Asymptotic Optimality Theory of Confidence Intervals of the Mean

Vikas Deep
NUS, Singapore

Achal Bassamboo
Kellogg, Northwestern University

Sandeep Juneja
Ashoka University, India

Abstract

We address the classical problem of constructing confidence intervals (CIs) for the mean of a distribution, given N i.i.d. samples, such that the CI contains the true mean with probability at least $1-\delta$, where $\delta \in (0,1)$. We characterize three distinct learning regimes based on the minimum achievable limiting width of any CI as the sample size $N_\delta \to \infty$ and $\delta \to 0$. In the first regime, where N_δ grows slower than $\log(1/\delta)$, the limiting width of any CI equals the width of the distribution's support, precluding meaningful inference. In the second regime, where N_δ scales as $\log(1/\delta)$, we precisely characterize the minimum limiting width, which depends on the scaling constant. In the third regime, where N_δ grows faster than $\log(1/\delta)$, complete learning is achievable, and the limiting width of the CI collapses to zero and CI converges to the true mean. We demonstrate that CIs derived from concentration inequalities based on Kullback-Leibler (KL) divergences achieve asymptotically optimal performance, attaining the minimum limiting width in both the sufficient and the complete learning regimes for distributions in three families: single-parameter exponential, bounded support and known bound on $(1+\epsilon)^{\text{th}}$ moment. Additionally, these results extend to one-sided CIs, with the width notion adjusted appropriately. Finally, we generalize our findings to settings with random per-sample costs, motivated by practical applications such as stochastic simulators and cloud service selection. Instead of a fixed sample size, we consider a cost budget C_δ, identifying analogous learning regimes and characterizing the optimal CI construction policy.

Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300.

1 INTRODUCTION AND MAIN CONTRIBUTION

The problem of constructing a confidence interval (CI) for the mean of a distribution with a coverage guarantee, ensuring that the mean lies within the CI with probability at least $1-\delta$ for a pre-specified $\delta \in (0,1)$, is well-studied in the statistics literature. This problem has significant applications in A/B testing, experimentation, data analytics, and simulation. Typically, this is achieved using concentration inequalities for the mean, given a sample size of N i.i.d. observations from a probability distribution (see Hoeffding (1994), Waudby-Smith and Ramdas (2024), Maurer and Pontil (2009), Audibert et al. (2009), Boucheron and Gassiat (2009), Catoni (2012), Chen et al. (2021) and Bennett (1962)). In this paper, we primarily aim to address an important element absent from the literature on CI. *While there are various methods to construct a confidence interval for the mean of a distribution, there are no results characterizing the optimal CI with minimum width.* In this paper, we characterize three learning regimes based on the minimum limiting width achievable for any CI construction method/policy as $N_\delta \to \infty$ when $\delta \to 0$, under a mild assumption of stability of policies (as defined in Definition 2). We would like to emphasize that we consider CI construction policies that provide non-asymptotic coverage guarantees for any fixed N and δ. The only aspect where we rely on asymptotics is in defining our notion of optimality. In a nutshell, under this assumption, the limiting width of CI of mean becomes a deterministic constant for any CI construction method making analysis tractable. The three regimes are defined as follows:

1) **No learning regime:** If $\lim_{\delta\to 0} \frac{N_\delta}{\log(1/\delta)} \to 0$, the limiting width of the CI is the length of the support of the mean for any stable CI construction policies. This implies that no learning is possible as the sample size is not sufficient.

2) **Sufficient learning regime:** If $\lim_{\delta\to 0} \frac{N_\delta}{\log(1/\delta)} \to$

k for $k \in (0, \infty)$, we have a sharp characterization of the lower bound on limiting width that involves terms with KL divergences for any stable CI construction policies. This lower bound on limiting width shrinks as k increases. Further, we show that the method π_1 (see section 5) that constructs CI via inverting concentration inequality based on KL divergences matches the lower bound of the limiting width. Hence, this lower bound result on the limiting width is tight as $\delta \to 0$ and $N_\delta \to \infty$.

3) **Complete learning regime:** If $\lim_{\delta \to 0} \frac{N_\delta}{\log(1/\delta)} \to \infty$, the limiting CI width trivially converges to zero. Furthermore, we prove that the method π_1 has a zero limiting width. In this regime, we further analyze the rate at which the width converges to zero and demonstrate that π_1 achieves the fastest rate of convergence under some technical assumptions.

Our results apply to both parametric and non-parametric families of probability distributions. For parametric cases, we consider a canonical single-parameter exponential family, while for non-parametric cases, we consider two such settings. The first one is the family of probability distributions with bounded support and the second one is the family of probability distributions with $(1+\epsilon)^{\text{th}}$ moment bounded and the bound is known. We assume only that the unknown underlying distribution belongs to this known family. If a CI construction method, that matches the lower bound on the limiting width in the sufficient learning regime and the complete learning regime, is called an asymptotically optimal CI construction method. Next, we state the recipe for asymptotically optimal CI construction policies.

1.1 Recipe for asymptotically optimal CI construction methods

To construct an asymptotically optimal CI construction method, when the probability distribution belongs to a canonical single-parametric exponential family, we utilize existing concentration bound in the literature on the deviation of the maximum likelihood estimator (MLE) of the mean from the true mean, measured in terms of the KL divergence (see Eq. (4)). Inverting this concentration inequality to construct a CI for the mean leads to an asymptotically optimal CI construction method. For the non-parametric setting, Agrawal (2022) and Orabona and Jun (2023) provide a confidence interval that is constructed via inverting a concentration inequality based on KL divergences. In this setting, we prove the optimality of the CI provided by them. In summary, we prove that the CIs constructed via inverting concentration inequalities based on KL divergences are the best CIs in a reasonable asymptotic regime.

1.2 Random sampling cost

The above results assume that the cost of each sample is fixed and equal to one. However, in many practical situations, this assumption may not hold. For instance, in simulation studies, the cost can represent the time required to simulate the performance of a system design, which may vary depending on system complexity and randomness (see Chick and Inoue (2001)). Similarly, in real-world applications, online bidding optimization in sponsored search (see Amin et al. (2012), Xia et al. (2015) and Tran-Thanh et al. (2014)), the cost of acquiring each sample can fluctuate due to factors such as network delays, resource availability, or dynamic market conditions. To include such cases, we extend our results to a setting where each sample is associated with a random cost drawn from an i.i.d. cost distribution. In this scenario, instead of having a fixed sample size N_δ, the constraint is a cost budget C_δ. We similarly identify three corresponding learning regimes and characterize the asymptotically optimal method for this setting as well. When both the rewards and costs of obtaining samples are random, the large deviations of the observed average reward for a given cost budget typically depend on the distributions of both variables (see Glynn and Juneja (2013)). However, in our analysis, we find that the limiting width of the CI depends only on the mean of the cost distribution and is otherwise invariant to the cost distribution.

In summary, our main contributions are as follows:

- We characterize three distinct learning regimes: no learning, sufficient learning, and complete learning, based on the relative scaling of the sample size N_δ with respect to the desired accuracy $1-\delta$. These regimes lead to markedly different minimum asymptotic limiting widths for any CI construction method, under a mild policy stability assumption.
- We derive sharp lower bounds on the limiting CI width in the sufficient learning regime and demonstrate that CIs constructed by inverting existing concentration inequalities involving KL divergences achieve these bounds.This result extends to the complete learning regime, where the CI width converges to zero, establishing that KL divergence-based CI construction is asymptotically optimal in our setting. In the complete learning regime, we further analyze the rate at which the width converges to zero and show that our proposed policy achieves the fastest rate of convergence under some technical assumptions. We emphasize that while CIs constructed by inverting KL-based concentration bounds are known in the

literature, their optimality was not established. A key contribution of our work is demonstrating this optimality.

