boldsymbol{\Delta}\|_p = 2\varepsilon_p$ and consider the two instances $\boldsymbol{s}^\pm := \boldsymbol{s}^\star \pm \boldsymbol{\Delta}$. Both lie in the prescribed neighborhood and satisfy $|\Delta_k| \leq 2\varepsilon_p \leq R_k/4$. Define their success events by

$$\mathcal{E}_\pm := \{\left\|\hat{\boldsymbol{s}} - \boldsymbol{s}^\pm\right\|_p \leq \varepsilon_p\}.$$

These events are disjoint, since $\|\boldsymbol{s}^+ - \boldsymbol{s}^-\|_p = 4\varepsilon_p > 2\varepsilon_p$. Consequently, for at least one sign $\eta \in \{+,-\}$,

$$\mathbb{P}_{\boldsymbol{s}^\star}(\mathcal{E}_\eta) \leq \tfrac{1}{2}, \qquad \mathbb{P}_{\boldsymbol{s}^\eta}(\mathcal{E}_\eta) \geq 1-\delta.$$

The data processing inequality and monotonicity of binary KL give

$$C_1 \sum_k w_k \Delta_k^2 \geq D_{\mathrm{KL}}(\mathbb{P}_{\boldsymbol{s}^\star}, \mathbb{P}_{\boldsymbol{s}^\eta}) \geq \mathrm{kl}\left(\frac{1}{2}, 1-\delta\right). \tag{7}$$

For $\delta \leq 1/16$, we have

$$\mathrm{kl}\left(\frac{1}{2}, 1-\delta\right) = \frac{1}{2}\log\frac{1}{4\delta(1-\delta)} \geq \frac{1}{2}\log\frac{1}{\delta} - \log 2 \geq \frac{1}{4}\log\frac{1}{\delta}.$$

If any $w_k = 0$, applying Eqn. (7) to $\boldsymbol{\Delta} = 2\varepsilon_p \boldsymbol{e}_k$ gives a contradiction. Hence every budget-efficient algorithm considered here has $w_k > 0$.

Case I. $1 \le p < 2$**.** Choose $\boldsymbol{\Delta}$ from **Case I** of Lemma C.2. Applying Eqn. (7) to the corresponding pair of instances yields

$$\frac{4C_1\varepsilon_p^2}{\left(\sum_k w_k^{-\frac{p}{2-p}}\right)^{\frac{2-p}{p}}} \ge \frac{1}{4}\log\frac{1}{\delta}.$$

Relaxing the expected counts to all nonnegative allocations satisfying the budget, we obtain

$$\begin{aligned}\varepsilon_p^2 &\ge \frac{\log(1/\delta)}{16C_1} \min_{\substack{N_{k,j}\ge 0 \\ \sum_{k,j} c_j N_{k,j} \le B}} \left(\sum_k \left(\sum_j \frac{N_{k,j}}{\sigma_{k,j}^2}\right)^{-\frac{p}{2-p}}\right)^{\frac{2-p}{p}} \\ &= \frac{\log(1/\delta)}{16C_1 B} \underbrace{\left(\sum_k (c_{j^*(k)}\sigma_{k,j^*(k)}^2)^{\frac{p}{2}}\right)^{\frac{2}{p}}}_{=\mathcal{A}^*_{\frac{2p}{2-p}}(\boldsymbol{\sigma},\boldsymbol{c})},\end{aligned}$$

where the last step follows from the allocation optimization in Appendix A.2.

Case II. $p \ge 2$**.** For each k, apply Eqn. (7) to the pair $\boldsymbol{s}^\star \pm 2\varepsilon_p \boldsymbol{e}_k$. This gives

$$4C_1\varepsilon_p^2 w_k \ge \frac{1}{4}\log\frac{1}{\delta}, \qquad \forall k \in [K].$$

By the budget constraint and the definition of $j^*(k)$,

$$\begin{aligned}\sum_k c_{j^*(k)}\sigma_{k,j^*(k)}^2 w_k &= \sum_{k,j} \mathbb{E}^{\mathcal{A}}_{\boldsymbol{s}^\star}[N_{k,j}(T)] \frac{c_{j^*(k)}\sigma_{k,j^*(k)}^2}{\sigma_{k,j}^2} \\ &\le \sum_{k,j} c_j \mathbb{E}^{\mathcal{A}}_{\boldsymbol{s}^\star}[N_{k,j}(T)] \le B.\end{aligned}$$

Multiplying the coordinatewise information constraints by $c_{j^*(k)}\sigma_{k,j^*(k)}^2$ and summing therefore yields

$$\varepsilon_p^2 \ge \frac{\log(1/\delta)}{16C_1 B}\sum_k c_{j^*(k)}\sigma_{k,j^*(k)}^2 = \frac{\log(1/\delta)}{16C_1 B}\mathcal{A}^*_\infty(\boldsymbol{\sigma},\boldsymbol{c}).$$

This argument also covers $p = \infty$; it requires disjointness only between the two opposite perturbations of a given coordinate. Absorbing numerical factors into the universal constants proves the theorem. □

Remark 1 (Why $\mathcal{A}^*_\infty$ appears for $p \ge 2$)**.** *For $p \ge 2$, each alternative used above perturbs only one coordinate. The resulting constraints give*

$$\varepsilon_p^2 \ge \frac{\log(1/\delta)}{32C_1 \min_k w_k}.$$

Thus, minimizing this lower-bound expression over feasible allocations amounts to maximizing $\min_k w_k$. Let $a_k := c_{j^*(k)}\sigma_{k,j^*(k)}^2$. Since the budget implies $\sum_k a_k w_k \le B$, the associated continuous relaxation satisfies

$$\max_{\substack{w_k\ge 0\ \forall k \\ \sum_k a_k w_k \le B}} \min_k w_k = \frac{B}{\sum_k a_k}.$$

Indeed, $\min_k w_k \le B/\sum_k a_k$, with equality when all w_k are equal. This allocation is realized in the continuous relaxation by assigning query k the budget fraction $a_k/\sum_\ell a_\ell$ and using only a judge minimizing $c_j\sigma_{k,j}^2$. Substitution yields the factor $\sum_k a_k = \mathcal{A}^*_\infty(\boldsymbol{\sigma},\boldsymbol{c})$, explaining why this factor appears for every $p \ge 2$.

The equalizing allocation minimizes the lower-bound expression obtained from these single-coordinate tests. The tests provide necessary conditions only, so they do not establish the optimal allocation for finite-p estimation. In particular, the allocation minimizing the aggregate objective $\mathcal{A}_p$ in Theorem 3.2 assigns query budgets proportional to $a_k^{p/(p+2)}$ for finite p.

C.1 Proof of Lemma C.2: Optimal Hard Instance

For notational simplicity, let $\boldsymbol{x} \coloneqq \boldsymbol{s} - \boldsymbol{s}^\star$. By the symmetry of the objective function and the constraint, we can assume without loss of generality that $x_k \geq 0$ for all $k \in [K]$. For $p = \infty$, the constraint requires $\max_k |x_k| \geq 2\varepsilon_p$, so the minimum is $4\varepsilon_p^2 \min_k w_k$, attained at $\boldsymbol{x} = 2\varepsilon_p \boldsymbol{e}_{k^\star}$ for $k^\star \in \arg\min_k w_k$. Henceforth assume $p < \infty$. Then our optimization becomes:

$$V^\star = \min_{\boldsymbol{x} \in \mathbb{R}^K_{\geq 0}} \left\{ F(\boldsymbol{x}) \triangleq \sum_k w_k x_k^2 \right\}, \text{ subj. to } \sum_k x_k^p \geq (2\varepsilon_p)^p.$$

Scaling any feasible $\boldsymbol{x}$ with $\|\boldsymbol{x}\|_p > 2\varepsilon_p$ down to the boundary strictly decreases $F(\boldsymbol{x})$. Thus, the minimum is attained where $\sum_k x_k^p = (2\varepsilon_p)^p$.

We introduce a (bijective) change of variables $y_k = x_k^p$ for $y_k \geq 0$, which maps the non-convex L_p exterior into a standard linear simplex (a convex set). Thus, our optimization becomes:

$$V^\star = \min_{\boldsymbol{y} \in \mathbb{R}^K_{\geq 0}} \left\{ \tilde{F}(\boldsymbol{y}) \triangleq \sum_k w_k y_k^{\frac{2}{p}} \right\}, \text{ subj. to } \sum_k y_k = (2\varepsilon_p)^p.$$

Note that the geometric properties of this optimization, and thus the nature of the optimal solution, depend entirely on the curvature of $\widetilde{F}(\boldsymbol{y})$, which bifurcates at $p = 2$.

Case I: $1 \leq p < 2$**.** In this regime, as the exponent of $\widetilde{F}(\cdot)$ satisfies $\frac{2}{p} > 1$, $\widetilde{F}(\boldsymbol{y})$ is a *strictly convex* function. Minimizing a strictly convex function subject to a linear equality and non-negativity constraints (which trivially satisfies Slater's condition) guarantees that strong duality holds (Boyd and Vandenberghe, 2004, Section 5.2.3).

We thus introduce Lagrangian multiplier vectors $\boldsymbol{\lambda} \in \mathbb{R}^K_{\geq 0}$ and $\nu \in \mathbb{R}$, and the corresponding Lagrangian

$$\mathcal{L}(\boldsymbol{y}, \boldsymbol{\lambda}, \nu) \coloneqq \tilde{F}(\boldsymbol{y}) - \boldsymbol{\lambda}^\top \boldsymbol{y} + \nu \left(\mathbf{1}^\top \boldsymbol{y} - (2\varepsilon_p)^p \right).$$

The Karush-Kuhn-Tucker (KKT) stationarity condition with respect to y_k requires:

$$\frac{\partial \mathcal{L}}{\partial y_k} = \frac{2}{p} w_k y_k^{\frac{2-p}{p}} - \lambda_k + \nu = 0.$$

First, we establish that ν must be strictly negative. Because the primal constraint requires $\sum_k y_k = (2\varepsilon_p)^p > 0$, there must be at least one index i where $y_i > 0$. By complementary slackness, the corresponding dual variable is $\lambda_i = 0$, which implies that $\nu = -\frac{2}{p} w_i y_i^{\frac{2-p}{p}}$. As $w_i > 0$, $y_i > 0$, and for $1 \leq p < 2$ the exponent $\frac{2-p}{p} > 0$, it follows that $\nu < 0$.

Next, we prove that the solution must be strictly interior (i.e., $y_k > 0$ for all k). Assume for contradiction that there exists some index j such that $y_j = 0$. Because $\frac{2-p}{p} > 0$, the j-th stationarity condition becomes

$$0 - \lambda_j + \nu = 0 \implies \lambda_j = \nu < 0,$$

a contradiction, i.e., any solution must be strictly interior. Again, by complementary slackness, we then must have that $\lambda_k = 0$ for all k, which implies that

$$y_k = \left(\frac{-p\nu}{2w_k} \right)^{\frac{p}{2-p}}, \quad \forall k \in [K].$$

Denoting $C \coloneqq (-\frac{p\nu}{2})^{\frac{p}{2-p}} > 0$, we can rewrite the solution as $y_k = C \cdot w_k^{-\frac{p}{2-p}}$. To determine the constant C, we apply the primal equality constraint $\sum_k y_k = (2\varepsilon_p)^p$:

$$\sum_k C \cdot w_k^{-\frac{p}{2-p}} = (2\varepsilon_p)^p \implies C = \frac{(2\varepsilon_p)^p}{\sum_k w_k^{-\frac{p}{2-p}}}.$$

Substituting C back into our expression for y_k yields the optimal coordinates in the transformed space:

$$y_k^\star = (2\varepsilon_p)^p \frac{w_k^{-\frac{p}{2-p}}}{\sum_i w_i^{-\frac{p}{2-p}}}.$$

Converting back to our original variables via $x_k^\star = (y_k^\star)^{\frac{1}{p}}$, we recover the dense allocation:

$$x_k^\star = 2\varepsilon_p \frac{w_k^{-\frac{1}{2-p}}}{\left(\sum_i w_i^{-\frac{p}{2-p}}\right)^{\frac{1}{p}}}.$$

Finally, we substitute $x_k^\star$ into the original objective $F(\boldsymbol{x})$ to find the minimum value $V^\star$:

$$V^\star = \sum_k w_k (x_k^\star)^2 = 4\varepsilon_p^2 \sum_k w_k \frac{w_k^{-\frac{2}{2-p}}}{\left(\sum_i w_i^{-\frac{p}{2-p}}\right)^{\frac{2}{p}}} = 4\varepsilon_p^2 \frac{\sum_k w_k^{-\frac{p}{2-p}}}{\left(\sum_i w_i^{-\frac{p}{2-p}}\right)^{\frac{2}{p}}} = \frac{4\varepsilon_p^2}{\left(\sum_k w_k^{-\frac{p}{2-p}}\right)^{\frac{2-p}{p}}}.$$

Restoring back to $\boldsymbol{s}^\star$ as $s_k^\star = s_k' \pm x_k^\star$ yields **Case I**.

Case II: $p \geq 2$**.** In this regime, as the exponent of $\widetilde{F}(\cdot)$ satisfies $\frac{2}{p} \leq 1$, $\widetilde{F}(\boldsymbol{y})$ is a *concave* function (strictly concave if $p > 2$). Also note that the domain $\mathcal{Y}$ is a simplex defined as follows:

$$\mathcal{Y} := \operatorname{conv}\left(\{(2\varepsilon_p)^p \boldsymbol{e}_1, \cdots, (2\varepsilon_p)^p \boldsymbol{e}_K\}\right), \quad \operatorname{Ext}(\mathcal{Y}) = \{(2\varepsilon_p)^p \boldsymbol{e}_1, \cdots, (2\varepsilon_p)^p \boldsymbol{e}_K\}.$$

Bauer's maximum principle (Bauer, 1958) then states that $\min_{\boldsymbol{y}\in\mathcal{Y}} \widetilde{F}(\boldsymbol{y})$ is attained at at least one extreme point of $\mathcal{Y}$.

At each vertex $(2\varepsilon_p)^p \boldsymbol{e}_k$, we have that

$$\tilde{F}\left((2\varepsilon_p)^p \boldsymbol{e}_k\right) = w_k \left((2\varepsilon_p)^p\right)^{\frac{2}{p}} = 4\varepsilon_p^2 w_k.$$

To find the global minimum across all vertices, we simply choose the index that minimizes the RHS, i.e., the minimum is obtained at $k^\star \in \arg\min_k w_k$. Converting back to $\boldsymbol{x}$, the optimal solution is the following 1-sparse vector:

$$\boldsymbol{x}^\star = 2\varepsilon_p \boldsymbol{e}_{k^\star}, \quad V^\star = 4\varepsilon_p^2 \min_k w_k.$$

Restoring back to $\boldsymbol{s}^\star$ as $s_k^\star = s_k' \pm x_k^\star$ yields **Case II**. □

D Proof of Theorem 5.2: Assouad-Type In-Expectation Lower Bound

By Yao's minimax principle (Yao, 1977), for any prior μ over $\left\{\boldsymbol{s} : \|\boldsymbol{s} - \boldsymbol{s}^\star\|_p \le \xi(B)\right\}$,

$$\inf_{\hat{\boldsymbol{s}}} \sup_{\boldsymbol{s}:\|\boldsymbol{s}-\boldsymbol{s}^\star\|_p \le \xi(B)} \mathbb{E}_{\boldsymbol{s}}^{\hat{\boldsymbol{s}}} \left[\left\|\hat{\boldsymbol{s}} - \boldsymbol{s}\right\|_p\right] \ge \inf_{\hat{\boldsymbol{s}}} \mathbb{E}_{\boldsymbol{s}\sim\mu}^{\hat{\boldsymbol{s}}} \left[\left\|\hat{\boldsymbol{s}} - \boldsymbol{s}\right\|_p\right],$$

and analogously for $\mathbb{E}_{\boldsymbol{s}\sim\mu}^{\hat{\boldsymbol{s}}} \left[\left\|\hat{\boldsymbol{s}} - \boldsymbol{s}\right\|_p^p\right]^{\frac{1}{p}}$ when $1 \le p < \infty$. We will thus lower bound the RHS (Bayes error) with an appropriate choice of μ.

