Instance-Optimal Estimation with Multiple LLM Judges on a Budget

Junghyun Lee [*1] **Sanghwa Kim** [*1] **Yassir Jedra** [2] **Alexandre Proutière** [3] **Se-Young Yun** [1]

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*. Given K prompt–response pairs, J judges with known costs, and unknown query–judge variances, the goal is to estimate a 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 error rate. Since this allocation depends on the unknown variances, we then address the practical unknown-variance setting by proposing EST-IVWE, 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. Our second and central theoretical contribution is a matching *local* minimax lower bound, which establishes the instance-optimality of the proposed algorithms. A key technical insight is that Fano-type high-probability arguments are too coarse for this problem: their packing construction loses the local variance structure that governs the optimal allocation. We instead use an Assouad-type in-expectation argument, based on local perturbations, which preserves this structure and yields the sharp allocation-dependent lower bound. Finally, we numerically validate the superiority of our approach over naïve uniform allocation on synthetic and HelpSteer2 datasets.

[1]Kim Jaechul Graduate School of Artificial Intelligence, KAIST, Daejeon, South Korea [2]Department of Electrical and Electronic Engineering, Imperial College London, London, UK [3]School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden. Correspondence to: Junghyun Lee <jh_lee00@kaist.ac.kr>, Sanghwa Kim <tkdghk9667@kaist.ac.kr>, Se-Young Yun <yunseyoung@kaist.ac.kr>.

Accepted to the 1st Workshop on Combining Theory and Benchmarks, CTB@ICML 2026, Seoul, South Korea. Copyright 2026 by the author(s).

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 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, by contrast, 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 (prompt-answer) queries $k \in [K] := \{1, \ldots, K\}$, and a set of available 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 $\mathbf{s} = (s_k)_{k\in[K]} \in \mathbb{R}^K$ 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), where the authors only considered $p = \infty$. 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 a *local* minimax lower bound, matching the performance guarantees of EST-IVWE, hence establishing its instance-optimality (Section 5). A key technical insight is that Fano-type high-probability arguments are too coarse for this problem: their packing construction loses the local variance structure that governs the optimal allocation. We instead use an Assouad-type in-expectation argument, based on local perturbations, which preserves this structure and yields the sharp allocation-dependent lower bound.

(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 & Deal, 1959; Norwood Jr & 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 & 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., question–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 random estimate $s_{k,j} \in [0, R]$. 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 $\mathbf{c} = (c_j)_{j\in[J]} \in \mathbb{R}^J_{>0}$, her goal is to provide an estimator of the score vector $\mathbf{s} = (s_k)_{k\in[K]} \in \mathbb{R}^K$ as accurate as possible, i.e., minimizing the ℓ_p-error, for a fixed $p \in [1, \infty]$ specified in advance.

The boundedness assumption 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 the (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 $\sigma_{k,j}^2 \leq R^2/4$ 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), $\sigma_{k,j}^2 \ll R^2/4$. 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 & Krause, 2018). With this, each problem instance is characterized by a triple $(\boldsymbol{s}, \boldsymbol{\sigma}, \boldsymbol{c}) \in \mathbb{R}^K \times \mathbb{R}_{>0}^{K\times J} \times \mathbb{R}_{>0}^J$, 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:

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

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

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 $\sigma_{k,j}^2$.

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:

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

$$\hat{s}_{k,j} := \frac{1}{N_{k,j}} \sum_{i=1}^{N_{k,j}} s_{k,j}^{(i)}, \ \forall j \in [J],$$

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:*

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

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

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, the optimal allocation is *sparse*: for each query k, the budget is assigned to a single optimal judge $j^*(k)$. This sparsity follows from the KKT conditions, analyzed in Appendix A.2, which show that an optimal allocation assigns budget only to judges 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 suboptimal judge $j \neq j^*(k)$ can only worsen, or at best leave unchanged, the resulting estimate. Thus, while the choice of p changes how budget is distributed across queries, the optimal judge selection within each query remains sparse.

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

Next, we consider the realistic setting in which the variances are unknown. We propose EST-IVWE, a simple adaptive algorithm whose performance approaches that of IVWE combined with the oracle allocation. The algorithm proceeds in two phases: the first estimates the variances, while the second allocates the remaining budget according to the empirical optimal allocation induced by these estimates. The pseudocode 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 $\boldsymbol{c} \in \mathbb{R}_+^J$,
Number of forced exploration per pair $N_0 > 0$,
Bias term for variance estimator $\tau > 0$

`// 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

$$\hat{\sigma}_{k,j}^2 := \frac{1}{2N_0(N_0-1)} \sum_{1\le n\ne n'\le N_0} \left(s_{k,j}^{(n)} - s_{k,j}^{(n')}\right)^2$$

`// Phase II: IVWE`

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

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 $(1-\alpha)\sigma_{k,j}^2 \le \hat{\sigma}_{k,j}^2 \le (1+\alpha)\sigma_{k,j}^2$ for some $\alpha \in (0, 1)$. Such a bound is highly desirable because, as seen in the form of the optimal allocation and error bounds (Theorem 3.2), the common error factors $(1 \pm \alpha)$ can be factored out, which provides a direct link between empirical and oracle allocations. While (Saha et al., 2026, Lemma 6) have successfully leveraged this property by assuming that the scores are Gaussian, where multiplicative bounds are attained by (Laurent & Massart, 2000, Eqn. (4.4) & (4.5))[1], this assumption is not applicable to our setting where scores are bounded. Instead, we utilize the empirical Bernstein inequality (Maurer & 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* $\bar{\sigma}_{k,j} = \hat{\sigma}_{k,j} + \tau$, which satisfies $\sigma_{k,j} \le \bar{\sigma}_{k,j} \le \sigma_{k,j} + 2\tau$ 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 $\tau = R\sqrt{\frac{2}{N_0-1}\log\frac{4KJ}{\delta}}$, $N_0 = (2B)^{\frac{1}{3}}\left(R^2 \log\frac{4KJ}{\delta}\right)^{\frac{2}{3}}$ when $p \ge 2$, and $N_0 = 2^{\frac{p}{2p+2}} R^{\frac{3p+2}{2p+2}} \left(\log\frac{4KJ}{\delta}\right)^{\frac{3p+2}{4p+4}} B^{\frac{p+2}{4p+4}}$ otherwise. Further suppose that $B \ge 2N_0\left(K\sum_{j\in[J]} c_j\right)$. 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:*

$$\varepsilon_p = \sqrt{\frac{2\mathcal{A}_p^*(\boldsymbol{\sigma}, \boldsymbol{c}) \log\frac{4K}{\delta}}{B}} + E_p(B),$$

where $E_p(B)$ is the lower-order error term defined as:

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

[1] This heavily relies on the Gaussianity assumption, as its derivation is based on the fact that squaring Gaussian random variables results in a χ^2-distribution.

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 & 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 EST-IVWE (Algorithm 1) matches the leading term of the oracle strategy that knows the variances in advance (Theorem 3.2), for all $p \in [1, \infty]$. Thus, to leading order, there is no price for not knowing the variances: a simple two-stage procedure suffices to recover the oracle rate.

(2) Relations to Graybill-Deal Estimator. Our problem of estimating s_k from multiple sources can be viewed as an instance of the *common mean problem*, a classical problem in statistics (Graybill & Deal, 1959; Pal et al., 2007). Our two-stage estimator is closely related to the *Graybill–Deal* estimator, a standard approach for common mean estimation, which uses all available samples to estimate both the empirical means and variances. However, because the same samples are reused for both quantities, the resulting empirical means and variance estimates are statistically correlated, making sharp performance analysis difficult. Indeed, even for Gaussian populations (Nair, 1980; Voinov, 1984), tight characterizations of the estimator's variance are technically involved, and admissibility is known only for certain classes of problems (Ghosh & Sinha, 1981; Pal & Lim, 1997; Sinha & Mouqadem, 1982). We avoid this dependence through *sample splitting*: the samples used to estimate the variances $\hat{\sigma}^2_{k,j}$ in Phase I are not reused to compute the final IVWE estimates $\hat{s}_k$ in Phase II. This decouples the randomness of the variance estimates from that of the mean estimates and yields a key property for the analysis: 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 & 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 different techniques yield different levels of sharpness. Fano-type high-probability arguments are too coarse to recover the optimal allocation dependence, whereas an Assouad-type in-expectation analysis preserves the local variance structure and yields the matching lower bound.

5.1. A High-Probability Lower Bound and Its Limitation

We begin by stating a high-probability minimax lower bound:

Theorem 5.1 (High Probability Lower Bound). *Let $p \in [1, \infty]$, $B > 0$, $\delta \in (0, e^{-2}]$, 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 $\max_j \sigma^2_{k,j} \leq \frac{R_k^2}{2}$. 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 $\|\boldsymbol{s}' - \boldsymbol{s}\|_p \leq 2\varepsilon_p$[a] must satisfy the following: for some absolute constants $C_1, C_2 > 0$,*

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

$$\varepsilon_p \geq \sqrt{\frac{\log \frac{1}{\delta}}{C_2 B} \mathcal{A}^*_{\infty}(\boldsymbol{\sigma}, \boldsymbol{c})} \quad \textit{for } p \geq 2.$$

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

Proof sketch. We provide the full proof in Appendix C, which we sketch here. We suppose that the learner has knowledge of the true $\boldsymbol{\sigma}$. The proof follows the multiple hypotheses testing framework (Jedra & Proutière, 2023) based on 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

means 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 $p = 2$ as the function $z \mapsto z^{2/p}$ is convex iff $p \geq 2$. This is crucial in constructing a hard lower bound instance for ℓ_p, as for $p \geq 2$ (resp. $1 \leq p < 2$), the hardest instance is sparse (resp. dense) (Lemma C.2). □

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 $1 \leq p < 2$ and $\mathcal{A}_\infty^*(\boldsymbol{\sigma}, \boldsymbol{c})$ for $p \geq 2$. This gap is not an artifact of the proof constants, but reflects a limitation of Fano-type packing arguments in this problem.

5.2. A Sharp In-Expectation Lower Bound

The preceding result suggests that high-probability lower bounds based on global packing do not capture the full allocation geometry of the problem. We now show that the gap can be closed by changing the lower-bound technique. Using an Assouad-type construction, we obtain a local in-expectation lower bound that matches $\mathcal{A}_p^*(\boldsymbol{\sigma}, \boldsymbol{c})$ for all $p \in [1, \infty)$.

Theorem 5.2 (In-Expectation Lower Bound)**.** *Let $B > 0$ and $(\boldsymbol{s}^\star, \boldsymbol{\sigma}, \boldsymbol{c})$ be any given problem instance with $\boldsymbol{s}^\star = (s_k^\star)_k$. Let $R_k := \min\{s_k^\star, R - s_k^\star\}$ and $v_k := c_{j^*(k)}\sigma^2_{k,j^*(k)}$, and define "threshold" $B_0 := \left(\sum_{k\in[K]} v_k^{\frac{p}{p+2}}\right) \max_{k\in[K]} \frac{v_k^{\frac{2}{p+2}}}{4R_k^2}$ and "radius" $\xi(B) := \sqrt{\frac{\mathcal{A}_p^*(\boldsymbol{\sigma},\boldsymbol{c})}{16B}}$. Then, there exists a universal constant $C_3 > 0$ such that for $B \geq B_0$,*

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

and furthermore for $1 \leq p < \infty$ and $B \geq pB_0$,

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

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 & 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).

$\mathbb{E}[\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_p]$ is *not* decomposable coordinate-wise, hindering the direct application of Assouad's lemma. 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 $\mathbb{E}[\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_p^p]^{1/p}$, as it is decomposable, 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 $\mathcal{A}_p^*(\boldsymbol{\sigma}, \boldsymbol{c})$. □

We make two remarks concerning our lower bounds and their comparison with the upper bounds derived earlier.