- We extend our results to more general settings, including random sampling costs, one-sided CIs, and non-parametric distributions. In each of these cases, we again characterize the three learning regimes and an asymptotically optimal CI construction policy. Intriguingly, in the random sampling cost setting, we find that the limiting CI width depends solely on the mean of the cost distribution.

2 LITERATURE REVIEW

There is a well-established duality between hypothesis testing and the construction of confidence intervals/regions (see Bickel and Doksum (2015)). In parametric settings, one can invert hypothesis tests to obtain confidence intervals. Classical theory shows that for a fixed sample size and a prescribed Type I error, there exist UMP (uniformly most powerful) tests that maximize power, often restricting attention to particular families (e.g., such as unbiased tests in exponential families). This optimality, when translated to CIs, is usually framed as minimizing expected width, subject to some restriction on the class of CI, such as unbiasedness (See Definition 9.3.7 in Casella and Berger (2024)). Further, Classic theory indicates that no CI can uniformly minimize random width at fixed N and δ (see page 250 in Bickel and Doksum (2015)). To overcome this negative result, we use asymptotic analysis. Our framework adopts a stability assumption, ensuring that the interval endpoints converge to deterministic values as $\lim_{\delta\to 0} N_\delta \to \infty$. This makes the limiting width a well-defined metric and allows us to show that no stable policy can outperform our KL-based intervals, asymptotically. In the context of inverting a test statistic, existing results often allow for thresholds that depend intricately on sample size or distributional assumptions. The classical literature, for the most part, focuses on parametric families. In contrast, we show that a uniform threshold can be used in KL-divergence-based constructions for both parametric and nonparametric families with bounded support, yielding asymptotically optimal intervals.

Additionally, there is a vast amount of literature on constructing CIs that proceeds by inverting finite-sample concentration inequalities. For the probability distributions with bounded support, one can use various concentration inequalities such as the Hoeffding and Bernstein inequalities (see Hoeffding (1994), Bennett (1962), Boucheron and Gassiat (2009), Audibert et al. (2009) and Maurer and Pontil (2009)). In general, using the Chernoff bound, one can construct a CI of the mean if the upper bound on MGF is known. For the heavy tail distributions, one can get a CI of the mean using Markov's inequality and Chebyshev's inequality (see Catoni (2012) and Chen et al. (2021)). Bootstrap methods (Efron and Tibshirani, 1994) are a popular approach for constructing confidence intervals, but their coverage guarantees are only asymptotically valid as $N \to \infty$. In our setting, we require exact finite-sample guarantees for any fixed N and δ, which standard bootstrap methods cannot provide.

Recently, Waudby-Smith and Ramdas (2024) transform a CI construction problem into a betting problem for the wealth maximization process for bounded probability distributions. A recent work Gupta et al. (2023) utilizes a different notion of optimality (mini-max) for the width of CI in location parameter families. We compare our policy's performance with some of these relevant existing methods in Section 8.

There appears to be only one work in the literature that aims to characterize a lower bound on the width of a CI in terms of a distribution-dependent complexity term: Shekhar and Ramdas (2023). As noted in Remark 4.4 in Shekhar and Ramdas (2023): "To the best of our knowledge, this is the first result providing an explicit characterization of the smallest achievable width of CI in terms of a distribution-dependent complexity term."

However, their bound is loose (in the sense that we strengthen it by a multiplicative constant asymptotically) and relies on the strong assumption that the width of the CI is deterministically bounded by a specific function for any given sample size N. In contrast, our lower bound on the limiting width is asymptotic but **remains tight** (in the sense the proposed CI construction policy matches the lower bound asymptotically), under a much milder assumption concerning the stability of the policies (see Section 8).

The paper is organized as follows: In the next section, we formally introduce the setup and notation and focus on the single-parameter exponential family. In Section 4, we present our results on the minimum limiting width for CIs in the three learning regimes: no learning, sufficient learning, and complete learning. In Section 5, we provide a recipe for constructing CIs using KL divergence-based concentration inequalities (see (4)) that achieve the minimum limiting width in the sufficient learning regime and zero limiting width in the complete learning regime. In Section 6, we extend our results to the setting when there is a random cost associated with the sample collection. In Section 7, we extend our results to the non-parametric distribution families with bounded support and known bound on $(1+\epsilon)^{\text{th}}$ moment. We also discuss the asymptotic opti-

mality of KL divergences-based CI construction policies in these settings. In Section 8, we present our numerical studies. In Section 9, we present our conclusions and outline future research directions. In Appendix I, we extend our results to the setting when we want to construct a one-sided CI of the mean.

3 SETUP AND NOTATION

Let $X_1, X_2, \ldots$ be i.i.d. copies of a random variable X with distribution ν and mean $m(\nu) = \mathbb{E}_\nu[X] = \mu$. For $n \geq 1$, let $\mathcal{F}_n$ denote the information contained in the σ-algebra generated by $\{X_k, k \leq n\}$. We now state a formal definition of confidence intervals.

3.1 Description of the policy space for construction of CI

Let Π_{CI} denote the collection of methods/policies for constructing CIs. Let $[\widehat{\mu}_L^\pi(N,\delta), \widehat{\mu}_R^\pi(N,\delta)]$ denote the estimated CI after observing $X_1, X_2, \ldots, X_N$ under a policy π for any given $\delta \in (0,1)$. For any policy $\pi \in \Pi_{\text{CI}}$ must satisfy the following for any given $\delta \in (0,1)$,

$$\forall n \in \mathbb{N} : \mathbb{P}_\nu(\mu \in [\widehat{\mu}_L^\pi(n,\delta), \widehat{\mu}_R^\pi(n,\delta)]) \geq 1 - \delta.$$

Here, $\mathbb{P}_\nu(\cdot)$ denotes the probability measure induced by the environment ν. Importantly, the above coverage requirement is non-asymptotic: it must hold for every fixed N and δ.

Recall that our objective is to characterize the width of CI, i.e., $\widehat{\mu}_R^\pi(N,\delta) - \widehat{\mu}_L^\pi(N,\delta)$. As stated in the Literature review section, for a given N and δ, a tight characterization of the width of CI is analytically intractable, and thus we consider the asymptotic regime where $N \to \infty$ as $\delta \to 0$ and aim to characterize the limiting width of CIs. Henceforth, we denote N as N_δ. To start the analysis, we first make a stability assumption over the space of policies which construct CIs, enabling us to derive a lower bound on the limiting width. Under this assumption, the limiting width of CIs becomes a deterministic constant for any CI construction policy. This assumption requires that for a given environment ν, the boundaries of the CI, $\widehat{\mu}_L^\pi(N_\delta,\delta), \widehat{\mu}_R^\pi(N_\delta,\delta)$ converge in probability to deterministic points as δ approaches 0 and N_δ approaches infinity. In the Appendix, we show that many popular policies for constructing CI are stable.