For both in-expectation lower bounds, we utilize the ***Assouad's lemma*** (Assouad, 1983; Yu, 1997)[6] combined with the Bretagnolle-Huber inequality (Bretagnolle and Huber, 1979), which we recall here:

Lemma D.1 (Assouad's lemma)**.** *Let $\mathcal{F}_K = \{-1, 1\}^K$ be the hypercube, $\{\mathbb{P}_{\boldsymbol{v}} : \boldsymbol{v} \in \mathcal{F}_K\}$ be a family of 2^K probability measures indexed by $\boldsymbol{v} \in \mathcal{F}_K$, and $w_k > 0$ be some fixed weights. Then, for any estimator $\hat{\boldsymbol{v}} \in \mathcal{F}_K$,*

$$\mathbb{E}_{\boldsymbol{v}\sim\mathrm{Unif}(\mathcal{F}_K)} \left[\sum_{k\in[K]} w_k \mathbb{1}[\hat{v}_k \ne v_k]\right] \ge \frac{1}{4} \sum_{k\in[K]} w_k e^{-\min\left\{D_{\mathrm{KL}}(\bar{\mathbb{P}}_{+,k}, \bar{\mathbb{P}}_{-,k}), D_{\mathrm{KL}}(\bar{\mathbb{P}}_{-,k}, \bar{\mathbb{P}}_{+,k})\right\}},$$

where $\bar{\mathbb{P}}_{i,k} \coloneqq \frac{1}{2^{K-1}} \sum_{\boldsymbol{v}\in\mathcal{F}_K : v_k = i} \mathbb{P}_{\boldsymbol{v}}$ for $i \in \{-, +\}$ is the marginal distribution for $v_k = i$.

Throughout, we denote $\mathbb{E}_{\boldsymbol{v}} = \mathbb{E}_{\boldsymbol{v}\sim\mathrm{Unif}(\mathcal{F}_K)}$.

Construction. Suppose that $(\boldsymbol{\sigma}, \boldsymbol{c})$ is given and fixed, and let $\boldsymbol{s}^\star \in \mathbb{R}^K$ be a given score vector. To apply Assouad's lemma, we define the prior μ over the score vectors as the uniform distribution over the following hypercube:

$$\mu = \mathrm{Unif}\left(\left\{\boldsymbol{s}(\boldsymbol{v}) \triangleq \boldsymbol{s}^\star + \boldsymbol{v} \odot \boldsymbol{\Delta} \;:\; \boldsymbol{v} \in \{-1, +1\}^K\right\}\right),$$

where $\boldsymbol{\Delta} \coloneqq (\Delta_1, \ldots, \Delta_K) \in \mathbb{R}^K_{\ge 0}$ is chosen a-priori as described for each lower bound.

We utilize the same construction as in Proposition C.1, with $R_k \coloneqq \min\{s_k^\star, R - s_k^\star\} > 0$, $\max_j \sigma_{k,j}^2 \le R_k^2/2$, and $\Delta_k \le R_k/4$. We set $\mathbb{P}_{\boldsymbol{v}} = \mathbb{P}_{\boldsymbol{s}^\star + \boldsymbol{v}\odot\boldsymbol{\Delta}}$.

We, again, consider the canonical bandit model (Lattimore and Szepesvári, 2020, Section 4.6) over the observations, as the estimator may be adaptive. Then, the following lemma, whose proof is provided in Appendix D.3, rigorously establishes the adaptive extension of the Assouad's lemma:

Lemma D.2. *With $\mathbb{P}_{\boldsymbol{v}}$'s as defined and $\bar{N}_{k,j}$ being the number of times (k, j) has been sampled, averaged over $\boldsymbol{v} \sim \mathrm{Unif}(\mathcal{F}_K)$, we have that*

$$\min\left\{D_{\mathrm{KL}}(\bar{\mathbb{P}}_{+,k}, \bar{\mathbb{P}}_{-,k}), D_{\mathrm{KL}}(\bar{\mathbb{P}}_{-,k}, \bar{\mathbb{P}}_{+,k})\right\} \le C_2 \sum_{j\in[J]} \frac{\bar{N}_{k,j}\Delta_k^2}{\sigma_{k,j}^2},$$

satisfying $\sum_{k,j} \bar{N}_{k,j} c_j \le B$.

From hereon and forth, we will denote $V_k \coloneqq \sum_{j\in[J]} \frac{\bar{N}_{k,j}}{\sigma_{k,j}^2}$, which is an algorithm-dependent quantity.

[6] This is derived by applying the Le Cam's Two-Point Method (Le Cam, 1973) coordinate-wise to the hypercube: see Yu (1997) and Tsybakov (2009, Chapter 2) for more detailed discussions.

D.1 Lower Bound for Expected ℓ_p-Error

We first consider $1 \le p < \infty$ and describe our a priori choice of $\mathbf{\Delta}$:

$$\Delta_k = \frac{1}{4\sqrt{V_k^\star}}, \quad V_k^\star := \frac{B\left(c_{j^*(k)}\sigma_{k,j^*(k)}^2\right)^{-\frac{2}{p+2}}}{\sum_{k\in[K]}\left(c_{j^*(k)}\sigma_{k,j^*(k)}^2\right)^{\frac{p}{p+2}}}.$$

Then it is easy to check that $\|\Delta\|_p = \frac{1}{4}\sqrt{\frac{\mathcal{A}_p^*(\boldsymbol{\sigma},\boldsymbol{c})}{B}} = \xi(B)$. Also, we have that $\Delta_k \le \frac{R_k}{4}$ for all k's if

$$B \ge \left(\sum_{k\in[K]}\left(c_{j^*(k)}\sigma_{k,j^*(k)}^2\right)^{\frac{p}{p+2}}\right)\max_{k\in[K]}\frac{\left(c_{j^*(k)}\sigma_{k,j^*(k)}^2\right)^{\frac{2}{p+2}}}{R_k^2},$$

as given in the statement.

Here, as $\mathbb{E}_{\boldsymbol{v}}\left[\left\|\hat{\boldsymbol{s}} - \boldsymbol{s}(\boldsymbol{v})\right\|_p\right]$ is not coordinate-wise decomposable for $p > 1$, we cannot directly utilize Assouad's lemma (Lemma D.1).[7] Instead, we utilize a concavity argument as follows.

For an arbitrary estimator $\hat{\boldsymbol{s}}$, let $\hat{v}_k = +1$ if $\hat{s}_k \ge s_k^\star$ and $\hat{v}_k = -1$ otherwise. The following lemma relates the two:

Lemma D.3. *$|\hat{s}_k - (\boldsymbol{s}(\boldsymbol{v}))_k| \ge \Delta_k \mathbb{1}[\hat{v}_k \ne v_k]$ for any $k \in [K]$ and $\boldsymbol{v} \in \mathcal{F}_K$.*

Proof. W.l.o.g. suppose that $v_k = +1$. If $\hat{v}_k = v_k = +1$, then the bound trivially holds, and thus suppose that $\hat{v}_k = -1 \ne v_k$. Then,

$$|\hat{s}_k - \boldsymbol{s}(\boldsymbol{v})| = |(\hat{s}_k - s_k^\star) - \Delta_k| = -(\hat{s}_k - s_k^\star) + \Delta_k \ge \Delta_k.$$

□

By Lemma D.3, $\|\hat{\boldsymbol{s}} - \boldsymbol{s}(\boldsymbol{v})\|_p \ge S^{1/p}$, where $S := \sum_{k\in[K]}(\Delta_k)^p \mathbb{1}[\hat{v}_k \ne v_k]$ satisfies $0 \le S \le \|\mathbf{\Delta}\|_p^p$. By Assouad's lemma (Lemma D.1), its expectation is bounded as follows:

$$\begin{aligned}\mathbb{E}[S] = \mathbb{E}_{\boldsymbol{v}}\left[\sum_{k\in[K]}(\Delta_k)^p\mathbb{1}[\hat{v}_k \ne v_k]\right] &\ge \frac{1}{4}\sum_{k\in[K]}(\Delta_k)^p \exp\left(-C_2\Delta_k^2 V_k\right) \quad \text{(Lemmas D.1 and D.2)} \\ &= \frac{1}{4}\|\mathbf{\Delta}\|_p^p \underbrace{\sum_{k\in[K]}\frac{(\Delta_k)^p}{\sum_k(\Delta_k)^p}\exp\left(-C_2\Delta_k^2 V_k\right)}_{(*)}.\end{aligned}$$

Note that from our choice of Δ_k's,

$$q_k := \frac{(\Delta_k)^p}{\sum_k(\Delta_k)^p} = \frac{(V_k^\star)^{-p/2}}{\sum_k(V_k^\star)^{-p/2}}, \quad \Delta_k^2 V_k = \frac{1}{16}\frac{V_k}{V_k^\star}.$$

One important observation is that

$$\begin{aligned}\sum_{k\in[K]} q_k \frac{V_k}{V_k^\star} &= \frac{1}{B}\sum_{k\in[K]} c_{j^*(k)}\sigma_{k,j^*(k)}^2 V_k \quad \text{(Direct computation)} \\ &= \frac{1}{B}\sum_{k\in[K]}\sum_{j\in[J]}\bar{N}_{k,j}\frac{c_{j^*(k)}\sigma_{k,j^*(k)}^2}{\sigma_{k,j}^2}\end{aligned}$$

[7] Keen readers may wonder if Jensen's inequality may do the trick, but then would probably despair after finding out that the direction is not as we want.

$$\leq \frac{1}{B} \sum_{k\in[K]} \sum_{j\in[J]} \bar{N}_{k,j} c_j \leq 1.$$

By Jensen's inequality, we have that

$$(*) \geq \exp\left(-\frac{C_2}{16} \sum_{k\in[K]} q_k \frac{V_k}{V_k^\star}\right) \geq \exp\left(-\frac{C_2}{16}\right).$$

Now, we present the concavity argument. For $p \geq 1$, the function $x \mapsto x^{\frac{1}{p}}$ is concave on $[0, \infty)$ and $f(0) = 0$, i.e., geometrically,[8] we have that for any $S \in [0, S_{\max}]$,

$$S^{\frac{1}{p}} \geq \frac{S_{\max}^{\frac{1}{p}} - 0^{\frac{1}{p}}}{S_{\max} - 0} S = S_{\max}^{\frac{1}{p}-1} S.$$

Plugging in $S_{\max} = \|\boldsymbol{\Delta}\|_p^p$, we have that

$$\mathbb{E}_{\boldsymbol{v}}\left[\left\|\hat{\boldsymbol{s}} - \boldsymbol{s}(\boldsymbol{v})\right\|_p\right] \geq \mathbb{E}\left[S^{\frac{1}{p}}\right] \geq \|\boldsymbol{\Delta}\|_p^{1-p} \mathbb{E}[S] \geq \frac{\exp\left(-\frac{C_2}{16}\right)}{4} \|\boldsymbol{\Delta}\|_p = \underbrace{\frac{\exp\left(-\frac{C_2}{16}\right)}{16}}_{\triangleq C_3} \sqrt{\frac{\mathcal{A}_p^*(\boldsymbol{\sigma}, \boldsymbol{c})}{B}}.$$

Constant-Probability Lower Bound. For $1 \leq p < \infty$, the preceding argument gives $0 \leq S \leq S_{\max} = \xi(B)^p$ and $\mathbb{E}[S] \geq 4C_3 S_{\max}$. For any $0 < t < 4C_3$, the reverse Markov inequality follows directly from

$$4C_3 S_{\max} \leq \mathbb{E}[S] \leq t S_{\max} + (1-t) S_{\max} \mathbb{P}(S \geq t S_{\max}),$$

and hence $\mathbb{P}(S \geq t S_{\max}) \geq (4C_3 - t)/(1-t)$. Setting $t = (4C_3 - \delta)/(1-\delta)$ yields the probability bound stated in Section 5, since $\left\|\hat{\boldsymbol{s}} - \boldsymbol{s}(\boldsymbol{v})\right\|_p \geq S^{1/p}$. This probability averages over the hypercube and observations, so at least one fixed vertex satisfies the same bound for each estimator. The conversion is consistent with the local-risk-to-quantile approach of Ma et al. (2026); it requires no independence between the coordinate errors.

The case $p = \infty$. We use the same hypercube construction with equal perturbations

$$\Delta_k = \xi(B) = \frac{1}{4}\sqrt{\frac{\mathcal{A}_\infty^*(\boldsymbol{\sigma}, \boldsymbol{c})}{B}}, \qquad k \in [K].$$

If $B \geq \mathcal{A}_\infty^*(\boldsymbol{\sigma}, \boldsymbol{c}) \max_k R_k^{-2}$, then $\Delta_k \leq R_k/4$ for every k, so the beta construction remains valid, and every vertex lies in the ℓ_∞-ball of radius $\xi(B)$ around $\boldsymbol{s}^\star$.

For this case, take

$$q_k := \frac{c_{j^*(k)} \sigma_{k,j^*(k)}^2}{\mathcal{A}_\infty^*(\boldsymbol{\sigma}, \boldsymbol{c})}, \qquad S := \sum_k q_k \mathbb{1}[\hat{v}_k \neq v_k],$$

using the same sign decoder as above. Since $\mathcal{A}_\infty^*(\boldsymbol{\sigma}, \boldsymbol{c}) = \sum_k c_{j^*(k)} \sigma_{k,j^*(k)}^2$, we have $\sum_k q_k = 1$ and $0 \leq S \leq 1$. The budget constraint gives

$$\xi(B)^2 \sum_k q_k V_k = \frac{1}{16B} \sum_k c_{j^*(k)} \sigma_{k,j^*(k)}^2 V_k \leq \frac{1}{16}.$$

Applying Lemmas D.1 and D.2 and Jensen's inequality, with expectation over the uniform hypercube and observations, yields

$$\begin{aligned}\mathbb{E}[S] &\geq \frac{1}{4} \sum_k q_k \exp\left(-C_2 \xi(B)^2 V_k\right) \\ &\geq \frac{1}{4} \exp\left(-C_2 \xi(B)^2 \sum_k q_k V_k\right) \geq \frac{1}{4} \exp\left(-\frac{C_2}{16}\right) = 4C_3.\end{aligned}$$

[8] For concave functions, any secant line lies below the function.

A single incorrect sign forces ℓ_∞-error at least $\xi(B)$, so

$$S \le \mathbb{1}[\|\hat{\boldsymbol{s}} - \boldsymbol{s}(\boldsymbol{v})\|_\infty \ge \xi(B)].$$

Consequently, for every budget-feasible estimator, at least one fixed vertex $\boldsymbol{s} = \boldsymbol{s}(\boldsymbol{v})$ satisfies

$$\mathbb{P}_{\boldsymbol{s}}^{\hat{\boldsymbol{s}}}\left(\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_\infty \ge \xi(B)\right) \ge 4C_3.$$

At the same vertex,

$$\mathbb{E}_{\boldsymbol{s}}^{\hat{\boldsymbol{s}}}\left[\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_\infty\right] \ge 4C_3\xi(B) = C_3\sqrt{\frac{\mathcal{A}_\infty^*(\boldsymbol{\sigma}, \boldsymbol{c})}{B}}.$$

This proves the expected-error lower bound for $p = \infty$ and gives its constant-probability counterpart directly.

D.2 Lower Bound for p-th Moment of the Error

For $1 \le p < \infty$, we choose $\boldsymbol{\Delta}$ slightly differently as $\Delta_k = \frac{\sqrt{p}}{4\sqrt{V_k^\star}}$ instead of $\Delta_k = \frac{1}{4\sqrt{V_k^\star}}$, with the same $V_k^\star$ as previous. This satisfies $\|\boldsymbol{\Delta}\|_p = \frac{1}{4}\sqrt{\frac{p\mathcal{A}_p^*(\boldsymbol{\sigma}, \boldsymbol{c})}{B}} = \sqrt{p}\xi(B)$. Under the given assumption on B, we also have that $\Delta_k \le \frac{R_k}{4}$.

Let $\hat{\boldsymbol{s}}$ be any estimator, with the same sign decoder $\hat{v}_k$ defined above. Then the absolute error of $\hat{s}_k$ to $(\boldsymbol{s}(\boldsymbol{v}))_k$ with the sign-error is closely related as described in Lemma D.3.