(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 & Özgür, 2024; Zhu & Lafferty, 2014). In such settings, restricting the algorithm class naturally gives rise to an optimal allocation problem. Our analysis reveals a related phenomenon, but also uncovers a separation between high-probability and in-expectation lower-bound techniques in their ability to capture the correct allocation geometry. We discuss this separation in more detail in the next subsection.

(2) In-Expectation Upper Bounds. Since our sharp lower bound is stated in expectation, one may ask whether it is directly comparable to the high-probability upper bounds in Theorems 3.2 and 4.1. We note that an upper bound of the same order, without logarithmic factors, can also be obtained in expectation via the matrix Bernstein inequality (Tropp, 2015, Theorem 6.1.1).[2]

5.3. Why Assouad is Sharp and Fano is Not

While our in-expectation lower bounds via Assouad's Lemma correctly capture the dense $\mathcal{A}_p^*$ allocation geometry (Theorem 5.2), our high-probability bounds via Fano's inequality are suboptimal: for $p \geq 2$, Fano-based bounds saturate at the 1-sparse $\mathcal{A}_\infty^*$ rate, and even for $1 \leq p < 2$, they fail to recover the exact $\mathcal{A}_p^*$ allocation (Theorem 5.1). This separation stems from two distinct bottlenecks inherent to Fano-type arguments (Tsybakov, 2009) in our specific problem setting.

(i) The information-theoretical bottleneck. Fano-type arguments rely on constructing a globally separated packing of alternatives. In our setting, this creates a bottleneck: the KL divergences are compressed into a single scalar constraint before the errors are aggregated across coordinates.

[2] Although the result is formulated for matrices, setting $d_1 = d_2 = 1$ recovers the desired scalar inequality.

As a result, the fine ℓ_p-geometry that determines the optimal budget allocation is lost.

(ii) The geometric barrier. Fano-type arguments inherently rely on global volumetric packing over all coordinates simultaneously (Lemma C.2). For $p \geq 2$, maximizing an ℓ_p packing distance subject to an ℓ_2-based KL-divergence constraint forms a convex maximization problem, making the optimal adversarial packing to concentrate entirely on the single hardest query. While the optimal packing does become dense for $1 \leq p < 2$, the aforementioned information loss still prevents it from recovering the true optimal allocation.

In contrast, the in-expectation Assouad-based argument circumvents these limitations because it is structurally distinct. Rather than relying on a global bottleneck, it *first* lower bounds the risk coordinate-wise, *then* aggregates these bounds, and *finally* optimizes over the algorithm's budget allocation. By constructing a hypercube of independent perturbations, it forces any valid algorithm to distribute its budget globally across all queries to defend against simultaneous variations.

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 space reasons, we only provide an overview of the corresponding upper and lower bounds, highlighting the main differences and 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 & 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 more substantially different and technically more interesting. We sketch the main idea. 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 covariance given by the inverse Fisher information, rearranging the resulting expression 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

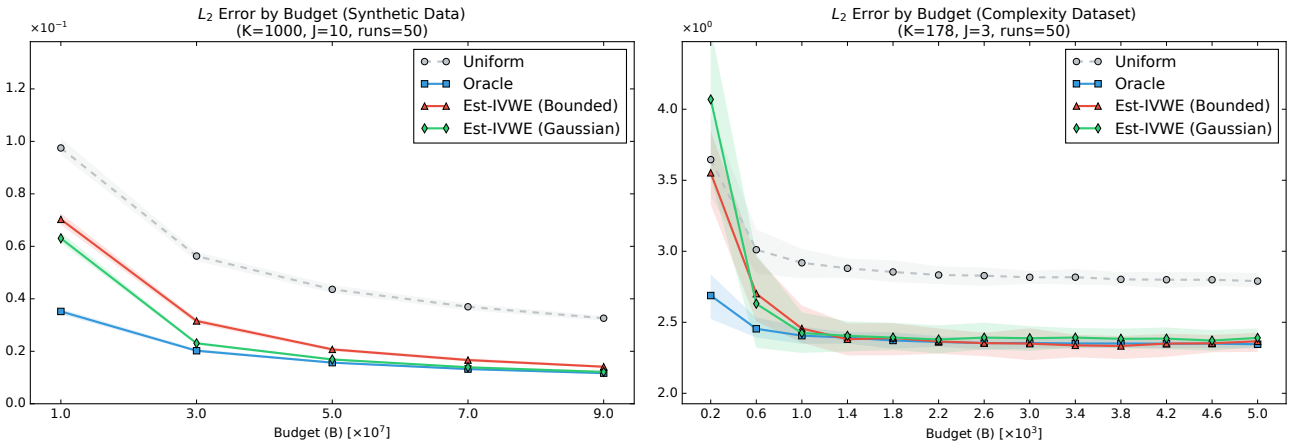
In this section, we numerically evaluate the performance of our proposed algorithms on both synthetic and real-world datasets.

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.

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 provided in Theorem 4.1 and Theorem F.4, rather than being heuristically tuned, in both experiments.

Results and Discussions. The results for the synthetic and HelpSteer2 datasets are shown in Figure 1a and Figure 1b, respectively. First, note that the 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, the Gaussian



(a) Synthetic data ($K = 1000$, $J = 10$)

(b) "Complexity" dataset ($K = 178$, $J = 3$)

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

variant 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. 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. Another interesting observation is that for HelpSteer2 datasets, the error curves eventually plateau at non-zero values. This phenomenon primarily stems from the finite evaluation pool used for resampling with replacement and the inherent biases present in the LLM judges (Gu et al., 2026).

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. The main algorithmic template is the inverse-variance weighted estimator (IVWE), whose error rate we analyze to derive the optimal oracle allocation. We developed a variant of IVWE, Est-IVWE, that does not require the knowledge of the variance profile via a two-stage approach: Stage I, the key ingredient, computes *optimistic* estimates of the variances, and Stage II is just IVWE with the estimates. We establish the *instance-wise* optimality of our algorithm by proving a matching, in-expectation lower bound for the error rate. Interestingly, we show that the standard Fano-type (high-probability) argument fails to capture the right allocation, while an Assouad-type (in-expectation) argument does, which may be of independent interest. We numerically validate the efficacy of IVWE and Est-IVWE on synthetic and HelpSteer2 datasets, showing its efficacy over naïve uniform allocation.

Several promising directions remain for future work. First, extending our framework to accommodate query-dependent and stochastic costs—such as varying token lengths per prompt—would better reflect real-world usage. Second, incorporating models of inherent judge bias could further refine evaluation accuracy. On the theoretical front, as noted in Section 5, deriving a matching high-probability lower bound remains an open challenge. Finally, scaling our empirical validation to larger, production-level LLM-as-a-judge pipelines is a crucial next step.

Acknowledgements

This work was supported in part by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) ([RS-2019-II190075, Artificial Intelligence Graduate School Program (KAIST), 5%], [RS-2024-00457882, AI Research Hub Project, 5%], and [RS-2022-II220311, Development of Goal-Oriented Reinforcement Learning Techniques for Contact-Rich Robotic Manipulation of Everyday Objects, 90%]).

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

Abeille, M., Faury, L., and Calauzènes, C. 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*, pp. 3691–3699. PMLR, 13–15 Apr 2021. URL https://proceedings.mlr.press/v130/abeille21a.html.

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

Anderson, T. W. 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.

Assouad, P. Deux remarques sur l'estimation. *Comptes rendus des séances de l'Académie des sciences. Série 1, Mathématique*, 1983. ISSN 0249-6291.

Bai, Y., Kadavath, S., Kundu, S., Askell, A., Kernion, J., Jones, A., Chen, A., Goldie, A., Mirhoseini, A., McKinnon, C., Chen, C., Olsson, C., Olah, C., Hernandez, D., Drain, D., Ganguli, D., Li, D., Tran-Johnson, E., Perez, E., Kerr, J., Mueller, J., Ladish, J., Landau, J., Ndousse, K., Lukosuite, K., Lovitt, L., Sellitto, M., Elhage, N., Schiefer, N., Mercado, N., DasSarma, N., Lasenby, R., Larson, R., Ringer, S., Johnston, S., Kravec, S., Showk, S. E., Fort, S., Lanham, T., Telleen-Lawton, T., Conerly, T., Henighan, T., Hume, T., Bowman, S. R., Hatfield-Dodds, Z., Mann, B., Amodei, D., Joseph, N., McCandlish, S., Brown, T., and Kaplan, J. Constitutional AI: Harmlessness from AI Feedback. *arXiv preprint arXiv:2212.08073*, 2022. URL https://arxiv.org/abs/2212.08073.

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

Bernstein, S. N. Sur l'extension du théoréme limite du calcul des probabilités aux sommes de quantités dépendantes. *Mathematische Annalen*, 97(1):1–59, Dec 1927. ISSN 1432-1807. doi:10.1007/BF01447859.

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

Boyd, S. and Vandenberghe, L. *Convex Optimization*. Cambridge University Press, 2004.

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

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

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

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

Chierichetti, F., Dasgupta, A., Kumar, R., and Lattanzi, S. Learning Entangled Single-Sample Gaussians. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '14, pp. 511–522, USA, 2014. Society for Industrial and Applied Mathematics. ISBN 9781611973389. doi:10.1137/1.9781611973402.38.

Cochran, W. G. 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. ISSN 1466-6162. doi:10.2307/2984123.

Cochran, W. G. The combination of estimates from different experiments. *Biometrics*, 10(1):101–129, 3 1954. ISSN 1541-0420. doi:10.2307/3001666.

Diakonikolas, I., Kane, D. M., Liu, S., and Pittas, T. Entangled Mean Estimation in High Dimensions. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, STOC '25, pp. 1680–1688, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400715105. doi:10.1145/3717823.3718162.

Efroimovich, S. Y. 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.

Fernandez, N., Kveton, B., Rossi, R. A., Lan, A., and Wang, Z. 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.

Fontaine, X., Perrault, P., Valko, M., and Perchet, V. 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*, pp. 3374–3383. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/fontaine21a.html.

Garivier, A., Ménard, P., and Stoltz, G. 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.

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

Ghosh, J. K. and Sinha, B. K. 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.

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

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

Gu, J., Jiang, X., Shi, Z., Tan, H., Zhai, X., Xu, C., Li, W., Shen, Y., Ma, S., Liu, H., Wang, S., Zhang, K., Lin, Z., Zhang, B., Ni, L., Gao, W., Wang, Y., and Guo, J. A survey on LLM-as-a-judge. *The Innovation*, pp. 101253, 2026. ISSN 2666-6758. doi:10.1016/j.xinn.2025.101253.