Definition 1 (Stability)**.** *Let $N_\delta \to \infty$ as $\delta \to 0$. For a given distribution ν with mean μ, a policy $\pi \in \Pi_{\text{CI}}$ is called **stable** if the CI it constructs, denoted by $[\widehat{\mu}_L^\pi(N_\delta,\delta), \widehat{\mu}_R^\pi(N_\delta,\delta)]$, satisfies: if*

$$\lim_{\delta \to 0} \frac{N_\delta}{\log(1/\delta)} \in [0, \infty],$$

then

$$\widehat{\mu}_L^\pi(N_\delta,\delta) \xrightarrow{p} \mu_L^\pi(\nu) \qquad \text{and} \qquad \widehat{\mu}_R^\pi(N_\delta,\delta) \xrightarrow{p} \mu_R^\pi(\nu),$$

where $\mu_L^\pi(\mu) \leq \mu$ and $\mu_R^\pi(\mu) \geq \mu$ are deterministic constants (with $-\infty$ and ∞ included).

We denote the collection of policies in the set Π_{CI} which are stable as Π_{CI}^s. It is worth noting that $\mu_R^\pi(\nu) - \mu_L^\pi(\nu)$ denotes the limiting width of CI for $\pi \in \Pi_{\text{CI}}^s$ in the asymptotic regime where the sample size, i.e, N_δ, scales to ∞ as $\delta \to 0$. We now assume that ν belongs to the canonical single-parameter exponential family **S**. In Section 7, we later generalize our results to non-parametric distributions.

Specifically, **S** is defined as: $\mathbf{S} = \left\{p_\theta : \theta \in \Theta \subseteq \mathbb{R}, \frac{dp_\theta(x)}{d\xi} = \exp(\theta \cdot x - b(\theta))\right\}$, where ξ is a fixed reference measure on $\mathbb{R}$, and $b(\theta)$ is a known, twice-differentiable, strictly convex function. The set Θ is: $\left\{\theta \in \mathbb{R} : \int_{\mathbb{R}} |x| \exp(\theta \cdot x) d\xi(x) < \infty\right\}$. This condition ensures that p_θ forms a well-defined probability distribution with a finite mean. We assume $\Theta = (\underline{\theta}, \overline{\theta})$ is an open interval. Common distributions in this family include Bernoulli, Poisson, Gaussian (with known variance), and Gamma (with a known shape parameter) (see Cappé et al. (2013) for more details on single-parameter exponential family). For $\nu = p_\theta \in \mathbf{S}$, the mean $\mu = \mathbb{E}_\nu[X]$ is unknown. It is known that $\mu(\theta) = b'(\theta)$, which, due to the strict convexity of $b(\theta)$, is a strictly increasing function. This allows us to define the inverse function $\theta(\mu)$. We assume that the support of the mean is denoted by $\mathbf{O} = (\underline{\mu}, \overline{\mu})$.

Divergence function: Let $KL(p_\theta, p_{\tilde{\theta}})$ denote the KL divergence. We define:

$$\begin{aligned} d(\mu, \tilde{\mu}) &= KL(p_{\theta(\mu)}, p_{\theta(\tilde{\mu})}) \\ &= b(\theta(\tilde{\mu})) - b(\theta(\mu)) - b'(\theta(\mu))\,(\theta(\tilde{\mu}) - \theta(\mu)). \end{aligned}$$

for $\mu, \tilde{\mu} \in \mathbf{O}$. Key properties of $d(\mu, \tilde{\mu})$ include its strict quasi-convexity in the second argument and the fact that $d(\mu, \tilde{\mu}) > 0$ for all $\tilde{\mu} \neq \mu$, with $d(\mu, \mu) = 0$. It is worth noting that, since for $\nu \in \mathbf{S}$, the distribution is uniquely determined by its mean μ. Therefore, for simplicity, we denote the limiting CI of a $\pi \in \Pi_{\text{CI}}^s$, $[\mu_L^\pi(\nu), \mu_R^\pi(\nu)]$, as $[\mu_L^\pi(\mu), \mu_R^\pi(\mu)]$.

4 MAIN RESULTS ON THE MINIMUM LIMITING WIDTH OF CONFIDENCE INTERVAL IN DIFFERENT REGIMES

Recall that ν, with mean μ, represents the true distribution from which samples are generated. In other

words, ν is the true underlying but unknown environment. To start the analysis of the minimum limiting width, we define an alternate environment $\tilde{\nu}$ such that $\mathbb{E}_{\tilde{\nu}}[X] = \tilde{\mu} \neq \mu$. Using the data processing inequality (see Cover and Thomas (1991)), it follows that for a given $\delta \in (0,1)$ and any alternate environment $\tilde{\nu}$ and any event $\mathcal{E}_\delta$, we have

$$N_\delta \cdot d(\mu, \tilde{\mu}) \geq \sup_{\mathcal{E}_\delta \in \mathcal{F}_{N_\delta}} \phi(\mathbb{P}_\nu(\mathcal{E}_\delta), \mathbb{P}_{\tilde{\nu}}(\mathcal{E}_\delta)), \tag{1}$$

where $\phi(p_1, p_2) \triangleq p_1 \log \frac{p_1}{p_2} + (1-p_1) \log\left(\frac{1-p_1}{1-p_2}\right)$ for $p_1, p_2 \in (0,1)$. To utilize the above result, consider any policy $\pi \in \Pi^s_{\text{CI}}$. We now define the set of alternate environments $K(\mu^\pi_L(\mu), \mu^\pi_R(\mu)) = K_1(\mu^\pi_L(\mu)) \cup K_2(\mu^\pi_R(\mu))$, where $K_1(\mu^\pi_L(\mu)) = \{\tilde{\nu} : \tilde{\nu} \in \mathbf{S}, \tilde{\mu} < \mu^\pi_L(\mu)\}$ and $K_2(\mu^\pi_R(\mu)) = \{\tilde{\nu} : \tilde{\nu} \in \mathbf{S}, \tilde{\mu} > \mu^\pi_R(\mu)\}$. We utilize (1) for $\tilde{\nu} \in K(\mu^\pi_L(\mu), \mu^\pi_R(\mu))$ and $\mathcal{E}_\delta = \{\tilde{\mu} \notin [\widehat{\mu}^\pi_L(N_\delta, \delta), \widehat{\mu}^\pi_R(N_\delta, \delta)]\}$, where recall that $\tilde{\mu} = m(\tilde{\nu})$. Using the fact that $\pi \in \Pi^s_{\text{CI}}$, we get that $\mathbb{P}_{\tilde{\nu}}(\mathcal{E}_\delta) \leq \delta$. Since $\tilde{\nu} \in K(\mu^\pi_L(\mu), \mu^\pi_R(\mu))$, it implies that $\tilde{\mu} > \mu^\pi_R(\mu)$ or $\tilde{\mu} < \mu^\pi_L(\mu)$. Further, as π is a stable policy and hence it implies $\mathbb{P}_\nu(\mathcal{E}_\delta) \approx 1$ for small δ. Using (1) and dividing both sides with $\log(1/\delta)$ and taking $\delta \to 0$, we get different lower bounds on the limiting width depending upon the scaling of N_δ. We now present the formal result and the rigorous proof is given in the Appendix A.