As $\mathbb{E}_{\boldsymbol{v}}\left[\|\hat{\boldsymbol{s}} - \boldsymbol{s}(\boldsymbol{v})\|_p^p\right]$ is coordinate-wise decomposable, applying Lemmas D.1 and D.3 and similar Jensen argument as previous, we have that

$$\begin{aligned}\mathbb{E}_{\boldsymbol{v}}\left[\|\hat{\boldsymbol{s}} - \boldsymbol{s}(\boldsymbol{v})\|_p^p\right] \ge \mathbb{E}_{\boldsymbol{v}}\left[\sum_{k\in[K]} (\Delta_k)^p \mathbb{1}[\hat{v}_k \ne v_k]\right] &\ge \frac{1}{4}\|\boldsymbol{\Delta}\|_p^p \exp\left(-\frac{C_2 p}{16}\right) \\ &= \frac{\exp\left(-\frac{C_2 p}{16}\right)}{4^{1+p}}\left(\frac{p\mathcal{A}_p^*(\boldsymbol{\sigma}, \boldsymbol{c})}{B}\right)^{\frac{p}{2}}\end{aligned}$$

Taking the $1/p$-th power on both sides, we have that

$$\mathbb{E}_{\boldsymbol{v}}\left[\|\hat{\boldsymbol{s}} - \boldsymbol{s}(\boldsymbol{v})\|_p^p\right]^{1/p} \ge \underbrace{\frac{\exp\left(-\frac{C_2}{16}\right)}{4^{1+\frac{1}{p}}}}_{=4^{1-\frac{1}{p}}C_3}\sqrt{\frac{p\mathcal{A}_p^*(\boldsymbol{\sigma}, \boldsymbol{c})}{B}},$$

and we are done. □

D.3 Proof of Lemma D.2: Adaptive Extension of Assouad's Lemma

For $\boldsymbol{u} \in \mathcal{F}_{K-1}$ and $i \in \{-, +\}$, let us denote $\mathbb{P}_{(i,\boldsymbol{u}),k}$ as the probability measure indexed by $\boldsymbol{v}$ defined as $v_k = i$ and $\boldsymbol{v}_{-k} = \boldsymbol{u}$. Also, let $\mathbb{P}_{i,k,j}$ be the probability measure for observations of (k, j) when $v_k = i$.

By the joint convexity of KL and the Divergence Decomposition Lemma (Lattimore and Szepesvári, 2020, Lemma 15.1), we have that

$$\begin{aligned}D_{\mathrm{KL}}(\bar{\mathbb{P}}_{+,k}, \bar{\mathbb{P}}_{-,k}) &\le \frac{1}{2^{K-1}}\sum_{\boldsymbol{u}\in\mathcal{F}_{K-1}} D_{\mathrm{KL}}(\mathbb{P}_{(+,\boldsymbol{u}),k}, \mathbb{P}_{(-,\boldsymbol{u}),k}) \\ &= \sum_{j\in[J]} \underbrace{\frac{1}{2^{K-1}}\sum_{\boldsymbol{u}\in\mathcal{F}_{K-1}} \mathbb{E}_{(+,\boldsymbol{u}),k}[N_{k,j}]}_{\triangleq \bar{N}_{+,k,j}} D_{\mathrm{KL}}(\mathbb{P}_{+,k,j}, \mathbb{P}_{-,k,j}) \\ &\le C_2 \sum_{j\in[J]} \bar{N}_{+,k,j}\frac{\Delta_k^2}{\sigma_{k,j}^2} \qquad \text{(Proposition C.1)}\end{aligned}$$

and similarly,

$$D_{\mathrm{KL}}(\bar{\mathbb{P}}_{-,k}, \bar{\mathbb{P}}_{+,k}) \leq C_2 \sum_{j\in[J]} \bar{N}_{-,k,j} \frac{\Delta_k^2}{\sigma_{k,j}^2}.$$

Thus,

$$\begin{aligned}\min\left\{D_{\mathrm{KL}}(\bar{\mathbb{P}}_{+,k}, \bar{\mathbb{P}}_{-,k}), D_{\mathrm{KL}}(\bar{\mathbb{P}}_{-,k}, \bar{\mathbb{P}}_{+,k})\right\} &\leq \frac{1}{2}\left\{D_{\mathrm{KL}}(\bar{\mathbb{P}}_{+,k}, \bar{\mathbb{P}}_{-,k}) + D_{\mathrm{KL}}(\bar{\mathbb{P}}_{-,k}, \bar{\mathbb{P}}_{+,k})\right\} \\ &\leq C_2 \Delta_k^2 \sum_{j\in[J]} \frac{\bar{N}_{+,k,j} + \bar{N}_{-,k,j}}{2} \frac{1}{\sigma_{k,j}^2}.\end{aligned}$$

Note that

$$\frac{\bar{N}_{+,k,j} + \bar{N}_{-,k,j}}{2} = 2^{-K} \sum_{\boldsymbol{v}\in\mathcal{F}_K} \mathbb{E}_{\boldsymbol{v}}[N_{k,j}] = \bar{N}_{k,j},$$

which satisfies the budget constraint

$$\sum_{k,j} \bar{N}_{k,j} c_j = 2^{-K} \sum_{\boldsymbol{v}\in\mathcal{F}_K} \mathbb{E}_{\boldsymbol{v}}\left[\sum_{k,j} N_{k,j} c_j\right] \leq B.$$

□

E Proof of Proposition C.1: Construction of Beta Distributions

E.1 Preliminaries and Construction

Beta Distribution. For given $\alpha, \beta > 0$, the probability density of $X \sim \mathrm{Beta}(\alpha, \beta)$ is given as

$$f_X(x; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad x \in [0, 1],$$

where $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ is the beta function, and $\Gamma(\cdot)$ is the gamma function. We note that its mean and variance are characterized in closed-form as follows:

$$\mathbb{E}[X] = \frac{\alpha}{\alpha+\beta}, \quad \mathrm{Var}[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Construction. Recall that $R_k = \min\{s_k^\star, R - s_k^\star\}$. Let $\boldsymbol{s}^\star, \boldsymbol{\Delta} \in \mathbb{R}^K$ be given, with $|\Delta_k| \le R_k/4$. Then, for each $\boldsymbol{s} = \boldsymbol{s}^\star + \boldsymbol{\Delta}$, we define the probability measure $\mathbb{P}_{\boldsymbol{s}} = (\mathbb{P}_{\boldsymbol{s},k})_{k\in[K]}$ as follows. For each k, define

$$q_{k,j} \coloneqq \frac{\sigma_{k,j}^2}{4R_k^2}, \quad d_k(\boldsymbol{\Delta}) \coloneqq \frac{\Delta_k}{2R_k},$$

and define

$$\alpha_k(\boldsymbol{\Delta}) \coloneqq \frac{1 - 4d_k(\boldsymbol{\Delta})^2 - 4q_{k,j}}{8q_{k,j}}(1 + 2d_k(\boldsymbol{\Delta})), \quad \beta_k(\boldsymbol{\Delta}) \coloneqq \frac{1 - 4d_k(\boldsymbol{\Delta})^2 - 4q_{k,j}}{8q_{k,j}}(1 - 2d_k(\boldsymbol{\Delta})).$$

Note that under the assumptions that $|\Delta_k| \le R_k/4$ and $\max_j \sigma_{k,j}^2 \le R_k^2/2$, we have that $\alpha_k(\boldsymbol{\Delta}), \beta_k(\boldsymbol{\Delta}) > 0$, and thus, the beta distribution $Y_k(\boldsymbol{\Delta}) \sim \mathrm{Beta}(\alpha_k(\boldsymbol{\Delta}), \beta_k(\boldsymbol{\Delta}))$ is well-defined. With this, $\mathbb{P}_{\boldsymbol{s},k}$ is defined as the measure of the random variable $X_k(\boldsymbol{\Delta}) \coloneqq s_k^\star - R_k + 2R_k Y_k(\boldsymbol{\Delta})$. One can immediately verify that $X_k(\boldsymbol{\Delta}) \in [s_k^\star - R_k, s_k^\star + R_k] \subseteq [0, R]$.

E.2 Proof of (i): Shifted Mean, Same Variance

Mean. By direct computation,

$$\mathbb{E}[Y_k(\Delta)] = \frac{\alpha_k(\boldsymbol{\Delta})}{\alpha_k(\boldsymbol{\Delta}) + \beta_k(\boldsymbol{\Delta})} = \frac{1}{1 + \frac{\beta_k(\boldsymbol{\Delta})}{\alpha_k(\boldsymbol{\Delta})}} = \frac{1}{1 + \frac{1-\frac{\Delta_k}{R_k}}{1+\frac{\Delta_k}{R_k}}} = \frac{\Delta_k + R_k}{2R_k},$$

and thus,

$$\mathbb{E}[X_k(\boldsymbol{\Delta})] = s_k^\star - R_k + 2R_k\mathbb{E}[Y_k(\Delta)] = s_k^\star + \Delta_k.$$

Variance. By direct computation,

$$\begin{aligned}
\mathrm{Var}[Y_k(\Delta)] &= \frac{\alpha_k(\boldsymbol{\Delta})\beta_k(\boldsymbol{\Delta})}{(\alpha_k(\boldsymbol{\Delta}) + \beta_k(\boldsymbol{\Delta}))^2(\alpha_k(\boldsymbol{\Delta}) + \beta_k(\boldsymbol{\Delta}) + 1)} \\
&= \left(\left(1 + \frac{\beta_k(\boldsymbol{\Delta})}{\alpha_k(\boldsymbol{\Delta})}\right)\left(1 + \frac{\alpha_k(\boldsymbol{\Delta})}{\beta_k(\boldsymbol{\Delta})}\right)(1 + \alpha_k(\boldsymbol{\Delta}) + \beta_k(\boldsymbol{\Delta}))\right)^{-1} \\
&= \left(\left(\frac{2R_k}{R_k + \Delta_k}\right)\left(\frac{2R_k}{R_k - \Delta_k}\right)\left(\frac{1 - 4d_k(\boldsymbol{\Delta})^2}{4q_{k,j}}\right)\right)^{-1} \\
&= \frac{R_k^2 - \Delta_k^2}{4R_k^2}\frac{\frac{\sigma_{k,j}^2}{R_k^2}}{1 - \frac{\Delta_k^2}{R_k^2}} = \frac{\sigma_{k,j}^2}{4R_k^2},
\end{aligned}$$

and thus,

$$\mathrm{Var}[X_k(\boldsymbol{\Delta})] = 4R_k^2\mathrm{Var}[Y_k(\Delta)] = \sigma_{k,j}^2.$$

E.3 Proof of (ii): KL of True(Null) and Perturbed Models

Because the KL divergence is invariant under affine transformations, it suffices to compute the KL divergence between the standard beta distributions $Y_k(\boldsymbol{\Delta})$ and $Y_k(\mathbf{0})$:

$$D_{\mathrm{KL}}\left(\mathbb{P}_{\boldsymbol{s}^\star+\boldsymbol{\Delta},k,j}, \mathbb{P}_{\boldsymbol{s}^\star,k,j}\right) = D_{\mathrm{KL}}\left(\mathrm{Beta}(\alpha_k(\boldsymbol{\Delta}), \beta_k(\boldsymbol{\Delta})) \parallel \mathrm{Beta}(\alpha_k(\mathbf{0}), \beta_k(\mathbf{0}))\right).$$

To simplify notation, let $q \coloneqq q_{k,j} = \frac{\sigma_{k,j}^2}{4R_k^2}$ and $d \coloneqq d_k(\boldsymbol{\Delta}) = \frac{\Delta_k}{2R_k}$. Recall our structural assumptions: $|\Delta_k| \le R_k/4$, which implies $|d| \le 1/8$, and $\max_j \sigma_{k,j}^2 \le R_k^2/2$, which implies $q \le 1/8$. The parameters for the alternate and true models respectively are $\boldsymbol{\theta}(d) \coloneqq (\alpha(d), \beta(d))$ and $\boldsymbol{\theta}(0) \coloneqq (\alpha(0), \beta(0))$, defined as:

$$\alpha(d) = \frac{1-4d^2-4q}{8q}(1+2d), \quad \beta(d) = \frac{1-4d^2-4q}{8q}(1-2d).$$

Step 1: Uniform Lower Bound on the Parameter Space. We first establish that the parameters are strictly bounded away from zero. Using $|d| \le 1/8$ and $q \le 1/8$, we can lower bound the common term:

$$1-4d^2-4q \ge 1-4\left(\frac{1}{64}\right)-4\left(\frac{1}{8}\right) = 1-\frac{1}{16}-\frac{1}{2} = \frac{7}{16}.$$

Similarly, $(1 \pm 2d) \ge 1-2(1/8) = \frac{3}{4}$. Thus, both $\alpha(d)$ and $\beta(d)$ are uniformly lower bounded by:

$$\alpha(d), \beta(d) \ge \frac{7/16}{8q}\left(\frac{3}{4}\right) = \frac{21}{512q} \triangleq \frac{C_{\min}}{q},$$

where $C_{\min} = \frac{21}{512}$. This lower bound identically holds for the null parameters $\boldsymbol{\theta}(0)$.

Step 2: Bounding the Parameter Perturbation. Next, we bound the squared Euclidean distance between the parameters $\boldsymbol{\theta}(d)$ and $\boldsymbol{\theta}(0)$ as a function of the shift d:

$$\begin{aligned}
\alpha(d) - \alpha(0) &= \frac{(1-4d^2-4q)(1+2d) - (1-4q)}{8q} \\
&= \frac{1+2d-4d^2-8d^3-4q-8dq-1+4q}{8q} \\
&= \frac{2d-4d^2-8d^3-8dq}{8q} \\
&= \frac{d}{4q}(1-2d-4d^2-4q).
\end{aligned}$$

Applying the absolute bounds $|d| \le 1/8$ and $q \le 1/8$ to the polynomial term yields:

$$|1-2d-4d^2-4q| \le 1+2|d|+4d^2+4q \le 1+\frac{1}{4}+\frac{1}{16}+\frac{1}{2} = \frac{29}{16}.$$

Therefore, $|\alpha(d)-\alpha(0)| \le \frac{29}{64}\frac{|d|}{q}$. By identical algebraic steps, $|\beta(d)-\beta(0)| \le \frac{29}{64}\frac{|d|}{q}$, and thus, the squared Euclidean distance is thus strictly bounded by:

$$\|\boldsymbol{\theta}(d)-\boldsymbol{\theta}(0)\|_2^2 \le 2\left(\frac{29}{64}\right)^2\frac{d^2}{q^2} = \frac{841}{2048}\frac{d^2}{q^2}.$$

Step 3: Exact KL via Bregman Divergence. The Beta distribution forms an exponential family, meaning the KL divergence between two Beta distributions can be written exactly as the Bregman divergence on their log-partition function $m(\alpha,\beta) \triangleq \log B(\alpha,\beta)$ (Wainwright and Jordan, 2008). By Taylor's theorem, there exists some intermediate parameter $\tilde{\boldsymbol{\theta}} = (\tilde{\alpha}, \tilde{\beta})$ lying on the line segment between $\boldsymbol{\theta}(d)$ and $\boldsymbol{\theta}(0)$ such that:

$$D_{\mathrm{KL}}(\mathbb{P}_{\boldsymbol{s}^\star+\boldsymbol{\Delta},k,j}, \mathbb{P}_{\boldsymbol{s}^\star,k,j}) = \frac{1}{2}(\boldsymbol{\theta}(d)-\boldsymbol{\theta}(0))^\top \boldsymbol{I}(\tilde{\boldsymbol{\theta}})(\boldsymbol{\theta}(d)-\boldsymbol{\theta}(0)),$$

where $\boldsymbol{I}(\tilde{\boldsymbol{\theta}}) = \nabla^2 A(\tilde{\boldsymbol{\theta}})$ is the Fisher Information Matrix (FIM) evaluated at $\tilde{\boldsymbol{\theta}}$ (Amari, 2016). The elements of the FIM for the Beta distribution are given by the trigamma function $\psi_1(z) \coloneqq \frac{d^2}{dz^2}\log\Gamma(z)$:

$$\boldsymbol{I}(\alpha,\beta) = \begin{pmatrix} \psi_1(\alpha)-\psi_1(\alpha+\beta) & -\psi_1(\alpha+\beta) \\ -\psi_1(\alpha+\beta) & \psi_1(\beta)-\psi_1(\alpha+\beta) \end{pmatrix}.$$