Guo, D., Yang, D., Zhang, H., Song, J., Wang, P., Zhu, Q., Xu, R., Zhang, R., Ma, S., Bi, X., Zhang, X., Yu, X., Wu, Y., Wu, Z. F., Gou, Z., Shao, Z., Li, Z., Gao, Z., Liu, A., Xue, B., Wang, B., Wu, B., Feng, B., Lu, C., Zhao, C., Deng, C., Ruan, C., Dai, D., Chen, D., Ji, D., Li, E., Lin, F., Dai, F., Luo, F., Hao, G., Chen, G., Li, G., Zhang, H., Xu, H., Ding, H., Gao, H., Qu, H., Li, H., Guo, J., Li, J., Chen, J., Yuan, J., Tu, J., Qiu, J., Li, J., Cai, J. L., Ni, J., Liang, J., Chen, J., Dong, K., Hu, K., You, K., Gao, K., Guan, K., Huang, K., Yu, K., Wang, L., Zhang, L., Zhao, L., Wang, L., Zhang, L., Xu, L., Xia, L., Zhang, M., Zhang, M., Tang, M., Zhou, M., Li, M., Wang, M., Li, M., Tian, N., Huang, P., Zhang, P., Wang, Q., Chen, Q., Du, Q., Ge, R., Zhang, R., Pan, R., Wang, R., Chen, R. J., Jin, R. L., Chen, R., Lu, S., Zhou, S., Chen, S., Ye, S., Wang, S., Yu, S., Zhou, S., Pan, S., Li, S. S., Zhou, S., Wu, S., Yun, T., Pei, T., Sun, T., Wang, T., Zeng, W., Liu, W., Liang, W., Gao, W., Yu, W., Zhang, W., Xiao, W. L., An, W., Liu, X., Wang, X., Chen, X., Nie, X., Cheng, X., Liu, X., Xie, X., Liu, X., Yang, X., Li, X., Su, X., Lin, X., Li, X. Q., Jin, X., Shen, X., Chen, X., Sun, X., Wang, X., Song, X., Zhou, X., Wang, X., Shan, X., Li, Y. K., Wang, Y. Q., Wei, Y. X., Zhang, Y., Xu, Y., Li, Y., Zhao, Y., Sun, Y., Wang, Y., Yu, Y., Zhang, Y., Shi, Y., Xiong, Y., He, Y., Piao, Y., Wang, Y., Tan, Y., Ma, Y., Liu, Y., Guo, Y., Ou, Y., Wang, Y., Gong, Y., Zou, Y., He, Y., Xiong, Y., Luo, Y., You, Y., Liu, Y., Zhou, Y., Zhu, Y. X., Huang, Y., Li, Y., Zheng, Y., Zhu, Y., Ma, Y., Tang, Y., Zha, Y., Yan, Y., Ren, Z. Z., Ren, Z., Sha, Z., Fu, Z., Xu, Z., Xie, Z., Zhang, Z., Hao, Z., Ma, Z., Yan, Z., Wu, Z., Gu, Z., Zhu, Z., Liu, Z., Li, Z., Xie, Z., Song, Z., Pan, Z., Huang, Z., Xu, Z., Zhang, Z., and Zhang, Z. DeepSeek-R1 incentivizes reasoning in LLMs through reinforcement learning. *Nature*, 645(8081):633–638, 2025. doi:10.1038/s41586-025-09422-z.

Jedra, Y. and Proutière, A. 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.

Kim, S., Suk, J., Longpre, S., Lin, B. Y., Shin, J., Welleck, S., Neubig, G., Lee, M., Lee, K., and Seo, M. 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*, pp. 4334–4353, Miami, Florida, USA, November 2024. Association for Computational Linguistics. doi:10.18653/v1/2024.emnlp-main.248.

Kirschner, J. and Krause, A. 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*, pp. 358–384. PMLR, 06–09 Jul 2018. URL https://proceedings.mlr.press/v75/kirschner18a.html.

Lattimore, T. and Szepesvári, C. *Bandit Algorithms*. Cambridge University Press, 2020.

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

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

Lee, J., Jang, K., Vojnović, M., Jun, K.-S., and Yun, S.-Y. GL-LowPopArt: A Nearly Instance-Wise Minimax-Optimal Estimator for Generalized Low-Rank Trace Re-

gression. 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.

Li, H., Dong, Q., Chen, J., Su, H., Zhou, Y., Ai, Q., Ye, Z., and Liu, Y. LLMs-as-Judges: A Comprehensive Survey on LLM-based Evaluation Methods. *arXiv preprint arXiv:2412.05579*, 2024. URL https://arxiv.org/abs/2412.05579.

Liang, Y. and Yuan, H. 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*, pp. 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.

Maurer, A. and Pontil, M. 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.

Nair, K. A. 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.

Neyman, J. 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. ISSN 0952-8385. doi:10.1111/j.2397-2335.1934.tb04184.x.

Norwood Jr, T. E. and Hinkelmann, K. 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.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P. F., Leike, J., and Lowe, R. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems*, volume 35, pp. 27730–27744. Curran Associates, Inc., 2022. URL https://openreview.net/forum?id=TG8KACxEON.

Pal, N. and Lim, W. K. 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. ISSN 0378-3758. doi:10.1016/S0378-3758(96)00202-9.

Pal, N., Lin, J.-J., Chang, C.-H., and Kumar, S. 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. ISSN 0167-9473. doi:10.1016/j.csda.2007.04.004.

Popoviciu, T. 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.

Rafailov, R., Sharma, A., Mitchell, E., Manning, C. D., Ermon, S., and Finn, C. Direct Preference Optimization: Your Language Model is Secretly a Reward Model. In *Advances in Neural Information Processing Systems*, volume 36, pp. 53728–53741. Curran Associates, Inc., 2023. URL https://openreview.net/forum?id=HPuSIXJaa9.

Raju, R. S., Jain, S., Li, B., Li, J. L., and Thakker, U. 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)*, pp. 167–181, Miami, Florida, USA, November 2024. Association for Computational Linguistics. doi:10.18653/v1/2024.customnlp4u-1.14.

Saha, A., Wagde, A., and Kveton, B. 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.

Simchowitz, M. and Foster, D. J. 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*, pp. 8937–8948. PMLR, 13–18 Jul 2020. URL https://proceedings.mlr.press/v119/simchowitz20a.html.

Sinha, B. K. and Mouqadem, O. Estimation of the common mean of two univariate normal populations. *Communications in Statistics - Theory and Methods*, 11(14): 1603–1614, 1982.

Tropp, J. A. An Introduction to Matrix Concentration Inequalities. *Foundations and Trends® in Machine Learning*, 8(1-2):1–230, 2015. ISSN 1935-8237. doi:10.1561/2200000048.

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

van Trees, H. L. *Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory*. John Wiley & Sons, New York, 1968. ISBN 978-0471899556.

Voinov, V. G. 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. ISSN 05815738. URL https://www.jstor.org/stable/25052354.

Wainwright, M. J. and Jordan, M. I. Graphical Models, Exponential Families, and Variational Inference. *Foundations and Trends® in Machine Learning*, 1(1–2):1–305, 2008. ISSN 1935-8237. doi:10.1561/2200000001.

Wang, Z., Dong, Y., Delalleau, O., Zeng, J., Shen, G., Egert, D., Zhang, J. J., Sreedhar, M. N., and Kuchaiev, O. HelpSteer 2: Open-source dataset for training top-performing reward models. In *Advances in Neural Information Processing Systems*, volume 37, pp. 1474–1501. Curran Associates, Inc., 2024. URL https://openreview.net/forum?id=PvVKUFhaNy.

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

Yu, B. Assouad, Fano, and Le Cam. In Pollard, D., Torgersen, E., and Yang, G. L. (eds.), *Festschrift for Lucien Le Cam: Research Papers in Probability and Statistics*, pp. 423–435. Springer New York, New York, NY, 1997. ISBN 978-1-4612-1880-7. doi:10.1007/978-1-4612-1880-7_29.

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

Zheng, L., Chiang, W.-L., Sheng, Y., Zhuang, S., Wu, Z., Zhuang, Y., Lin, Z., Li, Z., Li, D., Xing, E., Zhang, H., Gonzalez, J. E., and Stoica, I. 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.

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

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*[3]

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

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 $\mathbb{E}[X_j^{(i)}] = 0$, $|X_j^{(i)}| \leq \frac{R}{W_k \min_{j: \omega_{k,j} > 0} \sigma_{k,j}^2}$, 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 $k \in [K]$:

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

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

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

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

$$= \sqrt{2\log\frac{2K}{\delta}}\left(\sum_{k\in[K]} W_k^{-\frac{p}{2}}\right)^{\frac{1}{p}} + \frac{R\log\frac{2K}{\delta}}{3}\left(\sum_{k\in[K]}\left(W_k \min_{j:\omega_{k,j}>0}\sigma_{k,j}^2\right)^{-p}\right)^{\frac{1}{p}}$$

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

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

A.2. Proof of Theorem 3.2: Optimal Allocation

We first separate the cases into $p \in [1,\infty)$ and $p=\infty$, derive results for each case, and then conclude that, by convention, the optimal allocation can be represented in a single expression.

Optimal Allocation minimizing $\mathcal{A}_\infty$ Recall that the objective of the optimization problem $\mathcal{A}_p(\boldsymbol{\omega};\boldsymbol{\sigma},\boldsymbol{c})$ is given as:

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

When $p=\infty$, this objective becomes minimizing $\mathcal{A}_\infty(\boldsymbol{\omega};\boldsymbol{\sigma},\boldsymbol{c}) = \max_{k\in[K]}\left(\sum_{j\in[J]}\frac{\omega_{k,j}}{c_j\sigma_{k,j}^2}\right)^{-1}$, which can be calculated by convention. Since the inversion is decreasing, we instead solve the following optimization problem:

$$\max_{\boldsymbol{\omega}}\left\{\min_{k\in[K]}\sum_{j\in[J]}\frac{\omega_{k,j}}{c_j\sigma_{k,j}^2} : \sum_{(k,j)\in[K]\times[J]}\omega_{k,j}\leq 1,\quad \omega_{k,j}\geq 0\quad \forall k,j\right\}.$$

With a simple optimization trick, this problem can be converted to the linear optimization problem:

$$\max_{\boldsymbol{\omega},V}\left\{V : \sum_{k,j}\omega_{k,j}\leq 1;\quad V\leq\sum_{j\in[J]}\frac{\omega_{k,j}}{c_j\sigma_{k,j}^2}\quad\forall k;\quad \omega_{k,j}\geq 0\quad\forall k,j\right\}.$$

The Lagrangian $\mathcal{L}$ with multipliers λ_k (for the V constraints), λ_0 (for the budget constraint), and $\nu_{k,j}$ (for non-negativity) is:

$$\begin{aligned}\mathcal{L}\big(\boldsymbol{\omega},V,\lambda_0,\{\lambda_k\},\{\nu_{k,j}\}\big) &= V-\lambda_0\left(\sum_{k,j}\omega_{k,j}-1\right)-\sum_k\lambda_k\left(V-\sum_j\frac{\omega_{k,j}}{c_j\sigma_{k,j}^2}\right)+\sum_{k,j}\nu_{k,j}\omega_{k,j}\\ &= V\left(1-\sum_k\lambda_k\right)+\sum_{k,j}\left(\frac{\lambda_k}{c_j\sigma_{k,j}^2}-\lambda_0+\nu_{k,j}\right)\omega_{k,j}+\lambda_0,\end{aligned}$$

where $\lambda_k\geq 0$, $\lambda_0\geq 0$, and $\nu_{k,j}\geq 0$.

For an optimal solution $(\boldsymbol{\omega}^*,V^*)$ to exist, the KKT conditions consisting of stationarity conditions:

$$\begin{cases}\frac{\partial\mathcal{L}}{\partial V}=0 \implies \sum_k\lambda_k=1\\ \frac{\partial\mathcal{L}}{\partial\omega_{k,j}}=0 \implies \frac{\lambda_k}{c_j\sigma_{k,j}^2}-\lambda_0+\nu_{k,j}=0 \implies \nu_{k,j}=\lambda_0-\frac{\lambda_k}{c_j\sigma_{k,j}^2}\end{cases},$$

and complementary slackness conditions:

$$\begin{cases} \lambda_k \left(V - \sum_j \frac{\omega_{k,j}}{c_j \sigma_{k,j}^2} \right) = 0 \\ \lambda_0 \left(\sum_{k,j} \omega_{k,j} - 1 \right) = 0 \\ \nu_{k,j} \omega_{k,j} = 0 \end{cases}$$

must hold.

For a fixed $k \in [K]$, we have that $\lambda_0 - \frac{\lambda_k}{c_j \sigma_{k,j}^2} = \nu_{k,j} \geq 0$ for all $j \in [J]$ from above conditions, which is equivalent to $\lambda_k \leq \lambda_0 (c_j \sigma_{k,j}^2)$ for all $j \in [J]$. This implies that $\lambda_k \leq \lambda_0 \min_{j \in [J]} (c_j \sigma_{k,j}^2)$. If we allocate a query k to a judge j, i.e., $\omega_{k,j} > 0$, complementary slackness condition implies that $\nu_{k,j} = 0$ for that judge. For such a judge j, the equation $\lambda_k = \lambda_0 (c_j \sigma_{k,j}^2)$ holds. This equation and the inequality $\lambda_k \leq \lambda_0 \min_{j \in [J]} (c_j \sigma_{k,j}^2) = \lambda_0 c_{j^*(k)} \sigma_{k,j^*(k)}^2$ say that such judge j is one of the minimizer of $c_j \sigma_{k,j}^2$, i.e. $j = j^*(k)$. Unless the minimizer is unique, we select one among them with any consistent tie-breaking rule, and say that as $j^*(k)$. We only allocate the query k to this $j^*(k)$, then it is still the optimal solution of LP problem. For the other judges, the inequality $\lambda_k \leq \lambda_0 \min_{j \in [J]} (c_j \sigma_{k,j}^2)$ should be strict, i.e., $\omega_{k,j} = 0$ for $j \neq j^*(k)$.