Theorem 1. *For a given $\nu \in \mathbf{S}$ with mean μ, and any $\pi \in \Pi^s_{\text{CI}}$, the following holds:*

a) ***No learning regime :*** *If $\lim_{\delta\to 0} \frac{N_\delta}{\log(1/\delta)} \to 0$ then, $[\mu^\pi_R(\mu) - \mu^\pi_L(\mu)] = \overline{\mu} - \underline{\mu}$. Further, $\lim_{\delta\to 0} \widehat{\mu}^\pi_L(N_\delta, \delta) \xrightarrow{p} \underline{\mu}$ and $\lim_{\delta\to 0} \widehat{\mu}^\pi_R(N_\delta, \delta) \xrightarrow{p} \overline{\mu}$.*

b) ***Sufficient learning regime :*** *If $\lim_{\delta\to 0} \frac{N_\delta}{\log(1/\delta)} \to k$ for $k \in (0, \infty)$, then we have,*

$$[\mu^\pi_R(\mu) - \mu^\pi_L(\mu)] \geq \mu^*_R(\mu, k) - \mu^*_L(\mu, k), \tag{2}$$

where, $\mu^*_L(\mu, k) < \mu$ and $\mu^*_R(\mu, k) > \mu$ uniquely solve the following system of equations,

$$d(\mu, \mu^*_R(\mu, k)) = d(\mu, \mu^*_L(\mu, k)) = \frac{1}{k}. \tag{3}$$

For the case when the sample size scales at the rate of $\log(1/\delta)$, as $\delta \to 0$, we obtain a lower bound on the limiting width of the CI, given by $\mu^*_R(\mu, k) - \mu^*_L(\mu, k)$. It follows from the quasi-convexity of $d(\mu, x)$ in x, implying that $\mu^*_R(\mu, k) - \mu^*_L(\mu, k)$ decreases as k increases. In the proof, we first demonstrate that $\mu^\pi_R(\mu) \geq \mu^*_R(\mu, k)$ for any $\pi \in \Pi^s_{\text{CI}}$. We then show that $\mu^\pi_L(\mu) \leq \mu^*_L(\mu, k)$ for any $\pi \in \Pi^s_{\text{CI}}$. The key idea of the proof is that for $\mu^\pi_L(\mu) > \mu^*_L(\mu, k)$ and $\mu^\pi_R(\mu) < \mu^*_R(\mu, k)$, we get a contradiction with (1). Hence, $[\mu^*_L(\mu, k), \mu^*_R(\mu, k)]$ can be interpreted as a subset of the CI constructed by any policy $\pi \in \Pi^s_{\text{CI}}$ in the sufficient learning regime. Later, we show that our proposed policy π_1 has $[\mu^*_L(\mu, k), \mu^*_R(\mu, k)]$ as the limiting CI in the sufficient learning regime, proving that the above lower bound on the limiting width is tight.

Our main novelty and contribution in the lower bound proof lies in the introduction of the stability notion for CI construction methods, and in further leveraging this concept together with (1) (data processing inequality). This stability concept arises very naturally in the asymptotic regime we study, which itself has not been explored in the context of confidence interval width. Importantly, the stability assumption is very mild, as it is trivially satisfied by standard concentration bound-based methods, such as those based on Hoeffding or empirical Bernstein inequalities (see Appendix).

Further, due to the stability notion, we obtain a tighter lower bound than that of Proposition 4.3 of Shekhar and Ramdas (2023). To be precise, the authors in [4] derive one-sided lower bounds on the width using a similar argument to the data processing inequality, and then take the maximum of the left and right deviations. Ideally, one should be able to combine the lower bounds on the width from the left and right deviations, but unfortunately, it is not possible trivially. However, the stability assumption allows us to consider joint elimination of hypotheses on both sides, leading to a tighter bound that equals the total limiting width. This results in a factor of two improvement in the Gaussian case (as shown in Section 8). Thus, while our proof begins with a known inequality, the strength of our result comes from this careful asymptotic characterization using stability, which leads to tighter bounds.

4.1 Complete learning regime

It is worth noting that for the case when $\lim_{\delta\to 0} \frac{N_\delta}{\log(1/\delta)} \to \infty$, similar to the proof of above theorem, we get a trivial lower bound on the limiting width, i.e., $[\mu^\pi_R(\mu) - \mu^\pi_L(\mu)] \geq 0$ for any $\pi \in \Pi^s_{\text{CI}}$. Further, our proposed policy π_1 (see Section 5) has zero limiting width in this regime. Hence, we denote this regime as the complete learning regime. Additionally, in the complete learning regime, we characterize the rate at which the CI width converges to zero. Under certain technical assumptions on CI construction policies, we establish the fastest achievable convergence rate. Furthermore, we demonstrate that the CI width under our proposed policy π_1 attains this optimal rate as it approaches zero (see Appendix C for formal results).

Remark 1. *It is worth noting that, for any theoretical optimality guarantee for a CI procedure, one can think of four distinct formulations: (i) fixed N and δ;*

(ii) fixed N and $\delta \to 0$; (iii) fixed δ and $N \to \infty$; and (iv) $N \to \infty$ and $\delta \to 0$ jointly. Since classical theory shows that no CI can uniformly minimize random width for fixed N and δ (see Section 2), formulation (i) does not yield meaningful results. Formulations (ii) and (iii) are relatively trivial: if only $\delta \to 0$ while N remains fixed, the results fall into the no-learning regime; if only $N \to \infty$ while δ remains fixed, the results correspond to the complete learning regime, where the width shrinks to zero. The substantive and non-trivial setting is therefore formulation (iv), where $N \to \infty$ and $\delta \to 0$ jointly. We find that if $\lim_{\delta\to 0} \frac{N_\delta}{\log(1/\delta)} = k \in [0, \infty]$, the three different regimes, no learning ($k = 0$), sufficient learning ($k \in (0, \infty)$), and complete learning ($k = \infty$), arise naturally, and our results precisely characterize the minimum achievable limiting CI width in each case.

At last, we discuss the possible connections of our lower bound result with the classical Cramér–Rao lower bound in Appendix K

5 ASYMPTOTIC OPTIMALITY OF KL DIVERGENCE BASED CONSTRUCTION OF CI: DESCRIPTION OF METHOD π_1

In this section, we describe the method/policy π_1. It is well known that for $\nu \in \mathbf{S}$, the sample average is the MLE of the mean, $\hat{\mu}_n = \frac{\sum_{t=1}^n X_t}{n}$. For $\nu \in \mathbf{S}$, we utilize the concentration inequality based on KL divergences:

$$\mathbb{P}_\nu(n\, d(\hat{\mu}_n, \mu) \geq \beta(\delta)) \leq \delta. \tag{4}$$

Here $\beta(\delta) = \log(2/\delta)$ is a well-chosen function so the above holds. This can be derived from Lemma 4 in Ménard and Garivier (2017) or Theorem 4 in Busa-Fekete et al. (2019) and the proof is based on applying Markov's inequality and the structure of $d(\cdot, \cdot)$ function. To see, how we construct CI from the above concentration inequality, we formally define $\mu_L^{\pi_1}(n, \delta)$ and $\mu_R^{\pi_1}(n, \delta)$ as follows:

$$\mu_R^{\pi_1}(n, \delta) \triangleq \max\{q > \hat{\mu}_n : nd(\hat{\mu}_n, q) \leq \beta(\delta)\} \text{ and } \tag{5}$$

$$\mu_L^{\pi_1}(n, \delta) \triangleq \min\{q < \hat{\mu}_n : nd(\hat{\mu}_n, q) \leq \beta(\delta)\}.$$

Now we state the formal result related to the limiting width of the CI construction under policy π_1.

Theorem 2. *The policy π_1 has following properties:*

a) $\pi_1 \in \Pi_{\text{CI}}^s$.

b) If $\lim_{\delta\to 0} \frac{N_\delta}{\log(1/\delta)} \to k$ for $k \in (0, \infty)$, then we have, $\mu_R^{\pi_1}(\mu) = \mu_R^*(\mu, k)$ and $\mu_L^{\pi_1}(\mu) = \mu_L^*(\mu, k)$, where, $\mu_L^*(\mu, k) < \mu$ and $\mu_R^*(\mu, k) > \mu$ uniquely solve (3).

c) If $\lim_{\delta\to 0} \frac{N_\delta}{\log(1/\delta)} \to \infty$, then we have, $\mu_R^{\pi_1}(\mu) - \mu_L^{\pi_1}(\mu) = 0$.