For any vector $\boldsymbol{v} = (v_1, v_2)^\top$, the quadratic form is $v_1^2\psi_1(\alpha) + v_2^2\psi_1(\beta) - (v_1+v_2)^2\psi_1(\alpha+\beta)$. Since $\psi_1(z) > 0$ for all $z > 0$, the subtracted term is strictly negative, allowing us to upper-bound the quadratic form strictly by its diagonal elements:

$$\boldsymbol{v}^\top \boldsymbol{I}(\alpha,\beta)\boldsymbol{v} \le v_1^2\psi_1(\alpha) + v_2^2\psi_1(\beta) \le \|\boldsymbol{v}\|_2^2 \max(\psi_1(\alpha),\psi_1(\beta)).$$

Step 4: Trigamma Asymptotics and Final Assembly. We use the standard uniform upper bound for the trigamma function: $\psi_1(z) \le \frac{1}{z} + \frac{1}{z^2}$ for $z > 0$. Because $\tilde{\boldsymbol{\theta}}$ lies on the convex line segment between $\boldsymbol{\theta}(d)$ and $\boldsymbol{\theta}(0)$, and both endpoints are uniformly lower-bounded by $C_{\min}/q$, it must be that $\tilde{\alpha}, \tilde{\beta} \ge C_{\min}/q$. Thus,

$$\max(\psi_1(\tilde{\alpha}),\psi_1(\tilde{\beta})) \le \frac{q}{C_{\min}} + \frac{q^2}{C_{\min}^2} = q\left(\frac{1}{C_{\min}} + \frac{q}{C_{\min}^2}\right).$$

Since $q \le 1/8$ and $C_{\min} = \frac{21}{512}$, we can numerically upper bound the term in the parenthesis:

$$\frac{1}{C_{\min}} + \frac{q}{C_{\min}^2} \le \frac{512}{21} + \frac{1/8}{(21/512)^2} = \frac{512}{21} + \frac{32768}{441} = \frac{43520}{441} \triangleq C_\psi.$$

Substituting this maximum eigenvalue bound back into the Bregman divergence yields:

$$\begin{aligned} D_{\mathrm{KL}}(\mathbb{P}_{\boldsymbol{s}^\star+\boldsymbol{\Delta},k,j}, \mathbb{P}_{\boldsymbol{s}^\star,k,j}) &\le \frac{1}{2}\|\boldsymbol{\theta}(d)-\boldsymbol{\theta}(0)\|_2^2 \max(\psi_1(\tilde{\alpha}),\psi_1(\tilde{\beta})) \\ &\le \frac{1}{2}\left(\frac{841}{2048}\frac{d^2}{q^2}\right)(C_\psi q) = \frac{841\cdot 43520}{4096\cdot 441}\frac{d^2}{q} = \frac{71485}{3528}\frac{d^2}{q}. \end{aligned}$$

Finally, mapping back to the original parameterization $d^2 = \frac{\Delta_k^2}{4R_k^2}$ and $q = \frac{\sigma_{k,j}^2}{4R_k^2}$, the $4R_k^2$ denominators cancel out to yield the desired bound:

$$D_{\mathrm{KL}}(\mathbb{P}_{\boldsymbol{s}^\star+\boldsymbol{\Delta},k,j}, \mathbb{P}_{\boldsymbol{s}^\star,k,j}) \le \underbrace{\frac{71485}{3528}}_{\triangleq C_1}\frac{\Delta_k^2}{\sigma_{k,j}^2}.$$

□

E.4 Proof of (iii): KL of Two "Adjacent" Perturbed Models

Similarly, it suffices to bound

$$D_{\mathrm{KL}}\left(\mathbb{P}_{\boldsymbol{s}^\star+\boldsymbol{\Delta},k,j}, \mathbb{P}_{\boldsymbol{s}^\star-\boldsymbol{\Delta},k,j}\right) = D_{\mathrm{KL}}\left(\mathrm{Beta}(\alpha_k(d),\beta_k(d)) \,\|\, \mathrm{Beta}(\alpha_k(-d),\beta_k(-d))\right).$$

First, we evaluate the squared Euclidean perturbation between the endpoints $\boldsymbol{\theta}(d)$ and $\boldsymbol{\theta}(-d)$:

$$\alpha(d) - \alpha(-d) = \frac{1-4d^2-4q}{8q}\left[(1+2d)-(1-2d)\right] = \frac{1-4d^2-4q}{2}\frac{d}{q}.$$

Since $d^2 \ge 0$ and $q \ge 0$, we unconditionally have $1-4d^2-4q \le 1$. Therefore, $|\alpha(d)-\alpha(-d)| \le \frac{1}{2}\frac{|d|}{q}$. By symmetry, the exact same bound applies to $|\beta(d)-\beta(-d)|$, and thus, we have that

$$\|\boldsymbol{\theta}(d)-\boldsymbol{\theta}(-d)\|_2^2 \le 2\left(\frac{1}{2}\right)^2\frac{d^2}{q^2} = \frac{1}{2}\frac{d^2}{q^2}.$$

Following the exact same Bregman divergence formulation from **Step 3** above, there exists an intermediate point $\hat{\boldsymbol{\theta}}$ on the line segment connecting $\boldsymbol{\theta}(-d)$ and $\boldsymbol{\theta}(d)$. By the convexity of the

parameter space and the symmetry of our construction, the endpoints $\boldsymbol{\theta}(-d)$ and $\boldsymbol{\theta}(d)$ both identically satisfy the lower bound $C_{\min}/q$ derived in **Step 1**. Thus, the entire line segment, and consequently $\hat{\boldsymbol{\theta}}$, also satisfies this uniform lower bound.

As a result, the maximum eigenvalue bound of the Fisher Information Matrix remains $\max(\psi_1(\hat{\alpha}), \psi_1(\hat{\beta})) \leq C_\psi q$, utilizing the identical constant $C_\psi = \frac{43520}{441}$. Substituting this into the Bregman divergence quadratic form yields:

$$\begin{aligned} D_{\mathrm{KL}}(\mathbb{P}_{\boldsymbol{s}^\star+\boldsymbol{\Delta},k,j}, \mathbb{P}_{\boldsymbol{s}^\star-\boldsymbol{\Delta},k,j}) &\leq \frac{1}{2} \left\|\boldsymbol{\theta}(d) - \boldsymbol{\theta}(-d)\right\|_2^2 \max(\psi_1(\hat{\alpha}), \psi_1(\hat{\beta})) \\ &\leq \frac{1}{2} \left(\frac{1}{2}\frac{d^2}{q^2}\right) (C_\psi q) = \frac{C_\psi}{4}\frac{d^2}{q} = \frac{10880}{441}\frac{d^2}{q}. \end{aligned}$$

Mapping back to the original parameterization $d^2/q = \Delta_k^2/\sigma_{k,j}^2$, we obtain the non-asymptotic adjacent pair bound:

$$D_{\mathrm{KL}}(\mathbb{P}_{\boldsymbol{s}^\star+\boldsymbol{\Delta},k,j}, \mathbb{P}_{\boldsymbol{s}^\star-\boldsymbol{\Delta},k,j}) \leq \underbrace{\frac{10880}{441}}_{\triangleq C_2} \frac{\Delta_k^2}{\sigma_{k,j}^2}.$$

□

F Gaussian Scores

In this Appendix, we elaborate on how our results generalize to the case when the scores are Gaussian:

Assumption 1 (Gaussian Scores)**.** *If the learner queries $k \in [K]$ to a judge $j \in [J]$, she observes a noisy score $\tilde{s} \sim s_k + \varepsilon_{k,j}$, where $\varepsilon_{k,j} \sim \mathcal{N}(0, \sigma_{k,j}^2)$.*

This then allows us to utilize the following standard Gaussian tail bound:

Lemma F.1. *For $Z \sim \mathcal{N}(0, \sigma^2)$, $\mathbb{P}(|Z| \geq t) \leq 2\exp\left(-t^2/(2\sigma^2)\right)$.*

For clarity, we recall the inverse-variance weighted estimator:

Inverse-Variance Weighted Estimator (IVWE)

Input: An ***allocation*** $(N_{k,j})_{k,j} \in \mathbb{N}_0^{K\times J}$, variance profile $\boldsymbol{\sigma} = (\sigma_{k,j})_{k,j} \in \mathbb{R}_{>0}^{K\times J}$

1. For each (k,j), compute $\hat{s}_{k,j} \coloneqq \frac{1}{N_{k,j}} \sum_{i=1}^{N_{k,j}} s_{k,j}^{(i)}$, if $N_{k,j} \geq 1$; else, set $\hat{s}_{k,j} = 0$;
2. Aggregate the estimators per k as follows:

$$\hat{s}_k \coloneqq \left(\sum_{j=1}^{J} \frac{N_{k,j}}{\sigma_{k,j}^2}\right)^{-1} \sum_{j=1}^{J} \frac{N_{k,j}\hat{s}_{k,j}}{\sigma_{k,j}^2}.$$

Output: $\hat{\boldsymbol{s}} \coloneqq (\hat{s}_k)_{k\in[K]}$

F.1 Upper Bounds

Known Variances. First, when the variances are known, we have the following error bound:

Proposition F.2. *Let $p \in [1,\infty]$, $B \in \mathbb{R}_{>0}$ be a total budget, and $\delta \in (0,1)$ be a target confidence level. Also, let $(N_{k,j})_{k,j}$ be a fixed integer allocation satisfying $N_{k,j} \geq 0$, $\sum_{k,j} c_j N_{k,j} \leq B$, and $\sum_j N_{k,j} \geq 1$ for every k. Then, **IVWE** is (B,δ,ℓ_p)-budget efficient at any $(\boldsymbol{s},\boldsymbol{\sigma},\boldsymbol{c}) \in \mathbb{R}^K \times \mathbb{R}_{>0}^{K\times J} \times \mathbb{R}_{>0}^J$ with the following error rate ε_p:*

$$\varepsilon_p = \sqrt{\frac{2\mathcal{A}_p(\boldsymbol{\omega};\boldsymbol{\sigma},\boldsymbol{c}) \log\frac{2K}{\delta}}{B}},$$

*where $\boldsymbol{\omega} \coloneqq \left(\omega_{k,j} \triangleq \frac{c_j N_{k,j}}{B}\right)_{k,j}$, the **allocation objective** $\mathcal{A}_p(\boldsymbol{\omega};\boldsymbol{\sigma},\boldsymbol{c})$ is defined the same as in Theorem 3.2. Then, with the optimal allocation $\boldsymbol{\omega}^*$ that maximizes $\mathcal{A}_p(\boldsymbol{\omega};\boldsymbol{\sigma},\boldsymbol{c})$, we have $\varepsilon_p = \sqrt{\frac{2\mathcal{A}_p^*(\boldsymbol{\sigma},\boldsymbol{c}) \log\frac{2K}{\delta}}{B}}$.*

Proof Sketch. This follows from applying Lemma F.1 and taking a union bound over $k \in [K]$. □

A sharper confidence bound for $p = 2$. For a fixed allocation with at least one observation per query, the Gaussian structure gives a simple refinement without a union bound over queries. Write $v_k \coloneqq (\sum_j N_{k,j}/\sigma_{k,j}^2)^{-1}$ for the variance of $\hat{s}_k$. Since the coordinate errors are independent Gaussians, the weighted χ^2 bound of Laurent and Massart (2000, Lemma 1) gives, with probability at least $1-\delta$,

$$\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_2^2 \leq \sum_k v_k + 2\sqrt{\log\frac{1}{\delta}\sum_k v_k^2} + 2\log\frac{1}{\delta}\max_k v_k,$$

$$\leq \left(\sqrt{\sum_k v_k} + \sqrt{2 \max_k v_k \log \frac{1}{\delta}} \right)^2 . \tag{8}$$

The second inequality follows from $\sum_k v_k^2 \leq (\max_k v_k) \sum_k v_k$. This separates the overall estimation error from the additional confidence requirement. With one unit-cost judge, identical variances σ^2, and $N_k = B/K$, it gives the example in Section 5. More generally, the allocation minimizing $\sum_k v_k$ need not minimize $\max_k v_k$; this refinement alone therefore does not settle the optimal joint dependence on allocation and confidence.

Readers familiar with Gaussian process concentration may recognize this two-term structure from the classical Tsirelson–Ibragimov–Sudakov (TIS) inequality (Cirel'son et al., 1976; Katz-Samuels et al., 2020): an expected-supremum (Gaussian width) term and a dimension-free deviation term governed by the largest variance. Indeed, applying TIS to the process $\langle u, \hat{\boldsymbol{s}} - \boldsymbol{s} \rangle$ indexed by $\|u\|_2 \leq 1$ also yields the second bound in Eqn. (8), since its expected supremum is $\mathbb{E}[\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_2] \leq \sqrt{\sum_k v_k}$ and its largest variance is $\max_k v_k$.

Unknown Variances. Here, as we have briefly elaborated in Section 4, we can utilize the following multiplicative concentration for the sample variance of a Gaussian:

Lemma F.3 (Eqn. (4.3) & (4.4) of Laurent and Massart (2000)). *With $\hat{\sigma}^2_{k,j}$ as in line 3 of Algorithm 2 and $\epsilon_0 \triangleq \log \frac{2KJ}{\delta}$, suppose that N_0 is an integer satisfying $N_0 - 1 > 4\epsilon_0$, so that the denominators below are positive. Then*

$$\mathbb{P}\left(\frac{\hat{\sigma}^2_{k,j}}{1 + 2\sqrt{\frac{\epsilon_0}{N_0 - 1}} + \frac{2\epsilon_0}{N_0 - 1}} \leq \sigma^2_{k,j} \leq \frac{\hat{\sigma}^2_{k,j}}{1 - 2\sqrt{\frac{\epsilon_0}{N_0 - 1}}}, \quad \forall (k,j) \in [K] \times [J] \right) \geq 1 - \delta.$$

Algorithm 2: Estimate-then-IVWE (Gaussian)

Input: Total budget $B > 0$, Cost vector $\boldsymbol{c} \in \mathbb{R}^J_+$, Number of forced exploration per pair $N_0 \geq 2$

`// Phase I: Forced Exploration`

1 **for** *each* $(k,j) \in [K] \times [J]$ **do**

2 Pull arm (k,j) exactly N_0 times and observe noisy evaluation scores $\left\{ s^{(n)}_{k,j} \right\}_{n \in [N_0]}$;

3 Compute the empirical mean and variance estimator

$$\hat{s}_{k,j} := \frac{1}{N_0} \sum_{n=1}^{N_0} s^{(n)}_{k,j}, \quad \hat{\sigma}^2_{k,j} := \frac{1}{N_0 - 1} \sum_{n=1}^{N_0} \left(s^{(n)}_{k,j} - \hat{s}_{k,j} \right)^2 \tag{9}$$

`// Phase II: IVWE`

4 Compute the optimal allocation $\left(\widehat{\omega}^*_{k,j} \right)_{k,j}$ as in Theorem 3.2 using remaining budget $B' \triangleq B - N_0 K \sum_{j \in [J]} c_j - K(\max_{j \in [J]} c_j) > 0$, based on $\left\{ (\hat{\sigma}_{k,j})^2 \right\}_{k,j}$ Set $\widehat{N}^*_{k,j} = \left\lceil B' \widehat{\omega}^*_{k,j} / c_j \right\rceil$, and query each pair (k,j) exactly $\widehat{N}^*_{k,j}$ times;

5 **return IVWE** with $\left(\widehat{N}^*_{k,j} \right)_{k,j}$ as allocation and $\hat{\sigma}^2_{k,j}$ as variance proxies;

Then we have the following error rate, whose proof is provided in Appendix G.1:

Theorem F.4 (Error Rates of **Est-IVWE-Gaussian**). *Let $p \in [1, \infty]$, $B \in \mathbb{R}_{>0}$, and $\delta \in (0,1)$. Let us set $N_0 = 1 + \lceil 16 \log \frac{4KJ}{\delta} \rceil$ and suppose that $B > N_0 K \sum_{j \in [J]} c_j + K c_{\max}$. Then, **Est-IVWE-Gaussian** is (B, δ, ℓ_p)-budget efficient at any $(\boldsymbol{s}, \boldsymbol{\sigma}, \boldsymbol{c})$ with the following error rate*