Now, we claim that $\lambda_k > 0$ for all k. If there is a query k with $\lambda_k = 0$, it implies that the constraint $V \leq \sum_j \frac{\omega_{k,j}}{c_j \sigma_{k,j}^2}$ is non-binding for that k. In this case, by redistributing the resource allocated to this query k, we can raise the value of V. Consequently, at the optimum, the constraint $V \leq \sum_j \frac{\omega_{k,j}}{c_j \sigma_{k,j}^2}$ should be binding for all $k \in [K]$, i.e., $V^* = \frac{\omega_{k,j^*(k)}^*}{c_j \sigma_{k,j^*(k)}^2}$ for all $k \in [K]$.

Substituting $\omega_{k,j^*(k)} = V^* c_{j^*(k)} \sigma_{k,j^*(k)}^2$ into the budget constraint (which is tight, $\lambda_0 > 0$), then the optimal value of the optimization problem V^* satisfies

$$V^* \sum_k c_{j^*(k)} \sigma_{k,j^*(k)}^2 = 1.$$

Therefore, the optimal allocation is

$$\omega_{k,j}^* = \frac{\mathbb{1}[j = j^*(k)] c_{j^*(k)} \sigma_{k,j^*(k)}^2}{\sum_{k' \in [K]} c_{j^*(k')} \sigma_{k',j^*(k)}^2},$$

where $j^*(k) \triangleq \arg\min_{j \in [J]} c_j \sigma_{k,j}^2$; and the optimal value of the optimization problem is

$$V_*^{-1} = \left(\frac{\omega_{k,j}^*}{c_{j^*(k)} \sigma_{k,j^*(k)}^2} \right)^{-1} = \sum_{k \in [K]} c_{j^*(k)} \sigma_{k,j^*(k)}^2.$$

Optimal Allocation minimizing $\mathcal{A}_p$ We first solve the optimization problem for $p \in [1, \infty)$. Since the power of $2/p$ is monotone, we instead solve the optimization problem:

$$\min_{\boldsymbol{\omega}} \left\{ \sum_{k \in [K]} \left(\sum_{j \in [J]} \frac{\omega_{k,j}}{c_j \sigma_{k,j}^2} \right)^{-\frac{p}{2}} : \sum_{(k,j) \in [K] \times [J]} \omega_{k,j} \leq 1; \quad \omega_{k,j} \geq 0 \quad \forall k, j \right\}.$$

Let λ_0 be the multiplier for the budget constraint and $\nu_{k,j}$ be the multipliers for the non-negativity of $\omega_{k,j}$, then the Lagrangian is:

$$\mathcal{L}(\boldsymbol{\omega}, \lambda_0, \nu_{k,j}) = \sum_k \left(\sum_j \frac{\omega_{k,j}}{c_j \sigma_{k,j}^2} \right)^{-\frac{p}{2}} + \lambda_0 \left(\sum_{k,j} \omega_{k,j} - 1 \right) - \sum_{k,j} \nu_{k,j} \omega_{k,j}$$

The stationarity of KKT conditions requires the partial derivative of the Lagrangian with respect to decision variables $\omega_{k,j}$ to be zero:

$$\frac{\partial \mathcal{L}}{\partial \omega_{k,j}} = -\frac{p}{2} \left(\sum_j \frac{\omega_{k,j}}{c_j \sigma_{k,j}^2} \right)^{-\frac{p+2}{2}} \cdot \frac{1}{c_j \sigma_{k,j}^2} + \lambda_0 - \nu_{k,j} = 0.$$

Denoting $V_k = \frac{p}{2}\left(\sum_{j\in[J]} \frac{\omega_{k,j}}{c_j\sigma_{k,j}^2}\right)^{-\frac{p+2}{2}}$ and rearranging the terms, we obtain $\nu_{k,j} = \lambda_0 - \frac{V_k}{c_j\sigma_{k,j}^2}$. Furthermore, complementary slackness requires $\lambda_0(\sum_{k,j}\omega_{k,j} - 1) = 0$ and $\nu_{k,j}\omega_{k,j} = 0$ for all k, j.

For a fixed $k \in [K]$, from the dual feasibility $\nu_{k,j} \geq 0$, it follows that $V_k \leq \lambda_0 c_j \sigma_{k,j}^2$ for all $j \in [J]$, which implies that $V_k \leq \lambda_0 \min_{j\in[J]} c_j\sigma_{k,j}^2$. If $\omega_{k,j} > 0$, we have $\nu_{k,j} = 0$. For this $j = j^*(k)$.

$$\frac{p}{2}\left(\frac{\omega_{k,j^*(k)}}{c_{j^*(k)}\sigma_{k,j^*(k)}^2}\right)^{-\frac{p+2}{2}} = \lambda_0 c_{j^*(k)}\sigma_{k,j^*(k)}^2,$$

which implies that

$$\omega_{k,j^*(k)} = \left(\frac{2}{p}\lambda_0\right)^{-\frac{2}{p+2}}\left(c_{j^*(k)}\sigma_{k,j^*(k)}^2\right)^{\frac{p}{p+2}}.$$

Substituting this into the budget constraint, which is binding, then we get:

$$\left(\frac{2}{p}\lambda_0\right)^{-\frac{2}{p+2}}\sum_{k\in[K]}\left(c_{j^*(k)}\sigma_{k,j^*(k)}^2\right)^{\frac{p}{p+2}} = 1.$$

Therefore, the optimal allocation is

$$w_{k,j}^* = \frac{\mathbb{1}[j = j^*(k)]\left(c_{j^*(k)}\sigma_{k,j^*(k)}^2\right)^{\frac{p}{p+2}}}{\sum_{k'\in[K]}\left(c_{j^*(k')}\sigma_{k',j^*(k')}^2\right)^{\frac{p}{p+2}}}$$

where $j^*(k) \triangleq \arg\min_{j\in[J]} c_j\sigma_{k,j}^2$; and the optimal value of the $\mathcal{A}_p^*(\boldsymbol{\sigma}, \boldsymbol{c})$ is

$$\begin{aligned}
\left(\sum_{k\in[K]}\left(\sum_{j\in[J]}\frac{\omega_{k,j}^*}{c_j\sigma_{k,j}^2}\right)^{-\frac{p}{2}}\right)^{\frac{2}{p}} &= \left(\sum_{k\in[K]}\left(\frac{\omega_{k,j^*(k)}^*}{c_{j^*(k)}\sigma_{k,j^*(k)}^2}\right)^{-\frac{p}{2}}\right)^{\frac{2}{p}} \\
&= \left(\sum_{k\in[K]}\left(\frac{\left(c_{j^*(k)}\sigma_{k,j^*(k)}^2\right)^{-\frac{2}{p+2}}}{\sum_{k'\in[K]}\left(c_{j^*(k')}\sigma_{k',j^*(k')}^2\right)^{\frac{p}{p+2}}}\right)^{-\frac{p}{2}}\right)^{\frac{2}{p}} \\
&= \left(\frac{\sum_{k\in[K]}\left(c_{j^*(k)}\sigma_{k,j^*(k)}^2\right)^{\frac{p}{p+2}}}{\left(\sum_{k'\in[K]}\left(c_{j^*(k')}\sigma_{k',j^*(k')}^2\right)^{\frac{p}{p+2}}\right)^{-\frac{p}{2}}}\right)^{\frac{2}{p}} \\
&= \left(\sum_{k\in[K]}\left(c_{j^*(k)}\sigma_{k,j^*(k)}^2\right)^{\frac{p}{p+2}}\right)^{\frac{p+2}{p}}.
\end{aligned}$$

When $p = \infty$, we treat as $p/(p+2) = 1$ and $(p+2)/p = 1$ by convention. Therefore, this final expression can express the optimal allocation $\omega_{k,j}^*$ and minimum value $\mathcal{A}_p^*$ for all $p \in [1, \infty]$.

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 $\widehat{N}_k \coloneqq \frac{B' \widehat{\omega}^*_{k,\hat{j}^*(k)}}{c_{\hat{j}^*(k)}}$.

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 $k \in [K]$,

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

By Minkowski's inequality, we then have

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

We bound each (a) and (b) separately. We utilize the useful confidence bound on variance estimates given by Maurer & 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 & Pontil (2009))**.** *With $\hat{\sigma}^2_{k,j}$ as in line 3 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:

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. Here, we divide the pairs into "easy" ones where $\sigma_{k,j} \le \tau$, and "hard" ones where $\sigma_{k,j} > \tau$. For the easy pairs, we have that

$$\sigma_{k,j} \le \tau \le \overline{\sigma}_{k,j}.$$

For the hard pairs, we have that by the empirical Bernstein inequality (Lemma B.1), with probability at least $1-\delta$,

$$\sigma_{k,j} \le \hat{\sigma}_{k,j} + R\sqrt{\frac{2}{N_0-1}\log\frac{2KJ}{\delta}} = \overline{\sigma}_{k,j} - \tau + R\sqrt{\frac{2}{N_0-1}\log\frac{2KJ}{\delta}}.$$

Thus, as long as $\tau \geq R\sqrt{\frac{2}{N_0-1}\log\frac{2KJ}{\delta}}$, we have that $\sigma_{k,j} \leq \overline{\sigma}_{k,j}$. The other direction follows directly from our choice of τ and empirical Bernstein inequality (Lemma B.1), again. □

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

B.2. Bounding (a)

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

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

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 $c_{\hat{j}^*(k)}\overline{\sigma}^2_{k,\hat{j}^*(k)} \leq c_{j^*(k)}\overline{\sigma}^2_{k,j^*(k)}$ by construction of Algorithm 1. With these, we have that with probability at least $1-2\delta$,

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

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

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

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

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

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:

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

$$\leq \tau^{-2}\left(\sum_{k\in[K]}\left(c_{j^*(k)}\overline{\sigma}^2_{k,j^*(k)}\right)^{\frac{p}{p+2}}\right)\left(\left(\sum_{k\in[K]}\left(c_{j^*(k)}\overline{\sigma}^2_{k,j^*(k)}\right)^{\frac{p}{p+2}}\right)^2\right)^{\frac{1}{p}} \qquad (\textstyle\sum_k y_k^2 \leq (\sum_k y_k)^2 \text{ for } y_k \geq 0)$$

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

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

We have set $\tau = R\sqrt{\frac{2}{N_0-1}\log\frac{2KJ}{\delta}}$, and we denote $C := K\sum_{j\in[J]}c_j$. For simplicity, we divide the computation into two parts. In both cases, we assume that $B \geq 2CN_0$.