The above theorem shows that π_1 achieves the minimum limiting width in the sufficient learning regime and zero limiting width in the complete learning regime, making it an asymptotically optimal policy for CI construction of the mean. Next, we study the setting where samples are costly.

6 RANDOM SAMPLING COST

Motivated by the applications discussed in Section 1, this section focuses on the problem of constructing a CI for the mean when there is a random cost associated with obtaining samples. In this setting, instead of a fixed sample size N, we assume a cost budget of C units. As in the previous setting, we will scale C with δ, and henceforth denote the cost budget as C_δ. Each sample incurs a random cost c_i, where c_i for $i = 1, 2, 3, \ldots$ are i.i.d. copies from an unknown cost distribution $\mathcal{C}$ with positive support and mean $\bar{c} > 0$. We assume that the distributions ν and $\mathcal{C}$ are independent. Consequently, X_i and c_i for $i = 1, 2, 3, \ldots$ are also independent.

Our goal is to establish the lower bound on the limiting width of any CI construction policy that ensures the mean lies within the interval with probability $(1 - \delta)$ given the budget C_δ. To do this, we define a random time, interpreted as the maximum number of samples that can be collected within the given budget C_δ: $\tau_\delta = \sup\{n \in \mathbb{Z}^+ : \sum_{i=1}^n c_i \leq C_\delta\}$.

Let $[\widehat{\mu}_L^\pi(\tau_\delta, \delta), \widehat{\mu}_R^\pi(\tau_\delta, \delta)]$ denote the estimated CI after observing $X_1, X_2, \ldots, X_{\tau_\delta}$ under a policy π for any given $\delta \in (0, 1)$. Let $\hat{\Pi}_{\text{CI}}$ denote the collection of CI construction policies such that the following holds for any $\pi \in \hat{\Pi}_{\text{CI}}$ and for any given $\delta \in (0, 1)$, $\mathbb{P}_\nu(\mu \in [\widehat{\mu}_L^\pi(\tau_\delta, \delta), \widehat{\mu}_R^\pi(\tau_\delta, \delta)]) \geq 1 - \delta$.

Definition 2. *Let $C_\delta \to \infty$ as $\delta \to 0$. For a given distribution ν with mean μ, for any $\pi \in \hat{\Pi}_{\text{CI}}$, π is called a stable policy if following holds,*

$\lim_{\delta\to 0} \widehat{\mu}_L^\pi(\tau_\delta, \delta) \xrightarrow{p} \mu_L^\pi(\nu)$, $\lim_{\delta\to 0} \widehat{\mu}_R^\pi(\tau_\delta, \delta) \xrightarrow{p} \mu_R^\pi(\nu)$, *where $\mu_L^\pi(\nu) \leq \mu$ and $\mu_R^\pi(\nu) \geq \mu$ are constants (∞ and $-\infty$ included).*

We denote the collection of policies in the set $\hat{\Pi}_{\text{CI}}$ which are stable as $\hat{\Pi}_{\text{CI}}^s$. It is worth noting that $\mu_R^\pi(\nu) - \mu_L^\pi(\nu)$ denotes the limiting width of CI for $\pi \in \Pi_{\text{CI}}^s$ in the asymptotic regime where the sample size, i.e, $C_\delta \to \infty$ as $\delta \to 0$. We again assume that ν belongs to the canonical single-parameter exponential family $\mathbf{S}$. Hence, for simplicity, we denote $\mu_L^\pi(\nu)$ and $\mu_R^\pi(\nu)$ as $\mu_L^\pi(\mu)$ and $\mu_R^\pi(\mu)$, respectively. In Section 7, we generalize our results for non-parametric distributions. Now we state our key result on the lower bound on the limiting width of any $\pi \in \hat{\Pi}_{\text{CI}}^s$.

Theorem 3. *For a given $\nu \in \mathbf{S}$ with mean μ, a cost*

distribution $\mathcal{C}$ with mean $0 < \overline{c} < \infty$, and any $\pi \in \hat{\Pi}^s_{\rm CI}$, the following holds:

a) ***No learning regime :*** *If $\lim_{\delta\to 0} \frac{C_\delta}{\log(1/\delta)} \to 0$, then $[\mu^\pi_R(\mu) - \mu^\pi_L(\mu)] = \overline{\mu} - \underline{\mu}$. Further, $\lim_{\delta\to 0} \widehat{\mu}^\pi_L(\tau_\delta, \delta) \xrightarrow{p} \underline{\mu}$ and $\lim_{\delta\to 0} \widehat{\mu}^\pi_R(\tau_\delta, \delta) \xrightarrow{p} \overline{\mu}$.*

b) ***Sufficient learning regime :*** *If $\lim_{\delta\to 0} \frac{C_\delta}{\log(1/\delta)} \to k$ for $k \in (0, \infty)$, then*

$$[\mu^\pi_R(\mu) - \mu^\pi_L(\mu)] \geq \mu^*_R(\mu, k, \overline{c}) - \mu^*_L(\mu, k, \overline{c}), \quad (6)$$

where, $\mu^*_L(\mu, k, \overline{c}) < \mu$ and $\mu^*_R(\mu, k, \overline{c}) > \mu$ uniquely solve the following system of equations,

$$d(\mu, \mu^*_R(\mu, k, \overline{c})) = d(\mu, \mu^*_L(\mu, k, \overline{c})) = \frac{\overline{c}}{k}. \quad (7)$$

The presence of an additional average sample collection cost, $\overline{c}$, in (7) differentiates the above result from Theorem 1. Furthermore, the above result indicates that the limiting width of the confidence interval is invariant to the distribution of the cost and only depends on its mean. Apart from the ideas in the proof of Theorem 1, three additional key ideas are required to prove the above theorem. The first is the use of renewal theory to study the scaling of τ_δ with C_δ. The second is the fact that (1) holds for a stopping time, and in this setting, $\tau_\delta + 1$ is also a stopping time. The last idea is to extend (1) to the joint distribution of ν and $\mathcal{C}$.

Remark 2. *Analogous to Remark 1, here when $\lim_{\delta\to 0} \frac{C_\delta}{\log(1/\delta)} \to \infty$, we have a trivial lower bound on the limiting width, i.e., $[\mu^\pi_R(\mu) - \mu^\pi_L(\mu)] \geq 0$ for any $\pi \in \hat{\Pi}^s_{\rm CI}$. Further, our modified policy $\hat{\pi}_1$ (see below) has zero limiting width in this regime. Again, we denote this regime as the complete learning regime.*

In this extended setting, we propose a modified method, denoted as $\hat{\pi}_1$, which is identical to π_1, except with a modified $\beta(n, \delta)$ replacing $\beta(\delta)$ in (5). This modification is necessary for the following reason: the number of samples is random in this setting, and the concentration inequality used for a fixed number of samples, as given in (4), is no longer valid. To address this, we employ an anytime-valid concentration inequality. Specifically, we use $\beta(n, \delta)$ from (14) in Kaufmann and Koolen (2021), defined as: $\beta(n, \delta) = 3\log(1 + \log(n)) + \mathcal{T}(\log(1/\delta))$.