ε_p:

$$\varepsilon_p = \sqrt{\frac{13\mathcal{A}_p^*(\boldsymbol{\sigma}, \boldsymbol{c}) \log \frac{4K}{\delta}}{2\left(B - N_0 K \sum_j c_j - K c_{\max}\right)}}.$$

Additionally, if $B \geq 2\left(N_0 K \sum_j c_j + K c_{\max}\right)$*, this rate reduces to*

$$\varepsilon_p \leq \sqrt{\frac{13\mathcal{A}_p^*(\boldsymbol{\sigma}, \boldsymbol{c}) \log \frac{4K}{\delta}}{2B}} + \mathcal{O}\left(\frac{K \sum_j c_j \sqrt{\mathcal{A}_p^*(\boldsymbol{\sigma}, \boldsymbol{c})} \left(\log \frac{4KJ}{\delta}\right)^{3/2}}{B^{3/2}}\right).$$

F.2 Minimax Lower Bounds

We first have the following very similar-looking high-probability lower bound:

Theorem F.5 (Local & High Probability Lower Bound (Gaussian))**.** *Let* $p \in [1, \infty]$*,* $B > 0$*,* $\delta \in (0, e^{-2}]$*, and* $(\boldsymbol{s}, \boldsymbol{\sigma}, \boldsymbol{c})$ *be a given problem instance. Then, any algorithm that is* (B, δ, ℓ_p)*-budget efficient with an error rate of* $\varepsilon_p > 0$ *at any* $(\boldsymbol{s}', \boldsymbol{\sigma}, \boldsymbol{c})$ *with* $\|\boldsymbol{s}' - \boldsymbol{s}\|_p \leq 2\varepsilon_p$ *must satisfy the following: for some absolute constants* $C_1, C_2 > 0$*,*

$$\varepsilon_p \geq \sqrt{\frac{\mathrm{kl}(1-\delta, \delta)}{C_1 B} \mathcal{A}^*_{\frac{2p}{2-p}}(\boldsymbol{\sigma}, \boldsymbol{c})} \quad \text{for } 1 \leq p < 2, \qquad \varepsilon_p \geq \sqrt{\frac{\log \frac{1}{\delta}}{C_2 B} \mathcal{A}^*_\infty(\boldsymbol{\sigma}, \boldsymbol{c})} \quad \text{for } p \geq 2.$$

Proof. The proof is almost the same as that of Theorem 5.1, except we now construct the instances via Gaussians instead of the Beta distributions, resulting in constant-wise difference only. □

As in the bounded setting, this Fano-type change-of-measure argument captures the $\log \frac{1}{\delta}$ dependence. We complement it with global in-expectation lower bounds that recover $\mathcal{A}_p^*$, using different proof techniques that may be of independent interest.[9] Again, the $\inf_{\hat{\boldsymbol{s}}}$ is over all adaptive algorithms satisfying the budget constraint. In particular, the lower bound holds even when the learner is given the true variance profile $\boldsymbol{\sigma}$.

Theorem F.6 (Global, In-Expectation Minimax Lower Bounds)**.** *For each* $1 \leq p \leq \infty$*,*

$$\inf_{\hat{\boldsymbol{s}}} \sup_{\boldsymbol{s} \in \mathbb{R}^K} \mathbb{E}^{\hat{\boldsymbol{s}}}_{\boldsymbol{s} \sim \mu} \left[\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_p\right] \geq \sqrt{\frac{2}{\pi}} \sqrt{\frac{\mathcal{A}_p^*(\boldsymbol{\sigma}, \boldsymbol{c})}{B}}.$$

For $1 \leq p < \infty$*, we also have:*

$$\inf_{\hat{\boldsymbol{s}}} \sup_{\boldsymbol{s} \in \mathbb{R}^K} \mathbb{E}^{\hat{\boldsymbol{s}}}_{\boldsymbol{s} \sim \mu} \left[\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_p^p\right]^{\frac{1}{p}} \geq m_p \sqrt{\frac{\mathcal{A}_p^*(\boldsymbol{\sigma}, \boldsymbol{c})}{B}},$$

where[a] $m_p := \sqrt{2}\left(\frac{\Gamma((p+1)/2)}{\sqrt{\pi}}\right)^{\frac{1}{p}} = \mathbb{E}[|Z|^p]^{1/p}$ *for* $Z \sim \mathcal{N}(0, 1)$.

[a] $\Gamma(\cdot)$ is the usual gamma function.

Proof Sketch. The full proof is deferred to Appendix G.2. Instead of following the standard lower bound framework (Tsybakov, 2009; Yu, 1997), we heavily exploit the fact that everything is Gaussian. Specifically, this implies that the posterior risk is minimized at the posterior mean, and everything (e.g., posterior distribution, Fisher information) is computable in closed forms. □

[9] One may wonder whether it is possible to utilize Assouad's lemma (Assouad, 1983; Yu, 1997) as done in our Theorem 5.2. This is definitely possible that leads to a similar local minimax lower bound. Here, we intend to show a completely different argument for global minimax lower bounds.

When the *variances are known* and the continuous oracle allocation has integer sample counts, the lower bound for the p-th moment is constant-wise tight for every finite p, as shown by the following matching upper bound:

Theorem F.7. *Let $1 \le p < \infty$. When the variances are* known *and the continuous oracle allocation at budget B has integer sample counts, **IVWE** with optimal allocation achieves the following:*

$$\mathbb{E}\left[\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_p^p\right]^{\frac{1}{p}} \le m_p \sqrt{\frac{\mathcal{A}_p^*(\boldsymbol{\sigma}, \boldsymbol{c})}{B}}.$$

Proof. Recall that $\hat{s}_k - s_k \sim \mathcal{N}\left(0, \left(\sum_{j=1}^J \frac{N_{k,j}}{\sigma_{k,j}^2}\right)^{-1}\right)$. Thus, we have that

$$\mathbb{E}\left[\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_p^p\right]^{\frac{1}{p}} = \mathbb{E}\left[\sum_{k\in[K]} |\hat{s}_k - s_k|^p\right]^{\frac{1}{p}} \le m_p \left(\sum_{k\in[K]} \left(\sum_{j=1}^J \frac{N_{k,j}}{\sigma_{k,j}^2}\right)^{-\frac{p}{2}}\right)^{\frac{1}{p}},$$

where $m_p = \mathbb{E}[|Z|^p]^{1/p}$ for $Z \sim \mathcal{N}(0,1)$ as in Theorem F.6. Then, we conclude by Theorem 3.2. □

We lastly show another global lower bound for $\mathbb{E}^{\hat{\boldsymbol{s}}}_{\boldsymbol{s}\sim\mu}\left[\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_p^p\right]^{\frac{1}{p}}$ based on a completely different (and from our perspective, slightly less well-known) technique that may be of an independent interest:

Theorem F.8. *For $1 \le p < \infty$, we have:*

$$\inf_{\hat{\boldsymbol{s}}} \sup_{\boldsymbol{s}\in\mathbb{R}^K} \mathbb{E}^{\hat{\boldsymbol{s}}}_{\boldsymbol{s}\sim\mu}\left[\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_p^p\right]^{\frac{1}{p}} \ge \frac{\sqrt{2\pi e}}{C_{\mathrm{ME}}(p)} \sqrt{\frac{\mathcal{A}_p^*(\boldsymbol{\sigma}, \boldsymbol{c})}{B}},$$

where $C_{\mathrm{ME}}(p) \triangleq 2e^{\frac{1}{p}}\Gamma(p^{-1})p^{\frac{1}{p}-1}$ is the partition function of the L_p max-entropy distribution.

Proof Sketch. The full proof is deferred to Appendix G.3. The key technical tool is the ***van Trees inequality*** (Gill and Levit, 1995; van Trees, 1968), recently extended to L_p-norm by Chen and Özgür (2024), with cosine-squared prior (Tsybakov, 2009) and taking a limit to uninformative prior. □

Note that when $p = 2$, this bound is also constant-wise optimal: $m_2 = 1 = \left(\frac{\sqrt{2\pi e}}{C_{\mathrm{ME}}(2)}\right)^2$.

G Proofs of Theorems in Appendix F (Gaussian Scores)

G.1 Proof of Theorem F.4: Error Rate of Est-IVWE-Gaussian

Denote the lower and upper bounds on variances $\sigma_{k,j}^2$ after the exploration phase by

$$\underline{\sigma}_{k,j}^2 = \frac{\hat{\sigma}_{k,j}^2}{1 + 2\sqrt{\frac{\epsilon_0}{N_0 - 1}} + \frac{2\epsilon_0}{N_0 - 1}} \quad \text{and} \quad \overline{\sigma}_{k,j}^2 = \frac{\hat{\sigma}_{k,j}^2}{1 - 2\sqrt{\frac{\epsilon_0}{N_0 - 1}}},$$

where $\epsilon_0 = \log \frac{4KJ}{\delta}$. By Lemma F.3, we know that

$$\mathbb{P}\left(\underline{\sigma}_{k,j}^2 \le \sigma_{k,j}^2 \le \overline{\sigma}_{k,j}^2, \quad \forall (k,j) \in [K] \times [J]\right) \ge 1 - \frac{\delta}{2},$$

which we will assume to hold throughout the proof. We denote the ratio between the upper and lower bounds as $\kappa = \overline{\sigma}_{k,j}^2 / \underline{\sigma}_{k,j}^2$, which is the same across all pairs (k,j). Since $N_0 = 1 + \lceil 16 \log \frac{4KJ}{\delta} \rceil$, $\epsilon_0/(N_0 - 1) \le \frac{1}{16}$, we have that $\kappa \le \frac{13}{4}$.

Recall that **Est-IVWE-Gaussian** (Algorithm 2) designates an (predicted) optimal judge $\hat{j}^*(k) \triangleq \arg\min_{j \in [J]} c_j \hat{\sigma}_{k,j}^2$ for each query $k \in [K]$, then computes the optimal allocation with variance estimator $\hat{\sigma}_{k,j}^2$ and available continuous budget $B' = B - CN_0 - Kc_{\max} > 0$ with $C \coloneqq K \sum_{j \in [J]} c_j$ as follows:

$$\widehat{\omega}_{k,\hat{j}^*(k)}^* = \frac{\left(c_{\hat{j}^*(k)} \hat{\sigma}_{k,\hat{j}^*(k)}^2\right)^{\frac{p}{p+2}}}{\sum_{k' \in [K]} \left(c_{\hat{j}^*(k')} \hat{\sigma}_{k',\hat{j}^*(k')}^2\right)^{\frac{p}{p+2}}}, \quad \widehat{N}_{k,j}^* = \left\lceil \frac{B' \widehat{\omega}_{k,j}^*}{c_j} \right\rceil.$$

There is one positive allocation per query, so rounding costs at most $\sum_k c_{\hat{j}^*(k)} \le Kc_{\max}$; together with exploration, the total cost is at most B. Conditional on Phase I, the allocation and selected judges are fixed, and the fresh Phase-II sample mean $\hat{s}_k$ is Gaussian with variance $\sigma_{k,\hat{j}^*(k)}^2 / \widehat{N}_{k,\hat{j}^*(k)}^*$. Thus, by Lemma F.1 and a union bound, the following event has conditional probability at least $1 - \delta/2$:

$$\mathbb{P}\left(|\hat{s}_k - s_k| \le \sqrt{\frac{2\sigma_{k,\hat{j}^*(k)}^2 \log \frac{4K}{\delta}}{\widehat{N}_{k,\hat{j}^*(k)}^*}}, \quad \forall k \in [K]\right) \ge 1 - \frac{\delta}{2}.$$

Thus, by summing up p-th power of these terms over $k \in [K]$ and substituting the definition of $\widehat{\omega}_{k,j}^*$,the following holds with probability at least $1 - \delta$:

$$\begin{aligned}
\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_p &\le \sqrt{2 \log \frac{4K}{\delta}} \left(\sum_{k \in [K]} \left(\frac{\widehat{N}_{k,\hat{j}^*(k)}^*}{\sigma_{k,\hat{j}^*(k)}^2} \right)^{-\frac{p}{2}} \right)^{\frac{1}{p}} \\
&\le \sqrt{\frac{2 \log \frac{4K}{\delta}}{B'}} \left(\sum_{k \in [K]} \left(\frac{\widehat{\omega}_{k,\hat{j}^*(k)}}{c_{\hat{j}^*(k)} \sigma_{k,\hat{j}^*(k)}^2} \right)^{-\frac{p}{2}} \right)^{\frac{1}{p}} \\
&= \sqrt{\frac{2 \log \frac{4K}{\delta}}{B'}} \left(\sum_{k \in [K]} \left(\frac{1}{c_{\hat{j}^*(k)} \sigma_{k,\hat{j}^*(k)}^2} \cdot \frac{\left(c_{\hat{j}^*(k)} \hat{\sigma}_{k,\hat{j}^*(k)}^2\right)^{\frac{p}{p+2}}}{\sum_{k' \in [K]} \left(c_{\hat{j}^*(k')} \hat{\sigma}_{k',\hat{j}^*(k')}^2\right)^{\frac{p}{p+2}}} \right)^{-\frac{p}{2}} \right)^{\frac{1}{p}} \\
&= \sqrt{\frac{2 \log \frac{4K}{\delta}}{B'}} \sqrt{\sum_{k \in [K]} \left(c_{\hat{j}^*(k)} \hat{\sigma}_{k,\hat{j}^*(k)}^2\right)^{\frac{p}{p+2}}} \left(\sum_{k \in [K]} \left(\frac{\sigma_{k,\hat{j}^*(k)}}{\hat{\sigma}_{k,\hat{j}^*(k)}} \right)^p \left(c_{\hat{j}^*(k)} \hat{\sigma}_{k,\hat{j}^*(k)}^2\right)^{\frac{p}{p+2}} \right)^{\frac{1}{p}} \\
&\le \sqrt{\frac{1}{1 - 2\sqrt{\frac{\epsilon_0}{N_0 - 1}}}} \sqrt{\frac{2 \log \frac{4K}{\delta}}{B'}} \left(\sum_{k \in [K]} \left(c_{\hat{j}^*(k)} \hat{\sigma}_{k,\hat{j}^*(k)}^2\right)^{\frac{p}{p+2}} \right)^{\frac{p+2}{2p}},
\end{aligned}$$

where the last inequality holds since $\sigma^2_{k,j}/\hat{\sigma}^2_{k,j} \le \left(1 - 2\sqrt{\frac{\epsilon_0}{N_0 - 1}}\right)^{-1}$ for any $(k, j) \in [K] \times [J]$ by Lemma F.3. When $p = \infty$, we set $p/(p+2) = 1$ and $(p+2)/p = 1$ by convention, and the above inequality holds in that case as well:

$$\begin{aligned}
\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_\infty &\le \max_{k\in[K]} \sqrt{2\log\frac{4K}{\delta}\frac{\sigma^2_{k,\hat{j}^*(k)}}{\widehat{N}^*_{k,\hat{j}^*(k)}}} \\
&\le \max_{k\in[K]} \sqrt{\frac{2\log\frac{4K}{\delta}}{B'}\frac{c_{\hat{j}^*(k)}\sigma^2_{k,\hat{j}^*(k)}}{\widehat{\omega}^*_{k,\hat{j}^*(k)}}} \\
&= \sqrt{\frac{2\log\frac{4K}{\delta}}{B'}}\left(\max_{k\in[K]} \sqrt{\frac{\sigma^2_{k,\hat{j}^*(k)}}{\hat{\sigma}^2_{k,\hat{j}^*(k)}}\sum_{k'\in[K]} c_{\hat{j}^*(k')}\hat{\sigma}^2_{k',\hat{j}^*(k')}}\right) \\
&\le \sqrt{\frac{1}{1 - 2\sqrt{\frac{\epsilon_0}{N_0-1}}}}\sqrt{\frac{2\log\frac{4K}{\delta}}{B'}}\sqrt{\sum_{k\in[K]} c_{\hat{j}^*(k)}\hat{\sigma}^2_{k,\hat{j}^*(k)}}.
\end{aligned}$$