B.4.1. $p \geq 2$

Substituting $B' = B - N_0K\sum_{j\in[J]}c_j$,

$$\begin{aligned}
\varepsilon_p &\leq \sqrt{\frac{2\log\frac{2K}{\delta}}{B-CN_0}}\left(\sqrt{\mathcal{A}^*_p(\boldsymbol{\sigma},\boldsymbol{c})} + C^*_{p,1}R\sqrt{\frac{2}{N_0-1}\log\frac{2KJ}{\delta}}\right) \\
&\quad + \frac{R\log\frac{2K}{\delta}}{B-CN_0}\left(\mathcal{A}^*_p(\boldsymbol{\sigma},\boldsymbol{c})\left(R\sqrt{\frac{2}{N_0-1}\log\frac{2KJ}{\delta}}\right)^{-2} + (C^*_{p,1})^2\right) \\
&\leq \sqrt{\frac{2\mathcal{A}^*_p(\boldsymbol{\sigma},\boldsymbol{c})\log\frac{2K}{\delta}}{B-CN_0}} + RC^*_{p,1}\log\frac{2KJ}{\delta}\sqrt{\frac{1}{N_0(B-CN_0)}} \\
&\quad + \frac{\mathcal{A}^*_p(\boldsymbol{\sigma},\boldsymbol{c})}{R}\frac{N_0}{B-CN_0} + \frac{(C^*_{p,1})^2R\log\frac{2K}{\delta}}{B-CN_0}
\end{aligned}$$

Note that

$$\frac{1}{\sqrt{B-CN_0}} = \frac{1}{\sqrt{B}}\left(1-\frac{CN_0}{B}\right)^{-\frac{1}{2}} \overset{(*)}{\leq} \frac{1}{\sqrt{B}}\left(1+\frac{CN_0}{B}\right) = \frac{1}{\sqrt{B}} + \frac{CN_0}{B^{\frac{3}{2}}},$$

where $(*)$ follows from the simple algebraic inequality: $(1-x)^{-1/2} \leq 1+x$ for $x\in[0,1/2]$ (Lemma B.3, whose proof we provide at the end for completeness).

With the above inequality and $B \geq 2CN_0$, we have that

$$\begin{aligned}
\varepsilon_p &\leq \sqrt{\frac{2\mathcal{A}^*_p(\boldsymbol{\sigma},\boldsymbol{c})\log\frac{2K}{\delta}}{B}} + RC^*_{p,1}\log\frac{2KJ}{\delta}\sqrt{\frac{2}{N_0B}} + \frac{\mathcal{A}^*_p(\boldsymbol{\sigma},\boldsymbol{c})}{R}\frac{2N_0}{B} \\
&\quad + \frac{2(C^*_{p,1})^2R\log\frac{2K}{\delta}}{B} + \frac{CN_0\sqrt{2\mathcal{A}^*_p(\boldsymbol{\sigma},\boldsymbol{c})\log\frac{2K}{\delta}}}{B^{\frac{3}{2}}}.
\end{aligned}$$

A nearly optimal choice of N_0 here is attained by balancing[4] the red terms:

$$N_0 = 2^{-\frac{1}{3}}\left(R^2\log\frac{2KJ}{\delta}\right)^{\frac{2}{3}}B^{\frac{1}{3}},$$

[4]Here, we take into consideration the fact that $\mathcal{A}^*_p(\boldsymbol{\sigma},\boldsymbol{c})$ and $C^*_{p,1}$ are unknown to the learner. And even if the learner does have even an approximate knowledge of these terms, as this constitutes the lower-order terms, this is rather insignificant.

which yields

$$\begin{aligned}\varepsilon_p &\le \sqrt{\frac{2\mathcal{A}_p^*(\boldsymbol{\sigma},\boldsymbol{c})\log\frac{2K}{\delta}}{B}} + \frac{2^{\frac{2}{3}}R^{\frac{1}{3}}\left(C_{p,1}^* + \mathcal{A}_p^*(\boldsymbol{\sigma},\boldsymbol{c})\right)\left(\log\frac{2KJ}{\delta}\right)^{\frac{2}{3}}}{B^{\frac{2}{3}}} \\ &\quad + \frac{2(C_{p,1}^*)^2 R\log\frac{2K}{\delta}}{B} + \frac{2^{\frac{1}{6}}R^{\frac{4}{3}}C\left(\log\frac{2KJ}{\delta}\right)^{\frac{7}{6}}\sqrt{\mathcal{A}_p^*(\boldsymbol{\sigma},\boldsymbol{c})}}{B^{\frac{7}{6}}} \\ &= \sqrt{\frac{2\mathcal{A}_p^*(\boldsymbol{\sigma},\boldsymbol{c})\log\frac{2K}{\delta}}{B}} + \mathcal{O}\left(\frac{R^{\frac{1}{3}}\mathcal{A}_p^*(\boldsymbol{\sigma},\boldsymbol{c})\left(\log\frac{2KJ}{\delta}\right)^{\frac{2}{3}}}{B^{\frac{2}{3}}}\right).\end{aligned}$$

B.4.2. $1 \le p < 2$

Again substituting $B' = B - N_0 C$ and using $\tau^{\frac{2p}{p+2}} \ge \tau$ as $\tau \in (0,1)$,

$$\begin{aligned}\varepsilon_p &\le \sqrt{\frac{2\log\frac{2K}{\delta}}{B - CN_0}}\left(\sqrt{\mathcal{A}_p^*(\boldsymbol{\sigma},\boldsymbol{c})} + (C_{p,1}^* + C_{p,2}^*)\left(R^2\frac{2}{N_0-1}\log\frac{2KJ}{\delta}\right)^{\frac{p}{p+2}}\right) \\ &\quad + \frac{2R\log\frac{2K}{\delta}}{B - CN_0}\left(\mathcal{A}_p^*(\boldsymbol{\sigma},\boldsymbol{c})\left(R\sqrt{\frac{2}{N_0-1}\log\frac{2KJ}{\delta}}\right)^{-2} + (C_{p,1}^*)^2\right) \\ &\quad + \frac{2R\log\frac{2K}{\delta}}{B - CN_0}(C_{p,2}^*)^2\left(R^2\frac{2}{N_0-1}\log\frac{2KJ}{\delta}\right)^{\frac{p-2}{p+2}} \\ &\le \sqrt{\frac{2\mathcal{A}_p^*(\boldsymbol{\sigma},\boldsymbol{c})\log\frac{2K}{\delta}}{B - CN_0}} + \frac{2^{\frac{3p+2}{2p+4}}R^{\frac{2p}{p+2}}(C_{p,1}^* + C_{p,2}^*)\left(\log\frac{2KJ}{\delta}\right)^{\frac{3p+2}{2p+4}}}{N_0^{\frac{p}{p+2}}(B - CN_0)^{\frac{1}{2}}} \\ &\quad + \frac{\mathcal{A}_p^*(\boldsymbol{\sigma},\boldsymbol{c})}{R}\frac{N_0}{B - CN_0} + \frac{2(C_{p,1}^*)^2 R\log\frac{2K}{\delta}}{B - CN_0} \\ &\quad + \frac{2^{\frac{2p}{p+2}}(C_{p,2}^*)^2 R^{\frac{3p-2}{p+2}}\left(\log\frac{2KJ}{\delta}\right)^{\frac{2p}{p+2}} N_0^{\frac{2-p}{2+p}}}{B - CN_0}.\end{aligned}$$

Using the same inequality as above and $B \ge 2CN_0$, we have that

$$\begin{aligned}\varepsilon_p &\le \sqrt{\frac{2\mathcal{A}_p^*(\boldsymbol{\sigma},\boldsymbol{c})\log\frac{2K}{\delta}}{B}} + \frac{2^{\frac{2p+2}{p+2}}R^{\frac{2p}{p+2}}(C_{p,1}^* + C_{p,2}^*)\left(\log\frac{2KJ}{\delta}\right)^{\frac{3p+2}{2p+4}}}{N_0^{\frac{p}{p+2}}B^{\frac{1}{2}}} + \frac{\mathcal{A}_p^*(\boldsymbol{\sigma},\boldsymbol{c})}{R}\frac{2N_0}{B} \\ &\quad + \frac{4(C_{p,1}^*)^2 R\log\frac{2K}{\delta}}{B} + \frac{(C_{p,2}^*)^2 2^{\frac{3p+2}{p+2}}R^{\frac{3p-2}{p+2}}\left(\log\frac{2KJ}{\delta}\right)^{\frac{2p}{p+2}} N_0^{\frac{2-p}{2+p}}}{B} \\ &\quad + \frac{CN_0\sqrt{2\mathcal{A}_p^*(\boldsymbol{\sigma},\boldsymbol{c})\log\frac{2K}{\delta}}}{B^{\frac{3}{2}}}.\end{aligned}$$

Similarly, a nearly optimal choice of N_0 here is attained by balancing the red terms:

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

which yields

$$\begin{aligned}\varepsilon_p &\le \sqrt{\frac{2\mathcal{A}_p^*(\boldsymbol{\sigma},\boldsymbol{c})\log\frac{2K}{\delta}}{B}} + \frac{2^{\frac{3p+2}{2p+2}}\left(C_{p,1}^* + C_{p,2}^* + \mathcal{A}_p^*(\boldsymbol{\sigma},\boldsymbol{c})\right)R^{\frac{p}{2p+2}}\left(\log\frac{2KJ}{\delta}\right)^{\frac{3p+2}{4p+4}}}{B^{\frac{3p+2}{4p+4}}} \\ &\quad + \frac{4(C_{p,1}^*)^2 R\log\frac{2K}{\delta}}{B} + \cdots^{5}\end{aligned}$$

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

Since this bound holds with probability at least $1 - 2\delta$, the desired result follows by rescaling δ to $\delta/2$.

□

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 $[0, 1/2]$. □

[5]Due to its insignificance and awful form, we omit the further lower order terms, which are fourth and fifth terms with N_0 substituted in.

C. Proof of Theorem 5.1: High-Probability Lower Bound

The proof relies on the data processing inequality (Garivier et al., 2019), as utilized in Jedra & Proutière (2023) to derive a high-probability minimax lower bound.[6] The proof "bifurcates" at whether $1 \le p < 2$ or $p \ge 2$. For the former, we construct a single dense hard instance, while for the latter, we construct K 1-sparse hard instances and optimize for the algorithm's success probabilities, which deviates from prior lower bound techniques (Jedra & Proutière, 2023).

Throughout, the learner knows the given cost vector $\boldsymbol{c}$ and the true variance profile $\boldsymbol{\sigma}$ with $\max_{k,j} \sigma_{k,j}^2 \le \frac{R^2}{2}$.

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[7] 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 & 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 & 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

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

We first take the infimum over all $\boldsymbol{s}$ with $\|\boldsymbol{s} - \boldsymbol{s}^\star\|_p > 2\varepsilon_p$ satisfying $\Delta_k = s_k - s_k^\star \le \frac{R_k}{4}$ to optimize the RHS w.r.t. hard instances. This is characterized in the next lemma, whose proof is deferred to Appendix C.1:

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

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

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 $Z : \Omega \to [0, 1]$.

We divide into two cases:

Case I. $1 \leq p < 2$**.** In this case, we consider a single hard instance as in ***Case I*** of Lemma C.2. By construction, we then have that $\|\boldsymbol{s} - \boldsymbol{s}^\star\|_p = 2\varepsilon_p$.

Define an event $\mathcal{E} \triangleq \left\{\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_p \leq \varepsilon_p\right\}$. We first show that $\mathbb{P}_{\boldsymbol{s}^\star}(\mathcal{E}) < \delta$. To do so, suppose that $\mathcal{E}$ is true. Then,

$$2\varepsilon_p = \|\boldsymbol{s} - \boldsymbol{s}^\star\|_p \overset{(i)}{\leq} \|\boldsymbol{s} - \hat{\boldsymbol{s}}\|_p + \|\hat{\boldsymbol{s}} - \boldsymbol{s}^\star\|_p \overset{(ii)}{<} \varepsilon_p + \|\hat{\boldsymbol{s}} - \boldsymbol{s}^\star\|_p \Longrightarrow \|\hat{\boldsymbol{s}} - \boldsymbol{s}^\star\|_p > \varepsilon_p,$$

where (i) follows from triangle inequality and (ii) from the fact that $\mathcal{E}$ holds. But, as $\mathcal{A}$ is (B, δ, ℓ_p)-budget efficient, $\|\hat{\boldsymbol{s}} - \boldsymbol{s}^\star\|_p > \varepsilon_p$ holds with probability at most δ under $\mathbb{P}_{\boldsymbol{s}^\star}$.