Here, the function $\mathcal{T}(x) : \mathbb{R}^+ \to \mathbb{R}^+$ is given by:$\mathcal{T}(x) = 2\tilde{\psi}_{3/2}\left(\frac{x+\log(2\zeta(2))}{2}\right)$, where $\zeta(2) = \sum_{n=1}^\infty n^{-2}$ and,

$$\tilde{\psi}_y(x) = \begin{cases} e^{1/\psi^{-1}(x)}\psi^{-1}(x) & \text{if } x \geq \psi^{-1}(1/\ln y), \\ y(x - \ln\ln y) & \text{otherwise,} \end{cases}$$

for any $y \in [1, e]$ and $x \geq 0$. The function $\psi(u) = u - \ln u$ has an inverse $u = \psi^{-1}(z)$ for $z \geq 1$. As shown by Kaufmann and Koolen (2021), it holds that $\lim_{x\to\infty} \frac{\mathcal{T}(x)}{x} = 1$. Now we state the formal result that the limiting width of the CI under policy $\hat{\pi}_1$ matches the lower bound.

Theorem 4. *The policy $\hat{\pi}_1$ has following properties:*

a) $\hat{\pi}_1 \in \hat{\Pi}^s_{\rm CI}$.

b) If $\lim_{\delta\to 0} \frac{C_\delta}{\log(1/\delta)} \to k$ for $k \in (0, \infty)$, then we have, $\mu^{\hat{\pi}_1}_R(\mu) = \mu^*_R(\mu, k, \overline{c})$, and $\mu^{\hat{\pi}_1}_L(\mu) = \mu^*_L(\mu, k, \overline{c})$ where, $\mu^*_L(\mu, k, \overline{c}) < \mu$ and $\mu^*_R(\mu, k, \overline{c}) > \mu$ uniquely solve (7).

c) If $\lim_{\delta\to 0} \frac{C_\delta}{\log(1/\delta)} \to 0$, then we have, $\mu^{\hat{\pi}_1}_R(\mu) - \mu^{\hat{\pi}_1}_L(\mu) = 0$.

7 GENERALIZATION TO NON-PARAMETRIC DISTRIBUTIONS

In this section, we generalize our results for the case when the underlying distribution belongs to a non-parametric family. In particular, we consider two such settings. The first one is the family of probability distributions with bounded support in $[0, 1]$, we denote it as $\mathbf{B}$. The second one is the family of probability distributions with $(1+\epsilon)^{\rm th}$ moment bounded and the bound is known and given by a constant Γ. We denote this family as $\mathbf{H}$. It is worth noting that $KL_{\rm inf}(\nu, \mathbf{B}, x)$ is a well-studied in Honda and Takemura (2010) and Jourdan et al. (2022) for $\nu \in \mathbf{B}$. Further, Agrawal (2022) studies $KL_{\rm inf}(\nu, \mathbf{H}, x)$ for $\nu \in \mathbf{H}$. A primer on $KL_{\rm inf}(\nu, \mathbf{B}, x)$ and $KL_{\rm inf}(\nu, \mathbf{H}, x)$ and how they are computed is provided in the Appendix F,G. The analysis in this section is similar to the setting where $\nu \in \mathbf{S}$, with $d(\mu, x)$ now replaced by $KL_{\rm inf}(\nu, \mathbf{P}, x)$ (defined below).

Given a probability distribution family $\mathbf{P}$, an outcome distribution $\nu \in \mathbf{P}$ and $x \in \mathbf{R}$, let,

$$KL_{\rm inf}(\nu, \mathbf{P}, x) = \begin{cases} \inf\limits_{\substack{\kappa\in\mathbf{P} \\ m(\kappa)\geq x}} KL(\nu, \kappa), & \text{if } x \geq m(\nu) \\ \inf\limits_{\substack{\kappa\in\mathbf{P} \\ m(\kappa)\leq x}} KL(\nu, \kappa), & \text{if } x < m(\nu). \end{cases} \quad (8)$$

$KL_{\rm inf}(\nu, \mathbf{P}, x)$ is the minimum amongst the KL divergences between a given distribution ν and all distributions in the same family $\mathbf{P}$ which have higher mean than x if $x \geq m(\nu)$. It is similarly defined for $x < m(\nu)$.

7.1 Results on lower bound on limiting width of CI

For the setting, when $\nu \in \{\mathbf{B}, \mathbf{H}\}$, Theorem 3 extends as follows.

Theorem 5. *For a given $\nu \in \{\mathbf{B}, \mathbf{H}\}$ with mean $m(\nu) = \mu$, a cost distribution $\mathcal{C}$ with mean $0 < \overline{c} < \infty$, and any $\pi \in \hat{\Pi}^s_{\text{CI}}$, the following holds:*

a) ***No learning regime :*** *If $\lim_{\delta \to 0} \frac{C_\delta}{\log(1/\delta)} \to 0$, then $[\mu^\pi_R(\nu) - \mu^\pi_L(\nu)] = \overline{\mu} - \underline{\mu}$.*

Further, $\lim_{\delta \to 0} \widehat{\mu}^\pi_L(\tau_\delta, \delta) \xrightarrow{p} \underline{\mu}$ and $\lim_{\delta \to 0} \widehat{\mu}^\pi_R(\tau_\delta, \delta) \xrightarrow{p} \overline{\mu}$.

b) ***Sufficient learning regime :*** *If $\lim_{\delta \to 0} \frac{C_\delta}{\log(1/\delta)} \to k$ for $k \in (0, \infty)$, then*

$$[\mu^\pi_R(\nu) - \mu^\pi_L(\nu)] \geq \mu^*_R(\nu, k, \overline{c}) - \mu^*_L(\nu, k, \overline{c}), \quad (9)$$

$\mu^*_L(\nu, k, \overline{c}) < \mu$ and $\mu^*_R(\nu, k, \overline{c}) > \mu$ uniquely solve the following system of equations,

$$KL_{\inf}(\nu, \mathbf{P}, \mu^*_R(\nu, k, \overline{c})) = KL_{\inf}(\nu, \mathbf{P}, \mu^*_L(\nu, k, \overline{c})) = \frac{\overline{c}}{k}, \quad (10)$$

where, $\mathbf{P} = \mathbf{B}$ if $\nu \in \mathbf{B}$ and $\mathbf{P} = \mathbf{H}$ if $\nu \in \mathbf{H}$.

In this setting, when $\nu \in \{\mathbf{B}, \mathbf{H}\}$, we denote the limiting CI of a policy $\pi \in \hat{\Pi}^s_{\text{CI}}$ as $[\mu^\pi_L(\nu), \mu^\pi_R(\nu)]$ as $\delta \to 0$ and $C_\delta \to \infty$.

7.2 Asymptotic optimality of KLinf-based construction of CI

We use the following concentration inequality for the construction of the CI (see Agrawal (2022), Orabona and Jun (2023), and Jourdan et al. (2022)),

$$\mathbb{P}_\nu(\exists\, n \in \mathbb{N} : n\, KL_{\inf}(\hat{\nu}_n, \mathbf{B}, \mu) \geq \beta(n, \delta)) \leq \delta, \quad (11)$$

$\hat{\nu}_n$ denotes the empirical distribution after n samples and $\beta(n, \delta) = 1 + \log\left(\frac{2(1+n)}{\delta}\right)$. Similar to the π_1 and $\hat{\pi}_1$, we now utilize the above KLinf-based concentration inequality to get a CI construction method denoted as π^{b}_1. Let, $\mu^{\pi^{\text{b}}_1}_R(n, \delta) \triangleq \max\{q > m(\hat{\nu}_n) : n\, KL_{\inf}(\hat{\nu}_n, \mathbf{B}, q) \leq \beta(n, \delta)\}$, and $\mu^{\pi^{\text{b}}_1}_L(n, \delta) \triangleq \min\{q < m(\hat{\nu}_n) : \quad n\, KL_{\inf}(\hat{\nu}_n, \mathbf{B}, q) \leq \beta(n, \delta)\}$. It follows that the reported CI is $[\mu^{\pi^{\text{b}}_1}_L(n, \delta), \mu^{\pi^{\text{b}}_1}_R(n, \delta)]$. An equivalent version of Theorem 2 and Theorem 4 hold for our policy when $\nu \in \mathbf{B}$. The statements and proofs are given in the Appendix E. A similar CI construction method for $\nu \in \mathbf{H}$ holds using concentration bound in Agrawal (2022). Further, an equivalent version of Theorem 2 and Theorem 4 hold for the policy π^{b}_1 when $\nu \in \mathbf{H}$(with a minor technical assumption). See Appendix E.1 for the description of the CI construction method, Theorem statements and proofs. A numerical study for the heavy tailed case, i.e., $\nu \in \mathbf{H}$ is provided in Appendix H as well.