As $c_{\hat{j}^*(k)}\hat{\sigma}^2_{k,\hat{j}^*(k)} \le c_{j^*(k)}\hat{\sigma}^2_{k,j^*(k)}$ by the definition of $\hat{j}^*(k)$, we have

$$\begin{aligned}
&\left(\sum_{k\in[K]}\left(c_{\hat{j}(k)}\hat{\sigma}^2_{k,\hat{j}(k)}\right)^{\frac{p}{p+2}}\right)^{\frac{p+2}{2p}} \\
&\le \left(\sum_{k\in[K]}\left(c_{j^*(k)}\hat{\sigma}^2_{k,j^*(k)}\right)^{\frac{p}{p+2}}\right)^{\frac{p+2}{2p}} \\
&\le \sqrt{1 + 2\sqrt{\frac{\epsilon_0}{N_0-1}} + \frac{2\epsilon_0}{N_0-1}}\left(\sum_{k\in[K]}\left(c_{j^*(k)}\sigma^2_{k,j^*(k)}\right)^{\frac{p}{p+2}}\right)^{\frac{p+2}{2p}}. \qquad \text{(Lemma F.3)}
\end{aligned}$$

Combining the two events, with probability at least $1 - \delta$ we have

$$\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_p \le \sqrt{\frac{2\kappa\mathcal{A}^*_p(\boldsymbol{\sigma}, \boldsymbol{c})\log\frac{4K}{\delta}}{B'}} \le \sqrt{\frac{13\mathcal{A}^*_p(\boldsymbol{\sigma}, \boldsymbol{c})\log\frac{4K}{\delta}}{2(B - CN_0 - Kc_{\max})}},$$

where $\kappa \le 13/4$. This proves the exact bound for every $B > CN_0 + Kc_{\max}$. If $B \ge 2(CN_0 + Kc_{\max})$, we may additionally apply Lemma B.3 to obtain

$$\begin{aligned}
\varepsilon_p &\le \sqrt{\frac{13\mathcal{A}^*_p(\boldsymbol{\sigma}, \boldsymbol{c})\log\frac{4K}{\delta}}{2B}} + \frac{(CN_0 + Kc_{\max})\sqrt{\frac{13}{2}\mathcal{A}^*_p(\boldsymbol{\sigma}, \boldsymbol{c})\log\frac{4K}{\delta}}}{B^{3/2}} \\
&= \sqrt{\frac{13\mathcal{A}^*_p(\boldsymbol{\sigma}, \boldsymbol{c})\log\frac{4K}{\delta}}{2B}} + \mathcal{O}\left(\frac{C\sqrt{\mathcal{A}^*_p(\boldsymbol{\sigma}, \boldsymbol{c})}\left(\log\frac{4KJ}{\delta}\right)^{3/2}}{B^{3/2}}\right).
\end{aligned}$$

Here $Kc_{\max} \le C$ and $N_0 = \mathcal{O}(\log(4KJ/\delta))$, with universal constants. In particular, the expansion requires the stronger budget condition, while the unexpanded bound requires only a positive Phase-II budget.

G.2 Proof of Theorem F.6: Global In-Expectation Lower Bounds for Gaussian Scores

Again, we start with the Yao's minimax principle (Yao, 1977), but without the locality: for any prior μ over $\mathbb{R}^K$ (note that now, we consider the prior over the entire $\mathbb{R}^K$ instead of near some given $\boldsymbol{s}^\star$),

$$\inf_{\hat{\boldsymbol{s}}}\sup_{\boldsymbol{s}\in\mathbb{R}^K}\mathbb{E}^{\hat{\boldsymbol{s}}}_{\boldsymbol{s}}\left[\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_p\right] \ge \inf_{\hat{\boldsymbol{s}}}\mathbb{E}^{\hat{\boldsymbol{s}}}_{\boldsymbol{s}\sim\mu}\left[\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_p\right],$$

and analogously for $\mathbb{E}^{\hat{\boldsymbol{s}}}_{\boldsymbol{s}\sim\mu}\left[\|\hat{\boldsymbol{s}}-\boldsymbol{s}\|_p^p\right]^{\frac{1}{p}}$. We will thus lower bound the RHS (Bayes error) with an appropriate choice of μ.

In particular, we consider an isotropic Gaussian prior: for a $\lambda > 0$,

$$\mu \coloneqq \mathcal{N}(\mathbf{0}, \lambda^2 \boldsymbol{I}_K).$$

Let us fix an arbitrary, adaptive algorithm $\mathcal{A}$ that interacts with the true environment $\boldsymbol{s}^\star$ sequentially as follows: at each timestep t, when the learner selects query k_t and judge j_t, they observe a noisy score $\tilde{s}_t \coloneqq s_{k_t} + \varepsilon_{k_t,j_t}$. Let us denote $N_{k,j} \coloneqq \sum_t \mathbb{1}[k_t = k, j_t = j]$ as the number of times $\mathcal{A}$ queries the pair (k, j) throughout the interaction. Suppose that $\mathcal{A}$ interacts for T rounds, while satisfying the budget constraint almost surely. For each time t, let $\mathcal{H}_t \coloneqq (k_1, j_1, \tilde{s}_1, \cdots, k_t, j_t, \tilde{s}_t)$ be the interaction history til time t. Then, at the end of the interaction, $\mathcal{A}$ outputs an estimator $\hat{\boldsymbol{s}} = \hat{\boldsymbol{s}}(\mathcal{H}_T)$. For clarity, with a slight abuse of notation, we will distinguish between the two: $\mathcal{A}$ as an adaptive data collection subprocedure, and $\hat{\boldsymbol{s}}$ as an aggregation subprocedure that takes $\mathcal{H}_T$ as an input.

Then, using the fact that everything is Gaussian, we can compute the posterior distribution of $\boldsymbol{s} \mid \mathcal{H}_T$ in closed-form, as shown in the following lemma whose proof is deferred to the end:

Lemma G.1. *$\boldsymbol{s} \mid \mathcal{H}_T \sim \mathcal{N}(\mathbb{E}[\boldsymbol{s} \mid \mathcal{H}_T], \mathcal{I}(\mathcal{A})^{-1})$, where $\mathcal{I}(\mathcal{A}) = \operatorname{diag}\left((\mathcal{I}_k(\mathcal{A}))_{k\in[K]}\right)$ is the (diagonal) posterior Fisher information of the algorithm $\mathcal{A}$ defined as*

$$\mathcal{I}_k(\mathcal{A}) = \lambda^{-2} + \sum_{j\in[J]} \frac{N_{k,j}}{\sigma^2_{k,j}}.$$

Denoting $\mathcal{H}_T(\mathcal{A})$ to be the random full interaction history for $\mathcal{A}$,

$$\inf_{\mathcal{A}} \inf_{\hat{\boldsymbol{s}}} \mathbb{E}^{\mathcal{A},\hat{\boldsymbol{s}}}_{\boldsymbol{s}\sim\mu}\left[f\left(\hat{\boldsymbol{s}}(\mathcal{H}_T(\mathcal{A})) - \boldsymbol{s}\right)\right] = \inf_{\mathcal{A}} \inf_{\hat{\boldsymbol{s}}} \mathbb{E}^{\mathcal{A}}\left[\underbrace{\mathbb{E}^{\hat{\boldsymbol{s}}}_{\boldsymbol{s}\mid\mathcal{H}_T(\mathcal{A})}\left[f\left(\hat{\boldsymbol{s}}(\mathcal{H}_T(\mathcal{A})) - \boldsymbol{s}\right) \mid \mathcal{H}_T(\mathcal{A})\right]}_{\textit{conditional posterior risk}}\right],$$

where either $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_p$ or $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_p^p$.

We now argue that the posterior mean minimizes the conditional posterior risk, i.e., when $\hat{\boldsymbol{s}}(\mathcal{H}_T) = \mathbb{E}[\boldsymbol{s} \mid \mathcal{H}_T]$. To see this, let $Z \coloneqq S - \mathbb{E}[\boldsymbol{s} \mid \mathcal{H}_T]$. Then, conditioned on $\mathcal{H}_T$, Z is centered Gaussian and symmetric. Also, both choices of f are convex and even ($f(\boldsymbol{u}) = f(-\boldsymbol{u})$).

For any estimator $\hat{\boldsymbol{s}}$, define its posterior risk as

$$L(\hat{\boldsymbol{s}}) \coloneqq \mathbb{E}[f(\hat{\boldsymbol{s}} - \boldsymbol{s}) \mid \mathcal{H}_T] = \mathbb{E}[f(\boldsymbol{u} - Z) \mid \mathcal{H}_T],$$

where $\boldsymbol{u} \coloneqq \hat{\boldsymbol{s}} - \mathbb{E}[\boldsymbol{s} \mid \mathcal{H}_T]$. Then as L is convex, by Jensen's midpoint inequality, we have that

$$L(0) \le \frac{L(\boldsymbol{u}) + L(-\boldsymbol{u})}{2} = L(\boldsymbol{u}),$$

i.e., the minimum is attained at $\boldsymbol{u} = 0$. This implies that the posterior mean $\mathbb{E}[\boldsymbol{s} \mid \mathcal{H}_T]$ minimizes the conditional posterior risk.[10]

Combining this with Lemma G.1, we have that

$$\begin{aligned}
\inf_{\mathcal{A}} \inf_{\hat{\boldsymbol{s}}} \mathbb{E}^{\mathcal{A},\hat{\boldsymbol{s}}}_{\boldsymbol{s}\sim\mu}\left[f\left(\hat{\boldsymbol{s}}(\mathcal{H}_T(\mathcal{A})) - \boldsymbol{s}\right)\right] &\ge \inf_{\mathcal{A}} \mathbb{E}^{\mathcal{A}}\left[\mathbb{E}_{\boldsymbol{s}\mid\mathcal{H}_T(\mathcal{A})}\left[f\left(\mathbb{E}[\boldsymbol{s} \mid \mathcal{H}_T(\mathcal{A})] - \boldsymbol{s}\right) \mid \mathcal{H}_T(\mathcal{A})\right]\right] \\
&= \inf_{\mathcal{A}} \mathbb{E}_{\boldsymbol{x}\sim\mathcal{N}(\mathbf{0},\mathcal{I}(\mathcal{A})^{-1})}\left[f(\boldsymbol{x})\right] \qquad \text{(Lemma G.1)} \\
&= \inf_{\mathcal{A}} \mathbb{E}_{\boldsymbol{g}\sim\mathcal{N}(\mathbf{0},\boldsymbol{I}_K)}\left[f\left(\mathcal{I}(\mathcal{A})^{-\frac{1}{2}} \odot \boldsymbol{g}\right)\right].
\end{aligned}$$

When $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_p$, by Jensen's inequality,[11]

$$\mathbb{E}_{\boldsymbol{g}\sim\mathcal{N}(\mathbf{0},\boldsymbol{I}_K)}\left[\left\|\mathcal{I}(\mathcal{A})^{-\frac{1}{2}} \odot \boldsymbol{g}\right\|_p\right] = \mathbb{E}_{\boldsymbol{g}\sim\mathcal{N}(\mathbf{0},\boldsymbol{I}_K)}\left[\left\|\mathcal{I}(\mathcal{A})^{-\frac{1}{2}} \odot |\boldsymbol{g}|\right\|_p\right]$$

[10] This resembles Anderson's lemma (Anderson, 1955), which states that a centered Gaussian measure assigns maximal mass to a symmetric convex set when the set is centered at the origin.

[11] $\boldsymbol{g} \mapsto \|\boldsymbol{u} \odot \boldsymbol{g}\|_p$ is convex over $\mathbb{R}^K_{>0}$, for any $\boldsymbol{u} \in \mathbb{R}^K_{>0}$.

$$\geq \left\| \mathcal{I}(\mathcal{A})^{-\frac{1}{2}} \odot \mathbb{E}_{\boldsymbol{g}\sim\mathcal{N}(\mathbf{0},\boldsymbol{I}_K)}[\boldsymbol{g}] \right\|_p = \sqrt{\frac{2}{\pi}} \left\| \mathcal{I}(\mathcal{A})^{-\frac{1}{2}} \right\|_p,$$

where $|\cdot|$ is applied element-wise.

When $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_p^p$.

$$\mathbb{E}_{\boldsymbol{g}\sim\mathcal{N}(\mathbf{0},\boldsymbol{I}_K)} \left[\left\| \mathcal{I}(\mathcal{A})^{-\frac{1}{2}} \odot \boldsymbol{g} \right\|_p^p \right]^{\frac{1}{p}} = \sum_{k\in[K]} \left(\mathcal{I}(\mathcal{A})^{-\frac{1}{2}} \right) \mathbb{E}_{g\sim\mathcal{N}(0,1)}[|g|^p]^{\frac{1}{p}} = m_p \left\| \mathcal{I}(\mathcal{A})^{-\frac{1}{2}} \right\|_p.$$

Either way, we arrive at $\left\| \mathcal{I}(\mathcal{A})^{-\frac{1}{2}} \right\|_p$, which we lower bound as follows. Let $N_{k,j}$ denote the total number of times pair (k,j) is queried by algorithm $\mathcal{A}$ subject to the budget constraint, and let us reparametrize it by $\omega_{k,j} := \frac{c_j N_{k,j}}{B}$. Then, we have that for each allocation $\{\omega_{k,j}\}_{k,j}$,

$$\left\| \mathcal{I}(\mathcal{A})^{-\frac{1}{2}} \right\|_p = \left(\sum_{k\in[K]} \left(\lambda^{-2} + \sum_{j\in[J]} \frac{B\omega_{k,j}}{c_j \sigma_{k,j}^2} \right)^{-p/2} \right)^{1/p}, \quad \forall \lambda > 0.$$

Taking the limit as $\lambda \to \infty$, we have that

$$\left\| \mathcal{I}(\mathcal{A})^{-\frac{1}{2}} \right\|_p = \frac{1}{\sqrt{B}} \left(\sum_{k\in[K]} \left(\sum_{j\in[J]} \frac{\omega_{k,j}}{c_j \sigma_{k,j}^2} \right)^{-p/2} \right)^{1/p} = \sqrt{\frac{\mathcal{A}_p(\boldsymbol{\omega};\boldsymbol{\sigma},\boldsymbol{c})}{B}}.$$

Taking the infimum over $\mathcal{A}$, which is equivalent to taking the infimum over $\boldsymbol{\omega}$ under the budget constraint, we conclude with Theorem 3.2. □

Remark 2. *One could also utilize Gaussian hypercontractivity (specifically, Kahane's inequality) to transform a lower bound on* $\mathbb{E}\left[\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_p^p\right]^{1/p}$ *into a lower bound on* $\mathbb{E}\left[\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_p\right]$*, up to some Gaussian moment-dependent constants. But this leads to a bit more suboptimal constant.*

Proof of Lemma G.1. This follows from directly factorizing the likelihood and completing the square. Denoting $\pi_t(\cdot,\cdot)$ as the sampling kernel of the algorithm $\mathcal{A}$, we can drop terms independent of $\boldsymbol{s}$ by absorbing them into a proportionality constant. The likelihood is:

$$p(\mathcal{H}_T \mid \boldsymbol{s}) = \prod_{t=1}^{T} \left\{ \pi_t((k_t,j_t) \mid \mathcal{H}_{t-1}) \frac{1}{\sqrt{2\pi}\sigma_{k_t,j_t}} \exp\left(-\frac{(\tilde{s}_t - s_{k_t})^2}{2\sigma_{k_t,j_t}^2} \right) \right\}$$
$$\propto \exp\left(-\frac{1}{2} \sum_{t=1}^{T} \frac{(\tilde{s}_t - s_{k_t})^2}{\sigma_{k_t,j_t}^2} \right).$$

By grouping the observations by (k,j), and using the empirical mean $\bar{s}_{k,j} := \frac{1}{N_{k,j}} \sum_{t=1}^{T} \mathbb{1}[k_t = k, j_t = j]\tilde{s}_t$, the sum of squares decomposes as:

$$\sum_{t=1}^{T} \frac{(\tilde{s}_t - s_{k_t})^2}{\sigma_{k_t,j_t}^2} = \sum_{k,j} \frac{1}{\sigma_{k,j}^2} \left(\sum_{t:k_t=k,j_t=j} (\tilde{s}_t - \bar{s}_{k,j} + \bar{s}_{k,j} - s_k)^2 \right).$$