We then apply the data processing inequality (Lemma C.3) as follows:

$$D_{\mathrm{KL}}(\mathbb{P}_{\boldsymbol{s}^\star}, \mathbb{P}_{\boldsymbol{s}}) \geq \mathrm{kl}\left(\mathbb{P}_{\boldsymbol{s}^\star}(\mathcal{E}), \mathbb{P}_{\boldsymbol{s}}(\mathcal{E})\right) \overset{(*)}{\geq} \mathrm{kl}\left(\delta, 1-\delta\right) = \mathrm{kl}(1-\delta, \delta),$$

where $(*)$ follows from the monotonicity of $\mathrm{kl}(\cdot, \cdot)$[8], combined with the facts that $\mathbb{P}_{\boldsymbol{s}}(\mathcal{E}) \geq 1 - \delta$ ($\mathcal{A}$ is (B, δ)-budget efficient with error rate ε_p) and $\mathbb{P}_{\boldsymbol{s}^\star}(\mathcal{E}) \leq \delta$ as shown above.

Thus, we are left with

$$C_1 \sum_{k,j} \mathbb{E}_{\boldsymbol{s}^\star}^{\mathcal{A}}[N_{k,j}(T)] \frac{(s_k - s_k^\star)^2}{\sigma_{k,j}^2} \geq \mathrm{kl}\left(1-\delta, \delta\right).$$

[8] $q \mapsto \mathrm{kl}(p, q)$ is increasing in $[p, 1]$, and $p \mapsto \mathrm{kl}(p, q)$ is decreasing in $[0, 1/2]$ when $q \geq 1/2$.

We take the supremum over all (B, δ, ℓ_p)-budget efficient algorithms $\mathcal{A}$'s, which is equivalent to taking a supremum over all possible allocations $\{N_{k,j}\}_{k,j}$ satisfying the knapsack constraint $\sum_{k,j} N_{k,j} c_j \leq B$. Applying Lemma C.2 and rearranging for ε_p, we obtain:

$$\varepsilon_p^2 \geq \frac{\mathrm{kl}(1-\delta,\delta)}{4C_1} \min_{\{N_{k,j}\}:\sum_{k,j} N_{k,j} c_j \leq B} \left(\sum_k \left(\sum_{j=1}^J \frac{N_{k,j}}{\sigma_{k,j}^2} \right)^{-\frac{p}{2-p}} \right)^{\frac{2-p}{p}}$$

$$\overset{(**)}{\geq} \frac{\mathrm{kl}(1-\delta,\delta)}{4C_1 B} \underbrace{\left(\sum_{k\in[K]} \left(c_{j^*(k)} \sigma_{k,j^*(k)}^2 \right)^{\frac{p}{2}} \right)^{\frac{2}{p}}}_{=\mathcal{A}^*_{\frac{2p}{2-p}}(\boldsymbol{\sigma},\mathbf{c})},$$

where $(**)$ follows from similar reasoning as in Appendix A.2.

Case II. $p \geq 2$**.** In this case, we consider the following set of K hard instances:

$$\left\{ \boldsymbol{s}^{(k)} \triangleq \boldsymbol{s}^\star + 2\varepsilon_p \boldsymbol{e}_k \ : \ k \in [K] \right\},$$

which satisfies $D_{\mathrm{KL}}(\mathbb{P}_{\boldsymbol{s}^\star}, \mathbb{P}_{\boldsymbol{s}^{(k)}}) \leq 4C_1 \varepsilon_p^2 w_k$ for $k \in [K]$.

For each $k \in [K]$, define $\mathcal{R}^{(k)}$ as the event that the algorithm $\mathcal{A}$ outputs $\hat{\boldsymbol{s}}$ that is ε_p-close to $\boldsymbol{s}^{(k)}$:

$$\mathcal{R}^{(k)} := \left\{ \left\| \hat{\boldsymbol{s}} - \boldsymbol{s}^{(k)} \right\|_p \leq \varepsilon_p \right\},$$

and define the null event

$$\mathcal{R}^{(0)} := \left\{ \left\| \hat{\boldsymbol{s}} - \boldsymbol{s}^\star \right\|_p \leq \varepsilon_p \right\}.$$

Then, we have that $\mathcal{R}^{(k)} \cap \mathcal{R}^{(k')} = \emptyset$ for $k \neq k'$ as

$$\left\| \boldsymbol{s}^{(k)} - \boldsymbol{s}^\star \right\|_p = 2\varepsilon_p, \quad \left\| \boldsymbol{s}^{(k)} - \boldsymbol{s}^{(k')} \right\|_p = 2^{1+\frac{1}{p}} \varepsilon_p > 2\varepsilon_p,$$

where the last inequality holds as $p \geq 2$, Let $p^{(k)} := \mathbb{P}_{\boldsymbol{s}^\star}(\mathcal{R}^{(k)})$ be the (error) probability that $\mathcal{A}$ outputs $\hat{\boldsymbol{s}}$ that is closest to $\boldsymbol{s}^{(k)}$.

As $\mathcal{A}$ is (B, δ, ℓ_p)-budget efficient, we have that

$$\sum_{k\in[K]} p^{(k)} \leq 1 - \mathbb{P}_{\boldsymbol{s}^\star}(\mathcal{R}^{(0)}) \leq \delta, \quad \mathbb{P}_{\boldsymbol{s}^{(k)}}(\mathcal{R}^{(k)}) \geq 1 - \delta.$$

Then by applying the data processing inequality (Lemma C.3) per k, we have

$$\begin{aligned} 4C_1 \varepsilon_p^2 \sum_{k\in[K]} c_{j^*(k)} \sigma_{k,j^*(k)}^2 w_k &\geq \sum_{k\in[K]} c_{j^*(k)} \sigma_{k,j^*(k)}^2 D_{\mathrm{KL}}(\mathbb{P}_{\boldsymbol{s}^\star}, \mathbb{P}_{\boldsymbol{s}^{(k)}}) \\ &\geq \sum_{k\in[K]} c_{j^*(k)} \sigma_{k,j^*(k)}^2 \mathrm{kl}(\mathbb{P}_{\boldsymbol{s}^\star}(\mathcal{R}^{(k)}), \mathbb{P}_{\boldsymbol{s}^{(k)}}(\mathcal{R}^{(k)})) \\ &\overset{(*)}{\geq} \sum_{k\in[K]} c_{j^*(k)} \sigma_{k,j^*(k)}^2 \mathrm{kl}(p^{(k)}, 1-\delta), \end{aligned}$$

where $(*)$ follows from the monotonicity of $\mathrm{kl}(\cdot,\cdot)$.

Also note that due to the knapsack constraint, recalling the definition of w_k,

$$\sum_{k\in[K]} c_{j^*(k)} \sigma_{k,j^*(k)}^2 w_k = \sum_{k\in[K]} \sum_{j\in[J]} \frac{\mathbb{E}_{\boldsymbol{s}^\star}^{\mathcal{A}}[N_{k,j}(T)] c_{j^*(k)} \sigma_{k,j^*(k)}^2}{\sigma_{k,j}^2}$$

$$= \sum_{k\in[K]} \sum_{j\in[J]} \mathbb{E}^{\mathcal{A}}_{\boldsymbol{s}^\star}[c_j N_{k,j}(T)] \frac{c_{j^*(k)}\sigma^2_{k,j^*(k)}}{c_j \sigma^2_{k,j}}$$
$$\leq \sum_{k\in[K]} \sum_{j\in[J]} \mathbb{E}^{\mathcal{A}}_{\boldsymbol{s}^\star}[c_j N_{k,j}(T)] \leq B.$$

Combining everything, we then have

$$\varepsilon_p^2 \geq \frac{1}{4C_1 B} \sum_{k\in[K]} c_{j^*(k)}\sigma^2_{k,j^*(k)} \mathrm{kl}(p^{(k)}, 1-\delta).$$

As the last step of the lower bound proof valid for *any* (B, δ, ℓ_p)-budget efficient algorithm, we must find the infimum of this lower bound over all valid algorithms. This reduces to finding the distribution of error probabilities $\{p^{(k)}\}_{k\in[K]}$ that minimizes the right-hand side, subject to the algorithm's success constraints. Specifically, we solve the following convex optimization:

$$V^* \coloneqq \min_{\{p^{(k)}\}_{k\in[K]}} \sum_{k\in[K]} c_{j^*(k)}\sigma^2_{k,j^*(k)} \mathrm{kl}(p^{(k)}, 1-\delta), \text{ subj. to } \sum_{k\in[K]} p^{(k)} \leq \delta,\ 0 \leq p^{(k)} \leq 1. \tag{2}$$

Although a closed-form solution cannot be found, we show in Appendix C.2 that a tractable lower bound is available as follows: when $\delta \leq e^{-2}$,

$$V^* \geq \left(\frac{1}{2} - \frac{1}{e^2}\right) \underbrace{\left(\sum_{k\in[K]} c_{j^*(k)}\sigma^2_{k,j^*(k)}\right)}_{=\mathcal{A}^*_\infty(\boldsymbol{\sigma},\boldsymbol{c})} \log\frac{1}{\delta}.$$

Combining everything and redefining C_1', we arrive at

$$\varepsilon_p^2 \geq \left(\frac{1}{2} - \frac{1}{e^2}\right) \frac{\log\frac{1}{\delta}}{4C_1 B} \mathcal{A}^*_\infty(\boldsymbol{\sigma}, \boldsymbol{c}).$$

□

Remark 1 (Implicit Inter-Query Allocation for $p \geq 2$)**.** *Unlike the proof for $1 \leq p < 2$, the high-probability lower bound for $p \geq 2$ does not immediately yield a closed-form optimal allocation across queries. The tightness of the knapsack inequality does dictate the* intra-query *strategy: an optimal algorithm must allocate its budget exclusively to the most cost-efficient judge, setting $N_{k,j}(T) = 0$ for all $j \neq j^*(k)$. However, the* inter-query *allocation remains implicit. Because the final bound is obtained by solving a convex optimization over the algorithm's error distribution over hard instances rather than its physical budget variables, the specific budget fraction allocated to each k is not explicitly resolved here. We will formally recover the closed-form optimal allocation across all $p \geq 1$ in the subsequent in-expectation lower bound.*

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 $k \in [K]$. Then our optimization becomes:

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

As $F(\cdot)$ is strictly increasing in $\|\boldsymbol{x}\|_p$, the minimum is necessarily obtained on the boundary of the constraint, i.e., 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 $\tilde{F}(\boldsymbol{y})$, which bifurcates at $p = 2$.

Case I: $1 \le p < 2$**.** 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 & Vandenberghe, 2004, Section 5.2.3).

We thus introduce Lagrangian multiplier vectors $\boldsymbol{\lambda} \in \mathbb{R}^K_{\ge 0}$ 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 $1 \le p < 2$ 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

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

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**. □

C.2. Lower Bound for Eqn. (2): Infimum over Algorithms when $p \geq 2$

For notational convenience, let us denote $g_k := c_{j^*(k)} \sigma_{k,j^*(k)}^2 \geq 0$ for each $k \in [K]$, and $G := \sum_{k \in [K]} g_k = \mathcal{A}_\infty^*(\boldsymbol{\sigma}, \boldsymbol{c})$. With this, let us recall Eqn. (2):

$$V^* = \min_{\{p^{(k)}\}} \sum_{k \in [K]} g_k \mathrm{kl}(p^{(k)}, 1-\delta), \text{ subj. to } \sum_{k \in [K]} p^{(k)} \leq \delta, \ 0 \leq p^{(k)} \leq 1.$$

Invoking the well-known lower bound of Bernoulli KL (Garivier et al., 2019, Eqn. (11)), we have that for any feasible $\{p^{(k)}\}_{k \in [K]}$,

$$\begin{aligned}
\sum_{k \in [K]} g_k \mathrm{kl}(p^{(k)}, 1-\delta) &\geq \sum_{k \in [K]} g_k \left[(1-p^{(k)}) \log \frac{1}{\delta} - \log 2\right] \\
&= \left(G - \sum_{k \in [K]} g_k p^{(k)}\right) \log \frac{1}{\delta} - G \log 2 \\
&\geq \left(G - \left(\max_k g_k\right) \sum_{k \in [K]} p^{(k)}\right) \log \frac{1}{\delta} - G \log 2 \\
&\geq (G - G\delta) \log \frac{1}{\delta} - G \log 2 && (\{p^{(k)}\}_{k \in [K]} \text{ is feasible}) \\
&= G \left[(1-\delta) \log \frac{1}{\delta} - \log 2\right].
\end{aligned}$$

Lastly, if $\delta \leq e^{-2}$, then we further have that

$$V^* \geq G \left[(1 - e^{-2}) - \frac{1}{2}\right] \log \frac{1}{\delta} = G \left(\frac{1}{2} - \frac{1}{e^2}\right) \log \frac{1}{\delta}.$$

□

D. Proof of Theorem 5.2: In-Expectation Lower Bounds

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

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

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

For both in-expectation lower bounds, we utilize the ***Assouad's lemma*** (Assouad, 1983; Yu, 1997)[9] combined with the Bretagnolle-Huber inequality (Bretagnolle & 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 $\bar{\mathbb{P}}_{i,k} := \frac{1}{2^{K-1}}\sum_{\boldsymbol{v}\in\mathcal{F}_K : v_k = i} \mathbb{P}_{\boldsymbol{v}}$ 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 $\boldsymbol{\Delta} := (\Delta_1, \ldots, \Delta_K) \in \mathbb{R}_{\ge 0}^K$ is chosen a-priori as described for each lower bound.