Remark 3. *We provide a non-asymptotic analysis of the width of the CI for our policies under the cases where $\nu \in \mathbf{S}$ and $\mathbf{B}$ in Appendix J.*

8 NUMERICAL EXPERIMENTS

Our numerical study has two objectives. First, to demonstrate numerically that our asymptotic lower bound is sharper and tighter than that of Shekhar and Ramdas (2023) in our asymptotic regime. Second, to demonstrate the performance of the CI construction method π_1 and compare it with existing methods.

Observe that Proposition 4.3 of Shekhar and Ramdas (2023) presents a non-asymptotic lower bound on the width of any CI for a given N and δ. We first compare our asymptotic lower bound with the one presented by Shekhar and Ramdas (2023) in the asymptotic regime where $\delta \to 0$ and $N_\delta \to \infty$. In this regime, we plot both lower bounds versus the scaling constant k, where $\lim_{\delta \to 0} \frac{N_\delta}{\log(1/\delta)} = k$, for the case when $\nu = N(0, 1)$ with known variance. The results demonstrate that our asymptotic lower bound is twice as high as that of Shekhar and Ramdas (2023). This is shown as a function of the scaling constant in Figure 1.

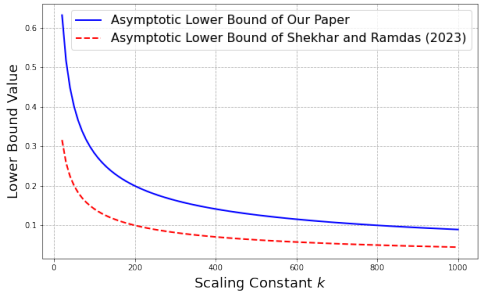

Figure 1: Comparison of our asymptotic lower bound given in Theorem 1 as a function of k, with the lower bound presented in Proposition 4.3 of Shekhar and Ramdas (2023) when $\lim_{\delta \to 0} \frac{N_\delta}{\log(1/\delta)} = k$. We assume that $\nu = N(0, 1)$ with known variance.

The explanation for the result observed in Figure 1 is as follows. For the Gaussian case with known variance σ, we have $\mu^*_R(\mu) = \mu + \sigma\sqrt{2/k}$ and $\mu^*_L(\mu) = \mu - \sigma\sqrt{2/k}$. Thus, our lower bound is $2\sigma\sqrt{2/k}$, while the bound from Shekhar and Ramdas (2023) is $\sigma\sqrt{2/k}$.

Hence, the factor of 2 applies in the Gaussian case. For other distributions, this ratio may vary; however, our bound remains asymptotically tighter. Moreover, recall that we have proven that the policy π_1 achieves the limiting width in the sufficient learning regime, which shows that our lower bound cannot be improved in the regime when $\delta \to 0$ and $N_\delta \to \infty$.

We now compare the performance of the method π_1 with three existing methods: the Hoeffding-based CI, the Empirical Bernstein (EB) CI (see Maurer and

Pontil (2009)), and the betting-based hedged CI (see Waudby-Smith and Ramdas (2024)). Experiments were conducted on Bernoulli distributions with means of 0.6 and 0.9 and δ is set to be 1%. For the betting method, we vary the discretization parameter $m \in \{1000, 3000, 5000\}$ (needed to discretize the [0,1] space). For each configuration, we generated 1000 i.i.d. datasets of size $N \in \{2000, 3000\}$, computed CI widths, and report the average width (max 95% CI width: 0.0001). Our method consistently yields narrower CIs than both Hoeffding and EB across settings. While betting-based hedged CIs slightly outperform ours at $N = 2000$, performance equalizes by $N = 3000$. Increasing discretization m offers diminishing returns beyond $m = 5000$. Further, our method remains computationally more efficient as compared to the betting-based hedged CIs due to its ease of computation.

Table 1: Average CI width for Bernoulli distributions with means 0.6 and 0.9 for $\delta = 0.01$

N	π_1	Betting (m=1000)	Betting (m=3000)	Betting (m=5000)	Hoeffding	EB
			Mean = 0.6			
2000	0.0712	0.0603	0.0596	0.0595	0.0728	0.0898
3000	0.0582	0.0592	0.0585	0.0584	0.0594	0.0712
			Mean = 0.9			
2000	0.0436	0.0378	0.0371	0.0369	0.0728	0.0606
3000	0.0356	0.0370	0.0363	0.0361	0.0594	0.0473

9 CONCLUSION AND FUTURE DIRECTIONS

In this paper, we explored the construction of confidence intervals (CIs) for the mean of a probability distribution, focusing on their minimum achievable limiting width in three distinct learning regimes. By characterizing these regimes, we provided a comprehensive understanding of the inherent limitations and potential of CI construction methods in both parametric and non-parametric settings. Finally, we demonstrated the asymptotic optimality of KL-based CI construction methods and extended our results to practically relevant scenarios where sample collection incurs random costs. While non-asymptotic results can capture more "granularity" and reveal finer structural properties, nonetheless, our asymptotic analysis offers more concise and elegant insights that have independent value. In particular, our stable policies assumption is asymptotic, is true quite generally and greatly simplifies the analysis. Future research directions could include extending the framework to other statistical estimation problems, such as variance and quantile estimation. At last, extending our results to higher-dimensional parameters, such as the mean of a multivariate distribution, likewise represents a compelling and nontrivial avenue for future research.

References

Agrawal, S. (2022). Bandits with heavy tails algorithms analysis and optimality.

Amin, K., Kearns, M., Key, P., and Schwaighofer, A. (2012). Budget optimization for sponsored search: Censored learning in mdps. *arXiv preprint arXiv:1210.4847*.

Audibert, J.-Y., Munos, R., and Szepesvári, C. (2009). Exploration–exploitation tradeoff using variance estimates in multi-armed bandits. *Theoretical Computer Science*, 410(19):1876–1902.

Bennett, G. (1962). Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association*, 57(297):33–45.

Bickel, P. J. and Doksum, K. A. (2015). *Mathematical statistics: basic ideas and selected topics, volumes I–II package*. Chapman and Hall/CRC.

Boucheron, S. and Gassiat, E. (2009). A bernstein-von mises theorem for discrete probability distributions.

Busa-Fekete, R., Fotakis, D., Szörényi, B., and Zampetakis, M. (2019). Optimal learning of mallows block model. In *Proceedings of the Conference on Learning Theory*, pages 529–532. PMLR.

Cappé, O., Garivier, A., Maillard, O.-A., Munos, R., and Stoltz, G. (2013). Kullback-leibler upper confidence bounds for optimal sequential allocation. *The Annals of Statistics*, pages 1516–1541.