Since the cross-terms sum to zero ($\sum(\tilde{s}_t - \bar{s}_{k,j}) = 0$), this simplifies to:

$$\sum_{k,j} \frac{1}{\sigma_{k,j}^2} \left(\sum_{t:k_t=k,j_t=j} (\tilde{s}_t - \bar{s}_{k,j})^2 + N_{k,j}(\bar{s}_{k,j} - s_k)^2 \right).$$

Absorbing the terms independent of s_k into the proportionality constant, we have:

$$p(\mathcal{H}_T \mid \boldsymbol{s}) \propto \exp\left(-\frac{1}{2} \sum_{k\in[K]} \sum_{j\in[J]} \frac{N_{k,j}}{\sigma_{k,j}^2} (s_k - \bar{s}_{k,j})^2 \right).$$

Now, applying Bayes' rule with the prior $\boldsymbol{s} \sim \mu = \mathcal{N}(0, \lambda^2 \boldsymbol{I}_K)$, the posterior is:

$$\begin{aligned} p(\boldsymbol{s} \mid \mathcal{H}_T) &\propto p(\mathcal{H}_T \mid \boldsymbol{s})\mu(\boldsymbol{s}) \\ &\propto \exp\left(-\frac{1}{2}\sum_{k,j}\frac{N_{k,j}}{\sigma_{k,j}^2}(s_k - \bar{s}_{k,j})^2\right)\exp\left(-\frac{1}{2}\sum_{k\in[K]}\frac{s_k^2}{\lambda^2}\right) \\ &= \prod_{k\in[K]}\exp\left(-\frac{1}{2}\left[\frac{s_k^2}{\lambda^2} + \sum_{j\in[J]}\frac{N_{k,j}}{\sigma_{k,j}^2}(s_k - \bar{s}_{k,j})^2\right]\right). \end{aligned}$$

Focusing on the exponent for each k, we expand and collect the s_k terms:

$$\begin{aligned} \frac{s_k^2}{\lambda^2} + \sum_{j\in[J]}\frac{N_{k,j}}{\sigma_{k,j}^2}(s_k^2 - 2s_k\bar{s}_{k,j} + \bar{s}_{k,j}^2) &= s_k^2\underbrace{\left(\frac{1}{\lambda^2} + \sum_{j\in[J]}\frac{N_{k,j}}{\sigma_{k,j}^2}\right)}_{=\mathcal{I}_k(\mathcal{A})} - 2s_k\sum_{j\in[J]}\frac{N_{k,j}}{\sigma_{k,j}^2}\bar{s}_{k,j} + C_k \\ &= \mathcal{I}_k(\mathcal{A})(s_k - \mu_k)^2 + C_k', \end{aligned}$$

where C_k and C_k' are terms independent of s_k, and $\mu_k \coloneqq \mathcal{I}_k(\mathcal{A})^{-1}\sum_j \frac{N_{k,j}}{\sigma_{k,j}^2}\bar{s}_{k,j}$. Thus, $p(\boldsymbol{s} \mid \mathcal{H}_T) \propto \prod_k \exp\left(-\frac{\mathcal{I}_k(\mathcal{A})}{2}(s_k - \mu_k)^2\right)$, which is exactly the density of a Gaussian distribution with mean $\boldsymbol{\mu}$ and a diagonal precision matrix whose k-th diagonal entry is $\mathcal{I}_k(\mathcal{A})$. □

G.3 Proof of Theorem F.8: Global In-Expectation Lower Bound for Gaussian Scores via van Trees Inequality

Our key technical tool is a generalization of van Trees inequality (Gill and Levit, 1995; van Trees, 1968) to L_p-norm due to Chen and Özgür (2024) via Efroimovich's inequality (Efroimovich, 1979):

Lemma G.2 (Scalar Version of Theorem 2.4 of Chen and Özgür (2024)). *Let $1 \le p < \infty$ and define $C_{\mathrm{ME}}(p) \triangleq 2e^{\frac{1}{p}}\Gamma(p^{-1})p^{\frac{1}{p}-1}$ to be the partition function of the L_p max-entropy distribution. For a closed interval $\Theta \subseteq \mathbb{R}$, suppose that we have a parametric family of distributions $\{\mathbb{P}_\theta\}_{\theta\in\Theta}$ with density $f(x;\theta)$, and let μ be a prior density over Θ that is absolutely continuous, **vanishes at the boundary**, and has finite Fisher information.*
*Then, for any estimator $\hat{\theta} : X \mapsto \hat{\theta}(X) \in \mathbb{R}$ where $X \sim P_\theta$, the **expected p-th moment risk** is lower-bounded as follows:*

$$\mathbb{E}_{\theta\sim\mu}\mathbb{E}_{\hat{\theta},X}\left[\left|\theta - \hat{\theta}(X)\right|^p\right] \ge \left(\frac{\sqrt{2\pi e}}{C_{\mathrm{ME}}(p)}\right)^p \left(\mathbb{E}_{\theta\sim\mu}[\mathcal{I}(\theta)] + \mathcal{J}(\mu)\right)^{-\frac{p}{2}},$$

*where the **Fisher information** terms are defined as*

$$\mathcal{I}(\theta) \coloneqq \mathbb{E}_X\left[-\frac{\partial^2}{\partial\theta^2}\log p(X \mid \theta)\right], \quad \mathcal{J}(\mu) \coloneqq \mathbb{E}_{\theta\sim\mu}\left[-\frac{\partial^2}{\partial\theta^2}\log\mu(\theta)\right].$$

Especially note that when $p = 2$, $C_{\mathrm{ME}}(2) = \sqrt{2\pi e}$, and the constant multiple becomes 1, as in the original van Trees inequality (Gill and Levit, 1995; van Trees, 1968).

And again, we begin By Yao's Minimax Principle (Yao, 1977): for *any* decomposable prior $\mu = \bigotimes_{k\in[K]}\mu_k$ over $\boldsymbol{s}$,

$$\begin{aligned} \inf_{\mathcal{A}}\sup_{\boldsymbol{s}}\mathbb{E}_{\boldsymbol{s}}^{\mathcal{A}}\left[\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_p^p\right] &\ge \inf_{\mathcal{A}}\mathbb{E}_{\boldsymbol{s}\sim\mu}\left[\mathbb{E}_{\boldsymbol{s}}^{\mathcal{A}}\left[\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_p^p\right]\right] \\ &= \inf_{\mathcal{A}}\sum_k \mathbb{E}_{s_k\sim\mu_k}\left[\mathbb{E}_{s_k}^{\mathcal{A}}\left[|\hat{s}_k - s_k|^p\right]\right], \end{aligned}$$

where the algorithm $\mathcal{A}$ must satisfy the knapsack constraint: if $\mathcal{A}$ samples each (k,j) pair $N_{k,j}$ times, then $\sum_j c_j \sum_k N_{k,j} \le B$. Let us fix an arbitrary algorithm $\mathcal{A}$, and the corresponding numbers of pulls $N_{k,j}$'s.

As first suggested in Borovkov (1998); Tsybakov (2009), we choose μ_k as the *cosine-squared* prior density, indexed by $r > 0$:[12]

$$\mu_k^{(r)}(s) = \frac{1}{r}\cos^2\left(\frac{\pi(s-s_k)}{2r}\right)\mathbb{I}_{[s_k-r,s_k+r]}(s).$$

Its prior Fisher information $\mathcal{J}(\mu_k^{(r)}) \triangleq \mathbb{E}_{s\sim\mu_k^{(r)}}\left[-\frac{\partial^2}{\partial s^2}\log\mu_k^{(r)}(s)\right]$ is computed as:

$$\begin{aligned}
\mathcal{J}(\mu_k^{(r)}) &= \int_{s_k-r}^{s_k+r}\mu_k^{(r)}(s)\left(-\frac{\partial^2}{\partial s^2}\left[\log\left(\frac{1}{r}\right)+2\log\cos\left(\frac{\pi(s-s_k)}{2r}\right)\right]\right)ds \\
&= \int_{-r}^{r}\frac{1}{r}\cos^2\left(\frac{\pi s}{2r}\right)\left(\frac{\pi^2}{2r^2}\sec^2\left(\frac{\pi s}{2r}\right)\right)ds \\
&= \int_{-r}^{r}\frac{\pi^2}{2r^3}ds = \frac{\pi^2}{r^2}.
\end{aligned}$$

The Fisher information of the Gaussian observations and algorithm $\mathcal{A}$ is computed as follows: denoting $p(\cdot|s,\sigma^2)$ as the probability density of $\mathcal{N}(s,\sigma^2)$,

$$\begin{aligned}
\mathbb{E}_{s_k\sim\mu_k^{(r)}}\left[\mathcal{I}(s_k)\right] &= \mathbb{E}^{\mathcal{A}}_{s_k\sim\mu_k^{(r)}}\left[-\frac{\partial^2}{\partial s_k^2}\left[\sum_{t=1}^{T}\mathbb{1}[k_t=k]\log p(\tilde{s}_t \mid s_k,\sigma^2_{k_t,j_t})\right]\right] \\
&= \mathbb{E}^{\mathcal{A}}_{s_k\sim\mu_k^{(r)}}\left[-\frac{\partial^2}{\partial s_k^2}\left[\sum_{j\in[J]}\sum_{t=1}^{T}\mathbb{1}[k_t=k,j_t=j]\left(-\frac{(\tilde{s}_t-s_k)^2}{2\sigma^2_{k,j}}+C\right)\right]\right] \\
&= \mathbb{E}^{\mathcal{A}}\left[\sum_{j\in[J]}\sum_{t=1}^{T}\mathbb{1}[k_t=k,j_t=j]\frac{1}{\sigma^2_{k,j}}\right] \\
&= \mathbb{E}^{\mathcal{A}}\left[\sum_{j\in[J]}\frac{N_{k,j}}{\sigma^2_{k,j}}\right] = \sum_{j\in[J]}\frac{\mathbb{E}^{\mathcal{A}}[N_{k,j}]}{\sigma^2_{k,j}}.
\end{aligned}$$

With this, we apply the L_p***-van Trees inequality*** (Lemma G.2) for each k, which gives

$$\sum_k \mathbb{E}_{s_k\sim\mu_k^{(r)}}\left[\mathbb{E}^{\mathcal{A}}_{s_k}\left[|\hat{s}_k-s_k|^p\right]\right] \geq \left(\frac{\sqrt{2\pi e}}{C_{\mathrm{ME}}(p)}\right)^p \sum_{k\in[K]}\left(\sum_{j\in[J]}\frac{\mathbb{E}^{\mathcal{A}}[N_{k,j}]}{\sigma^2_{k,j}}+\frac{\pi^2}{r^2}\right)^{-\frac{p}{2}}.$$

Chaining everything, we then have that

$$\sup_{\boldsymbol{s}}\mathbb{E}^{\mathcal{A}}_{\boldsymbol{s}}\left[\|\hat{\boldsymbol{s}}-\boldsymbol{s}\|_p^p\right] \geq \left(\frac{\sqrt{2\pi e}}{C_{\mathrm{ME}}(p)}\right)^p \sum_{k\in[K]}\left(\sum_{j\in[J]}\frac{B\omega_{k,j}}{c_j\sigma^2_{k,j}}+\frac{\pi^2}{r^2}\right)^{-\frac{p}{2}},$$

where as in the proof of Theorem 5.1, we define $\omega_{k,j} = \frac{\mathbb{E}^{\mathcal{A}}[N_{k,j}]c_j}{B}$.

As this holds for any $r > 0$, taking the limit on both sides as $r \to \infty$ and $1/p$-th power, we have

$$\sup_{\boldsymbol{s}}\mathbb{E}^{\mathcal{A}}_{\boldsymbol{s}}\left[\|\hat{\boldsymbol{s}}-\boldsymbol{s}\|_p^p\right]^{\frac{1}{p}} \geq \frac{\sqrt{2\pi e}}{C_{\mathrm{ME}}(p)}\left(\sum_{k\in[K]}\left(\sum_{j\in[J]}\frac{B\omega_{k,j}}{c_j\sigma^2_{k,j}}\right)^{-\frac{p}{2}}\right)^{\frac{1}{p}}.$$

Taking the infimum over $\mathcal{A}$ (and thus over $\boldsymbol{\omega}$) and invoking Theorem 3.2, we are done. □

[12] One can verify using calculus of variation that this prior density is the "optimal" choice that minimizes $\mathcal{J}(\mu)$ among all continuously differentiable densities supported on $[-r,r]$ that vanish at the boundaries.

H Additional Experiments

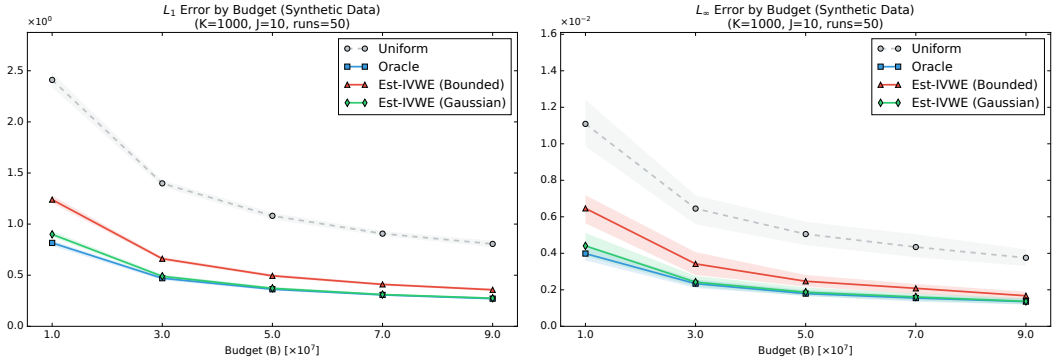
This appendix provides additional experimental results that extend the analysis presented in Section 7. We provide further empirical evidence regarding the optimality of our algorithms across different norms and datasets.

H.1 Details and Additional Results in Experiments on Synthetic Data

For each query $k \in [K]$, the ground-truth score s_k is sampled uniformly from the interval $[0.1, 0.9]$, then the true variance $\sigma^2_{k,j}$ for each judge j on query k is sampled uniformly from $[10^{-4}, 0.9 s_k(1 - s_k)]$. The upper bound of the interval is set to 90% of the maximum possible variance for a bounded distribution on $[0, 1]$ with a mean s_k. Finally, the cost of each judge is sampled uniformly from the interval $[0.5, 1.5]$.

We use $p \in \{1, 2, \infty\}$, budgets $B \in \{10^7, 3 \times 10^7, 5 \times 10^7, 7 \times 10^7, 9 \times 10^7\}$, and 50 independent runs. All experiments use random seed 0.

In Figure 2, we present the ℓ_1 and ℓ_∞-errors of the synthetic experiments described in Section 7 ($K = 1000, J = 10$). Consistent with the ℓ_2-error in the main text, both EST-IVWE (Bounded) and EST-IVWE (Gaussian) show a clear convergence toward the ORACLE as the budget B increases. This confirms that our adaptive allocation strategy effectively minimizes the ℓ_p-error for any p, outperforming the non-adaptive UNIFORM strategy.



(a) ℓ_1-Error. (b) ℓ_∞-Error.

Figure 2: Additional Experimental results on synthetic datasets. All results are averaged over 50 independent runs, with the 10% and 90% quantiles (shaded regions)

H.2 Details and Additional Results in Experiments on Real-data

Our preprocessing pipeline establishes a reliable baseline by first performing consensus filtering. We select only those queries where all three judges provide an identical mode score via majority voting. This ensures that the established s_k serves as a credible proxy for the true score. Furthermore, we ensure statistical significance in variance estimation by retaining query-judge pairs that have at least 25 evaluation samples for all judges. Following this filtering, we obtain a subset of queries for four distinct categories: Complexity, Correctness, Helpfulness, and Verbosity. Heteroskedasticity is modeled by empirically calculating the true variance $\sigma^2_{k,j}$ as the mean squared deviation from the consensus score s_k.