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

We, again, consider the canonical bandit model (Lattimore & 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 $V_k := \sum_{j\in[J]} \frac{\bar{N}_{k,j}}{\sigma_{k,j}^2}$, which is an algorithm-dependent quantity.

[9] 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 describe our a priori choice of $\boldsymbol{\Delta}$:

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

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 $\Delta_k \le \frac{R_k}{2}$ for all k's if

$$B \ge \left(\sum_{k\in[K]}\left(c_{j^*(k)}\sigma_{k,j^*(k)}^2\right)^{\frac{p}{p+2}}\right)\max_{k\in[K]}\frac{\left(c_{j^*(k)}\sigma_{k,j^*(k)}^2\right)^{\frac{2}{p+2}}}{4R_k^2},$$

as given in the statement.

Here, as $\mathbb{E}_{\boldsymbol{v}}\left[\|\hat{\boldsymbol{s}} - \boldsymbol{s}(\boldsymbol{v})\|_p\right]$ is not coordinate-wise decomposable, we cannot directly utilize Assouad's lemma (Lemma D.1).[10] Instead, we utilize a concavity argument as follows.

Let $S := \sum_{k\in[K]}(\Delta_k)^p\mathbb{1}[\hat{v}_k \neq v_k]$ be a random variable such that $S \le \|\boldsymbol{\Delta}\|_p^p$. By Assouad's lemma (Lemma D.1), its expectation is bounded as follows:

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

Note that from our choice of Δ_k's,

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

One important observation is that

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

By Jensen's inequality, we have that

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

[10] 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.

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,[11] 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

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

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

Here, 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 $\Delta_k \leq \frac{R_k}{2}$.

Let $\hat{\boldsymbol{s}}$ be any estimator, and let us denote $\hat{v}_k \coloneqq \operatorname{sign}(\hat{s}_k - s_k^\star)$. Then we first relate the absolute error of $\hat{s}_k$ to $(\boldsymbol{s}(\boldsymbol{v}))_k$ with the sign-error as follows:

Lemma D.3. *$|\hat{s}_k - (\boldsymbol{s}(\boldsymbol{v}))_k| \geq \Delta_k \mathbb{1}[\hat{v}_k \neq v_k]$ 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 $\hat{v}_k = -1 \neq v_k$. Then,

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

□

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

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

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

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

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 $v_k = i$.

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

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

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

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 $\Delta_k \leq R_k/4$. 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 $\Delta_k \leq R_k/4$ and $\max_j \sigma_{k,j}^2 \leq R_k^2/2$, 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 $X_k(\boldsymbol{\Delta}) \coloneqq s_k^\star - R_k + 2R_kY_k(\boldsymbol{\Delta})$. 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,

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

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: $\Delta_k \leq R_k/4$, which implies $|d| \leq 1/8$, and $\max_j \sigma_{k,j}^2 \leq R_k^2/2$, which implies $q \leq 1/8$. 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 $|d| \leq 1/8$ and $q \leq 1/8$, we can lower bound the common term:

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

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

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

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 $|d| \leq 1/8$ and $q \leq 1/8$ to the polynomial term yields:

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

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

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

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 & 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 $\psi_1(z) := \frac{d^2}{dz^2}\log\Gamma(z)$:

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

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

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

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:

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

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 $\varepsilon_{k,j} \sim \mathcal{N}(0, \sigma^2_{k,j})$.*

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 $\boldsymbol{\sigma} = (\sigma_{k,j})_{k,j} \in \mathbb{R}^{K\times J}_{>0}$

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

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

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

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 any given allocation. Then, **IVWE** is (B, δ, ℓ_p)-budget efficient at any $(\boldsymbol{s}, \boldsymbol{\sigma}, \boldsymbol{c}) \in [0,R]^K \times \mathbb{R}^K_{\geq 0} \times \mathbb{R}^K_{\geq 0}$ 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 $\boldsymbol{\omega} := \left(\omega_{k,j} \triangleq \frac{c_j N_{k,j}}{B}\right)_{k,j}$, 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, 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 $k \in [K]$. □

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 & Massart (2000))**.** *With $\hat{\sigma}^2_{k,j}$ as in line 3 of Algorithm 2 and $\epsilon_0 \triangleq \log \frac{2KJ}{\delta}$,*

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

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 $N_0 > 0$

`// 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 $\left\{s_{k,j}^{(n)}\right\}_{n \in [N_0]} \subset [a, b]$;

  Compute the empirical mean and variance estimator

$$\hat{s}_{k,j} := \frac{1}{N_0} \sum_{n=1}^{N_0} s_{k,j}^{(n)}, \quad \hat{\sigma}_{k,j}^2 := \frac{1}{N_0 - 1} \sum_{n=1}^{N_0} \left(s_{k,j}^{(n)} - \hat{s}_{k,j}\right)^2 \tag{3}$$

`// Phase II: IVWE`

Compute the optimal allocation $\left(\widehat{N}_{k,j}^*\right)_{k,j}$ using the remaining budget $B' \triangleq B - N_0 K \sum_{j \in [J]} c_j$, based on the variance estimators $\left\{(\hat{\sigma}_{k,j})^2\right\}_{k,j}$;

return IVWE with $\left(\widehat{N}_{k,j}^*\right)_{k,j}$ as allocation and $\hat{\sigma}_{k,j}^2$ 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 $N_0 = 1 + \left\lceil 16 \log \frac{4KJ}{\delta} \right\rceil$ and suppose that $B \geq N_0 K \sum_{j \in [J]} c_j$. 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:*

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

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 $p \in [1, \infty]$, $B > 0$, $\delta \in (0, 1/2]$, 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. □

Again, the high-probability lower bound fails to yield the correct allocation objective, and thus, we turn to global in-expectation lower bounds, which involves quite different proof techniques that may be of independent interest.[12] 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 $\boldsymbol{\sigma}$.

[12] 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.

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

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

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

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

where[a] $m_p \coloneqq \sqrt{2} \left(\frac{\Gamma((p+1)/2)}{\sqrt{\pi}} \right)^{\frac{1}{p}} = \mathbb{E}[|Z|^p]^{1/p}$ *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. □

We show that the above is actually constant-wise tight(!) across *all* p's by complementing this with a matching in-expectation upper bound when *variances are known*:

Theorem F.7. *When the variances are* known*,* ***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 $\hat{s}_k - s_k \sim \mathcal{N}\left(0, \left(\sum_{j=1}^J \frac{N_{k,j}}{\sigma^2_{k,j}}\right)^{-1}\right)$. Thus, we have that

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

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 & Levit, 1995; van Trees, 1968), recently extended to L_p-norm by Chen & 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 $\sigma^2_{k,j}$ after the exploration phase by

$$\underline{\sigma}^2_{k,j} = \frac{\hat{\sigma}^2_{k,j}}{1 + 2\sqrt{\frac{\epsilon_0}{n_0-1}} + \frac{2\epsilon_0}{n_0-1}} \quad \text{and} \quad \overline{\sigma}^2_{k,j} = \frac{\hat{\sigma}^2_{k,j}}{1 - 2\sqrt{\frac{\epsilon_0}{n_0-1}}},$$

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

$$\mathbb{P}\left(\underline{\sigma}^2_{k,j} \le \sigma^2_{k,j} \le \underline{\sigma}^2_{k,j}\right) \ge 1 - \frac{\delta}{2},$$

which we will assume to hold throughout the proof. We denote the ratio between the upper and lower bounds as $\kappa = \overline{\sigma}^2_{k,j}/\underline{\sigma}^2_{k,j}$, 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 $\hat{j}^*(k) \triangleq \arg\min_{j\in[J]} c_j \hat{\sigma}^2_{k,j}$ for each query $k \in [K]$, then computes the optimal allocation with variance estimator $\hat{\sigma}^2_{k,j}$ and remaining budget $B' = B - CN_0$ with $C \coloneqq K \sum_{j\in[J]} c_j$ as follows:

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

Since $\hat{s}_k$ is the sample mean of Gaussian samples, $\hat{s}_k$ itself is also Gaussian. Thus, by Lemma F.1, we have that

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

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 $1 - \delta$:

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

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:

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

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 everything, we conclude as follows: with probability at least $1-\delta$,

$$\begin{aligned}
\varepsilon_p &\le \sqrt{\frac{2\kappa\mathcal{A}^*_p(\boldsymbol{\sigma},\boldsymbol{c})\log\frac{4K}{\delta}}{B'}} \\
&\le \sqrt{13\mathcal{A}^*_p(\boldsymbol{\sigma},\boldsymbol{c})\log\frac{4K}{\delta}}\left(\frac{1}{\sqrt{2B}}+\frac{CN_0}{\sqrt{2}B^{\frac{3}{2}}}\right) \qquad \text{(Lemma B.3, } \kappa \le 13/4) \\
&= \sqrt{\frac{13\mathcal{A}^*_p(\boldsymbol{\sigma},\boldsymbol{c})\log\frac{4K}{\delta}}{2B}} + \underbrace{\frac{C\left(1+\lceil 16\log\frac{4KJ}{\delta}\rceil\right)\sqrt{\frac{13}{2}\mathcal{A}^*_p(\boldsymbol{\sigma},\boldsymbol{c})\log\frac{4K}{\delta}}}{B^{\frac{3}{2}}}}_{=\mathcal{O}\left(\frac{\sqrt{\mathcal{A}^*_p(\boldsymbol{\sigma},\boldsymbol{c})}\left(\log\frac{KJ}{\delta}\right)^{\frac{3}{2}}}{B^{\frac{3}{2}}}\right)}.
\end{aligned}$$

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 $\lambda > 0$,

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

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 $\tilde{s}_t := s_{k_t} + \varepsilon_{k_t, j_t}$. Let us denote $N_{k,j} := \sum_t \mathbb{1}[k_t = k, j_t = j]$ 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 $\mathcal{H}_t := (k_1, j_1, \tilde{s}_1, \cdots, k_t, j_t, \tilde{s}_t)$ 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*

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

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 $Z := S - \mathbb{E}[\boldsymbol{s} \mid \mathcal{H}_T]$. 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

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

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

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

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.[13]

Combining this with Lemma G.1, we have that

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

When $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_p$, by Jensen's inequality,[14]

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

[13] 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.

[14] $\boldsymbol{g} \mapsto \|\boldsymbol{u} \odot \boldsymbol{g}\|_p$ 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:

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

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 $\mu_k := \mathcal{I}_k(\mathcal{A})^{-1}\sum_j\frac{N_{k,j}}{\sigma_{k,j}^2}\bar{s}_{k,j}$. 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 & Levit, 1995; van Trees, 1968) to L_p-norm due to Chen & Özgür (2024) via Efroimovich's inequality (Efroimovich, 1979):

Lemma G.2 (Scalar Version of Theorem 2.4 of Chen & 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*

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

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 & 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 $\boldsymbol{s}$,

$$\inf_{\mathcal{A}}\sup_{\boldsymbol{s}}\mathbb{E}_{\boldsymbol{s}}^{\mathcal{A}}\left[\|\hat{\boldsymbol{s}}-\boldsymbol{s}\|_p^p\right]\ge\inf_{\mathcal{A}}\mathbb{E}_{\boldsymbol{s}\sim\mu}\left[\mathbb{E}_{\boldsymbol{s}}^{\mathcal{A}}\left[\|\hat{\boldsymbol{s}}-\boldsymbol{s}\|_p^p\right]\right]$$

$$= \inf_{\mathcal{A}} \sum_k \mathbb{E}_{s_k \sim \mu_k} \left[\mathbb{E}^{\mathcal{A}}_{s_k} \left[|\hat{s}_k - s_k|^p \right] \right],$$

where the algorithm $\mathcal{A}$ must satisfy the knapsack constraint: if $\mathcal{A}$ samples each (k, j) pair $N_{k,j}$ times, then $\sum_j c_j \sum_k N_{k,j} \leq B$. 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 $r > 0$:[15]

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

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

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

[15] 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.

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

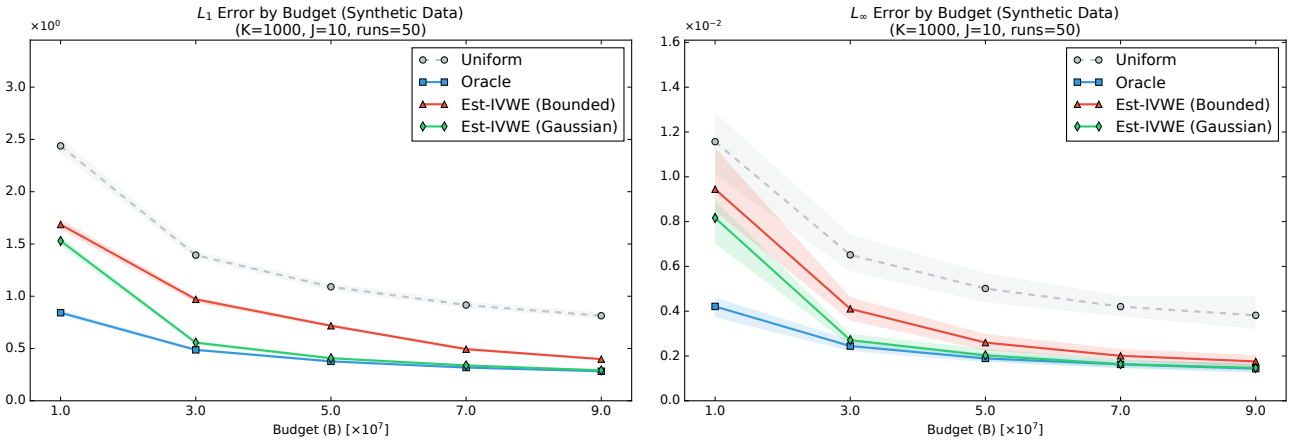
H. Additional Experiments

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 $[10^{-4}, 0.9s_k(1-s_k)]$. 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 $[0.5, 1.5]$.

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 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, $j^*(k) = \arg\min_{j\in[J]} c_j\sigma^2_{k,j}$, 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 $\sigma^2_{k,j}$ 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 $\sigma^2_{k,j}$ tends to be. In *Bad is Expensive* scenario, the cost is set to be proportional to $\sum_{k\in[K]} \sigma^2_{k,j}$; while in *Bad is Cheap* scenario, the cost is set to be proportional to $\sum_{k\in[K]} \sigma^{-2}_{k,j}$. In both cases, the costs are normalized so that the mean cost over J judges becomes 0.1.

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 $\{c_j \sigma^2_{k,j}\}_{j\in[J]}$ 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.

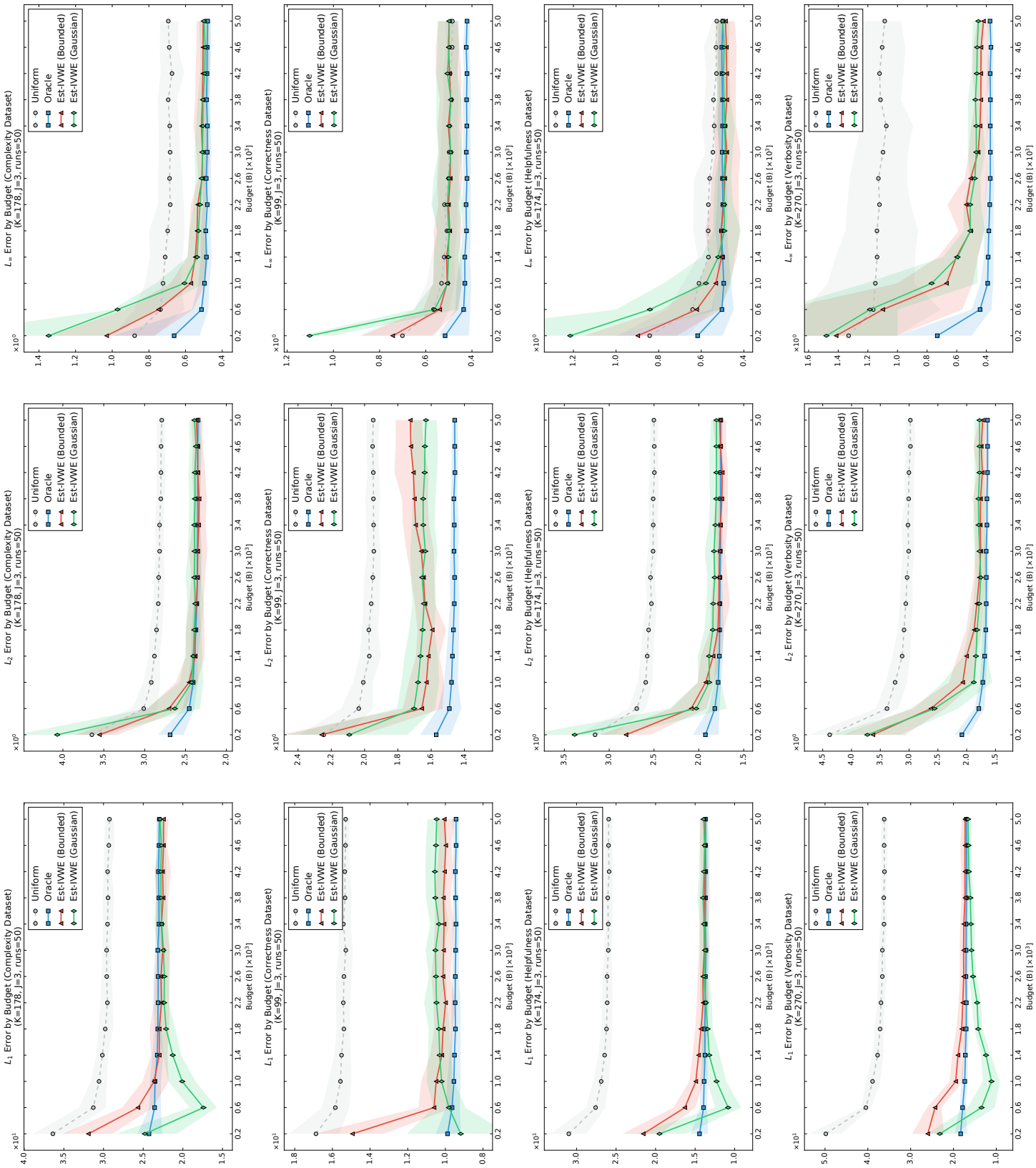


Figure 3. Full experimental results for the datasets (Complexity, Correctness, Helpfulness, Verbosity). Each column corresponds to an error metric ($\ell_1, \ell_2, \ell_\infty$).

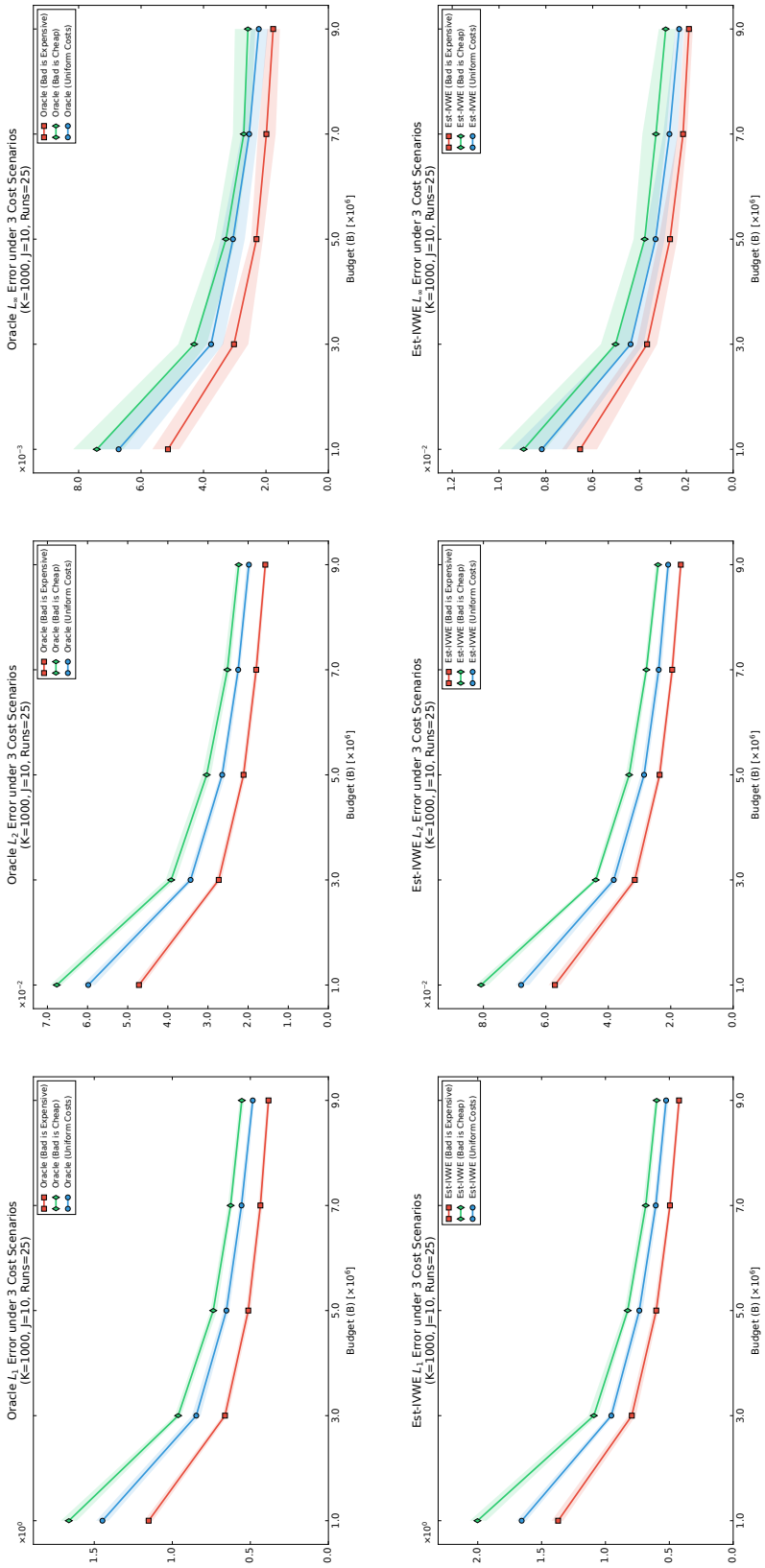


Figure 4. Impact of Cost Structure on Performances of **Est-IVWE** and **Oracle**