We use budgets from 200 to $5,000$ in increments of 400, $p \in \{1, 2, \infty\}$, uniform judge costs of 0.1, 50 independent runs, and random seed 0. Judge evaluations are simulated by sampling with replacement from the corresponding evaluation pools.

We provide the complete experimental results for all four datasets (Complexity, Correctness, Helpfulness, and Verbosity) preprocessed as described above. Figure 3 illustrates the ℓ_1, ℓ_2, and ℓ_∞ errors for each dataset. Across all twelve scenarios, the EST-IVWE variants consistently show the relative

superiority over the UNIFORM baseline, performing as well as the ORACLE strategy as the budget B increases.

H.3 Sensitivity Analysis on Cost-Quality Correlations

To evaluate the robustness of our framework beyond the uniform cost setting employed in Section 7, we investigate how varying correlations between a judge's cost and their evaluation accuracy (inverse variance) affect convergence behavior. Theoretically, the selection of an optimal judge is governed by the cost-weighted variance, $j^*(k) = \arg\min_{j\in[J]} c_j\sigma_{k,j}^2$, which balances the trade-off between per-sample information gain and budget consumption. We conducted a series of ablation studies using ORACLE and EST-IVWE (Bounded) for $p \in \{1, 2, \infty\}$ under three cost structures: (i) *Bad is Expensive*, where variance and cost are positively correlated; (ii) *Bad is Cheap*, where low-quality judges are priced cheaper; and (iii) *Uniform Costs*, where the costs are all the same. Precisely, for a fixed query k, the variance $\sigma_{k,j}^2$ for each judge $j \in [J]$ is sampled from the interval $[j, J + j - 1] \times s_k(1 - s_k)/(2J)$ as the larger the index j, the larger its variance $\sigma_{k,j}^2$ tends to be. In *Bad is Expensive* scenario, the cost is set to be proportional to $\sum_{k\in[K]} \sigma_{k,j}^2$; while in *Bad is Cheap* scenario, the cost is set to be proportional to $\sum_{k\in[K]} \sigma_{k,j}^{-2}$. In both cases, the costs are normalized so that the mean cost over J judges becomes 0.1.

We use $K = 1000$, $J = 10$, $p \in \{1, 2, \infty\}$, budgets $B \in \{10^6, 3 \times 10^6, 5 \times 10^6, 7 \times 10^6, 9 \times 10^6\}$, 25 independent runs, and random seed 0.

As illustrated in Figure 4, our empirical findings suggest that while the specific cost structure influences the constant factors of the convergence rate, it does not alter the fundamental asymptotic decay of the estimation error. The *Bad is Cheap* scenario represents the most challenging regime for adaptive allocation; when a judge's high variance is offset by a significantly lower cost, the cost-weighted variances $\{c_j\sigma_{k,j}^2\}_{j\in[J]}$ across different judges become numerically similar. This reduced "gap" between the optimal judge and noisier alternatives necessitates a more extensive initial exploration (N_0) to ensure correct identification. Nevertheless, the results confirm that the algorithm remains robust across diverse cost scenarios, consistently achieving error reduction without failing to identify the optimal judges as the budget scales.

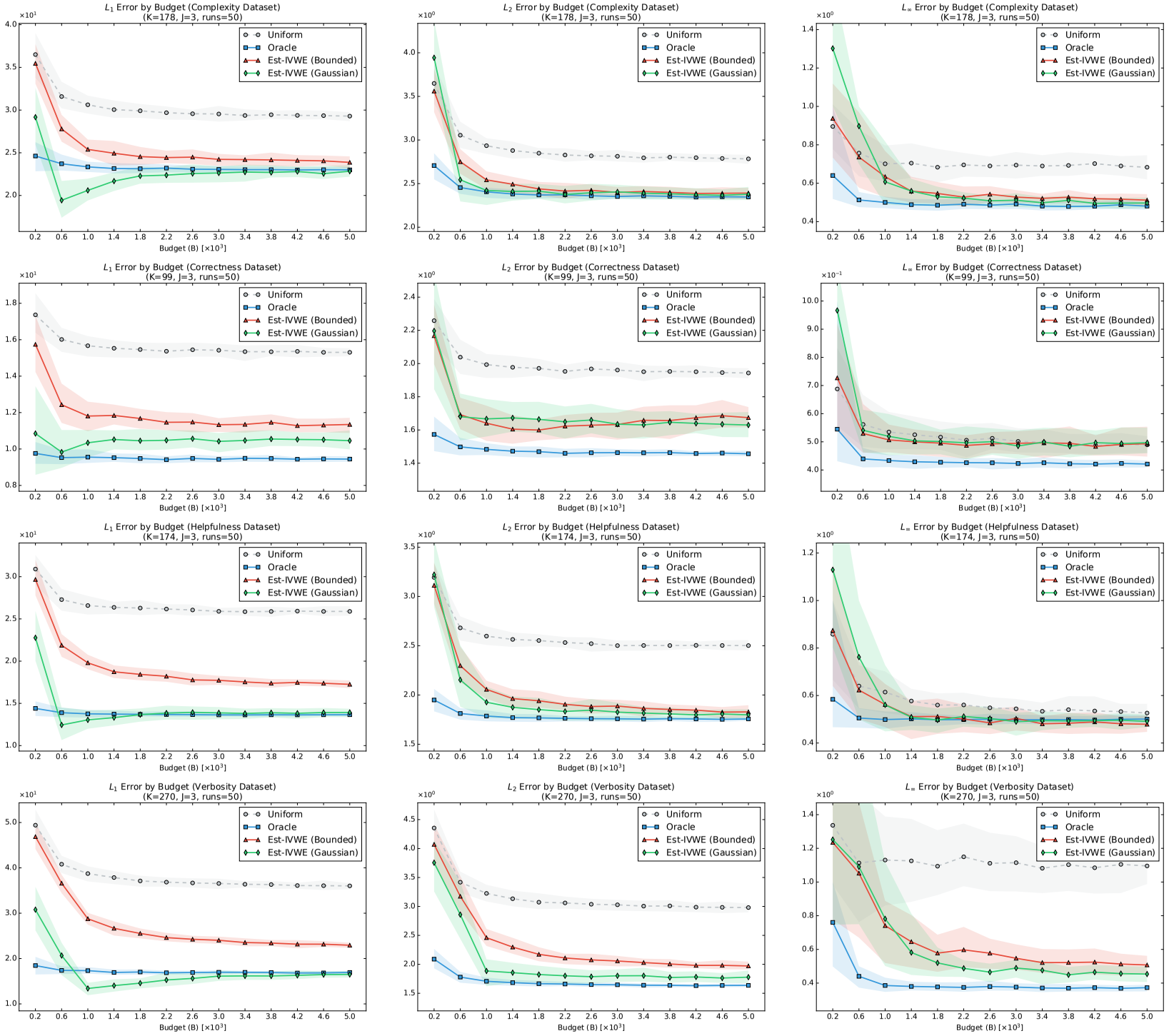


Figure 3: Full experimental results for the datasets (Complexity, Correctness, Helpfulness, Verbosity). Each column corresponds to an error metric ($\ell_1, \ell_2, \ell_\infty$).

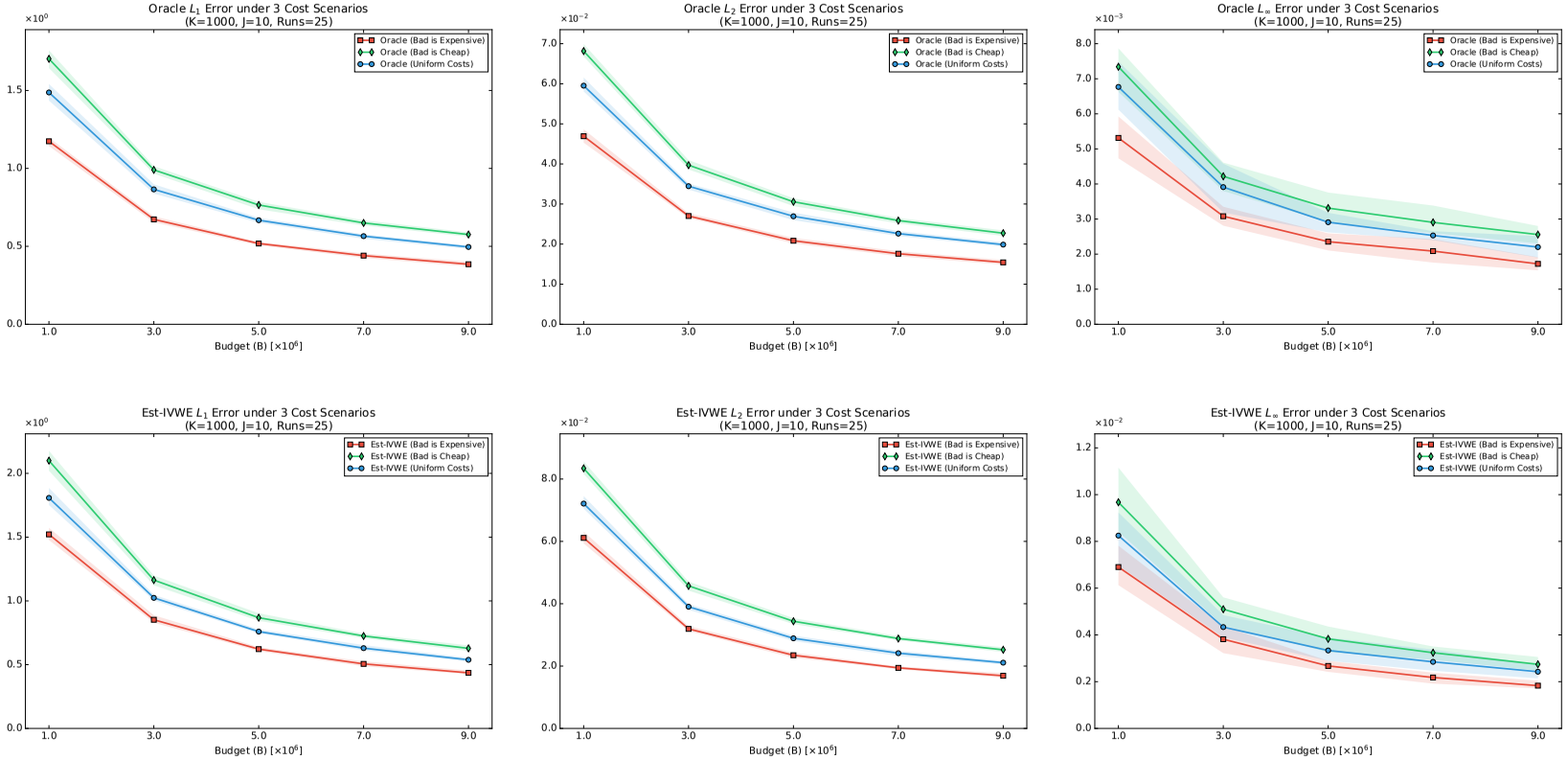


Figure 4: Impact of Cost Structure on Performances of **Est-IVWE** and **Oracle**

I Limitations and Future Directions

Here, we elaborate on the limitations and suggest some future directions for addressing them.

Query-independent costs. Our analyses rely on the assumption that querying to judge j incurs a deterministic cost c_j regardless of the query k. One way to approximately enforce this assumption is to impose fixed input and output token lengths for each judge j, and define c_j as the corresponding worst-case cost under theses limits. This recovers our deterministic cost setting, although the realized cost may be smaller when generation terminates early before reaching the limit.

A more realistic model would allow the cost $c_{k,j}$ of allocating query k to judge j to depend on both the query and the generated output. In such a setting, the costs may be *stochastic* and revealed in an online fashion during the generation. For example, under per-token API pricing, the input-side cost is known in advance, whereas the output-side cost depends on the stochastic generation length. This turns the hard deterministic budget constraint considered in our work into a stochastic, knapsack-type constraint. This setting is related to online learning and allocation under budget constraints (Balseiro et al., 2023; Slivkins et al., 2024), where primal-dual algorithms attain $O(\sqrt{T})$ regret.

However, stochastic costs make it non-trivial to adapt the usual fixed-budget best arm identification-type approaches, and the dual algorithms do not transfer directly either, since our objective is estimation rather than regret minimization. One potential direction is to *reduce* the budget-constrained estimation problem to an online learning problem (see, e.g., Clerico et al. (2025); Lee et al. (2024); Rakhlin and Sridharan (2017) for related reduction principle) and then combine the resulting formulation with primal–dual allocation methods. Whether such a reduction can preserves the instance-wise allocation optimality identified in this work remains an open question.

Unbiasedness. Some judge biases can be corrected on the fly: swapping the order of responses counters position bias (Zheng et al., 2023), and length-controlled win rates post-hoc debias verbosity (Dubois et al., 2024). Statistically, however, if each judge carries a non-vanishing residual bias, then without further information (e.g., the density ratio between the biased and unbiased judge distributions (Clémençon and Laforgue, 2022)), the true score is no longer identifiable, and the problem may not be solvable in general.

One natural structure to impose is that more powerful models have means closer to the "true" score *and* lower variance. Our framework currently captures only the latter, through the heteroskedastic variance profile. This kind of cost–accuracy trade-off is known as *multi-fidelity* structure (Peherstorfer et al., 2018; Poiani et al., 2024): one has access to M simulators, with costlier simulators providing more accurate simulations, and this literature develops techniques for minimizing cumulative querying cost while retaining a target notion of optimality. Adapting these approaches to our estimation setting is a promising and challenging direction.

Cross-query structure. Our framework treats the scores and variances across queries and judges as uncorrelated. Consequently, an observation from one query provides no information about another, and EST-IVWE estimates the variance of every query–judge pair separately through forced exploration. In practice, however, related queries may share features or latent structure. For example, if scores or variances are governed by a shared low-dimensional model based on query features, observations from one query could provide information about many others and substantially reduce the required exploration. Moreover, dependencies may arise among evaluations themselves; for example, Balasubramanian et al. (2026) model statistical dependencies among LLM judges using an Ising model. Exploiting such shared structure and dependence would require replacing our pairwise variance-based formulation with a joint statistical model, and developing the corresponding budget-optimal allocation remains an interesting open direction.

Randomness of repeated evaluations. Our analysis assumes that repeated evaluations of each query–judge pair are independent draws from a stationary distribution with mean s_k and variance $\sigma^2_{k,j}$. This assumption is natural only when judges are queried with a positive temperature, as is sometimes done to obtain stochastic judgments and characterize evaluation uncertainty. However, when LLM judges are queried at temperature 0, repeated evaluations can be identical or nearly identical. In such settings, repeated sampling provides little or no information for estimating the variance $\sigma^2_{k,j}$, which is essential to our allocation strategy.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: Our main claims in the abstract and introduction are about IVWE, EST-IVWE, and lower bounds; and they are in 3, 4 and 5. Additionally, experimental results are given in 7 and H

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: It is presented in Section 8 along with future work.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: All theoretical results clearly state their assumptions, and complete proofs are provided in Appendices A to E and G

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: All experimental setups, datasets (Synthetic and HelpSteer2), preprocessing procedures, evaluation metrics, random seed, software dependencies, and execution commands are fully detailed in Section 7, Appendix H, and our GitHub repository.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: The code and reproduction instructions are publicly available. However, the repeated LLM-judge evaluations are third-party assets provided by Saha et al. (2026) and cannot be redistributed. Their format and generation procedure are documented in their GitHub repository.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Data generation processes and preprocessing steps for the HelpSteer2 dataset are clearly specified in Appendix H.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: All experimental results are averaged over 50 independent runs, and 10% and 90% quantiles are reported as shaded regions in Figure 1.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [No]

   Justification: The experiments involve lightweight simulations and evaluations on static datasets, which do not require specialized or large-scale compute resources.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: This research is conducted, in every respect, with NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: This work focuses on fundamental theoretical resource allocation and statistical estimation; we do not foresee direct societal impacts.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.

- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: This paper does not release any data or models.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: We utilize a processed version of the HelpSteer2 dataset (`https://github.com/NVIDIA/NeMo-Aligner/blob/main/LICENSE`) as provided by Saha et al. (2026), and we have properly cited both the original dataset and the authors of the processed version in Section 7.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.

- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: This paper does not release new essetss.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: This paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: This paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.

- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: The core method development in this research does not involve LLMs.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.