Casella, G. and Berger, R. L. (2024). *Statistical Inference*. CRC Press.

Catoni, O. (2012). Challenging the empirical mean and empirical variance: a deviation study. In *Annales de l'IHP Probabilités et statistiques*, volume 48, pages 1148–1185.

Chen, P., Jin, X., Li, X., and Xu, L. (2021). A generalized catoni's m-estimator under finite α-th moment assumption with $\alpha \in (1, 2)$. *Electronic Journal of Statistics*, 15(2):5523–5544.

Chick, S. E. and Inoue, K. (2001). New two-stage and sequential procedures for selecting the best simulated system. *Operations Research*, 49(5):732–743.

Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory*. Wiley-Interscience, New York.

Efron, B. and Tibshirani, R. J. (1994). *An introduction to the bootstrap*. Chapman and Hall/CRC.

Garivier, A. and Cappé, O. (2011). The kl-ucb algorithm for bounded stochastic bandits and beyond. In *Proceedings of the 24th annual conference on learning theory*, pages 359–376. JMLR Workshop and Conference Proceedings.

Glynn, P. W. and Juneja, S. (2013). Asymptotic simulation efficiency based on large deviations. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 23(3):1–16.

Gupta, S., Lee, J., Price, E., and Valiant, P. (2023). Minimax-optimal location estimation. *Advances in Neural Information Processing Systems*, 36:900–915.

Hoeffding, W. (1994). Probability inequalities for sums of bounded random variables. *The collected works of Wassily Hoeffding*, pages 409–426.

Honda, J. and Takemura, A. (2010). An asymptotically optimal bandit algorithm for bounded support models. In *COLT*, pages 67–79. Citeseer.

Jourdan, M., Degenne, R., Baudry, D., de Heide, R., and Kaufmann, E. (2022). Top two algorithms revisited. *Advances in Neural Information Processing Systems*, 35:26791–26803.

Kaufmann, E. (2020). *Contributions to the Optimal Solution of Several Bandit Problems*. PhD thesis, Université de Lille.

Kaufmann, E. and Koolen, W. M. (2021). Mixture martingales revisited with applications to sequential tests and confidence intervals. *Journal of Machine Learning Research*, 22(246):1–44.

Maurer, A. and Pontil, M. (2009). Empirical bernstein bounds and sample variance penalization. *arXiv preprint arXiv:0907.3740*.

Ménard, P. and Garivier, A. (2017). A minimax and asymptotically optimal algorithm for stochastic bandits. In *Proceedings of the International Conference on Algorithmic Learning Theory*, pages 223–237. PMLR.

Orabona, F. and Jun, K.-S. (2023). Tight concentrations and confidence sequences from the regret of universal portfolio. *IEEE Transactions on Information Theory*.

Shekhar, S. and Ramdas, A. (2023). On the near-optimality of betting confidence sets for bounded means. *arXiv preprint arXiv:2310.01547*.

Tran-Thanh, L., Stavrogiannis, L., Naroditskiy, V., Robu, V., Jennings, N. R., and Key, P. (2014). Efficient regret bounds for online bid optimisation in budget-limited sponsored search auctions. In *uai2014, 30th Conf. on Uncertainty in AI*.

Waudby-Smith, I. and Ramdas, A. (2024). Estimating means of bounded random variables by betting. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(1):1–27.

Xia, Y., Li, H., Qin, T., Yu, N., and Liu, T.-Y. (2015). Thompson sampling for budgeted multi-armed bandits. *arXiv preprint arXiv:1505.00146*.

z

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes]
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Yes]

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes]
   (b) Complete proofs of all theoretical results. [Yes]
   (c) Clear explanations of any assumptions. [Yes]

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes]
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Not Applicable]
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes]
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Not Applicable]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [Not Applicable]
   (b) The license information of the assets, if applicable. [Not Applicable]
   (c) New assets either in the supplemental material or as a URL, if applicable. [Yes]
   (d) Information about consent from data providers/curators. [Not Applicable]
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. [Not Applicable]
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

Supplementary Material

Discussion on stable policies: Fix a distribution ν with mean μ. Recall a policy π is called stable if

$$\lim_{\delta\to 0} \widehat{\mu}_L^{\pi}(N_\delta, \delta) \xrightarrow{p} \mu_L^{\pi}(\nu), \quad \lim_{\delta\to 0} \widehat{\mu}_R^{\pi}(N_\delta, \delta) \xrightarrow{p} \mu_R^{\pi}(\nu),$$

where $\mu_L^{\pi}(\nu) \le \mu$ and $\mu_R^{\pi}(\nu) \ge \mu$ are constants ($-\infty$ and ∞ included). For instance, a policy using a symmetric confidence interval based on the central limit theorem inherently meets this assumption. Furthermore, for bounded outcome distributions, classical confidence interval approaches based on Hoeffding's and Bernstein's inequalities, as well as Maurer and Pontil's Empirical Bernstein (MP-EB) CI Maurer and Pontil (2009), also satisfy the stability assumption. To verify this, we analyze the asymptotic behavior of the CI boundaries as $N_\delta \to \infty$ and $\delta \to 0$ for three CI constructions. The key observation is that the CI boundaries converge almost surely to constants depending on the relationship between N_δ and $\log(1/\delta)$. Specifically, we consider three cases based on $\lim_{\delta\to 0} \frac{N_\delta}{\log(1/\delta)} = k$, where $k \in \{0, (0,\infty), \infty\}$.

Hoeffding CI. $C_N^{(H)} = \left[\widehat{\mu}_N \pm \frac{w_N^{(H)}}{2}\right] := \left[\widehat{\mu}_N - \frac{w_N^{(H)}}{2}, \widehat{\mu}_N + \frac{w_N^{(H)}}{2}\right]$, where $\widehat{\mu}_N = \frac{\sum_{i=1}^N X_i}{N}$ and $w_N^{(H)} = 2\sqrt{\frac{\log(2/\delta)}{2N}}$. For Hoeffding's CI, by the strong law of large numbers, $\widehat{\mu}_N \xrightarrow{a.s.} \mu$ as $N \to \infty$. The width term satisfies:

- If $k = \infty$: $w_N^{(H)} = 2\sqrt{\frac{\log(2/\delta)}{2N}} \to 0$, thus $\mu_L^{\pi}(\nu) = \mu_R^{\pi}(\nu) = \mu$.
- If $k \in (0,\infty)$: $w_N^{(H)} \to \sqrt{\frac{2}{k}}$ (constant), thus $\mu_L^{\pi}(\nu) = \mu - \frac{1}{2}\sqrt{\frac{2}{k}}$ and $\mu_R^{\pi}(\nu) = \mu + \frac{1}{2}\sqrt{\frac{2}{k}}$.
- If $k = 0$: $w_N^{(H)} \to \infty$, thus $\mu_L^{\pi}(\nu) = -\infty$ and $\mu_R^{\pi}(\nu) = +\infty$.

Bernstein CI. $C_N^{(B)} = \left[\widehat{\mu}_N \pm \frac{w_N^{(B)}}{2}\right]$, where $w_N^{(B)} = 2\sigma\sqrt{\frac{2\log(2/\delta)}{N}} + \frac{4\log(2/\delta)}{3N}$. For Bernstein's CI (see Shekhar and Ramdas (2023) for above formulation), the analysis is similar.

- If $k = \infty$: $w_N^{(B)} \to 0$, thus $\mu_L^{\pi}(\nu) = \mu_R^{\pi}(\nu) = \mu$.
- If $k \in (0,\infty)$: $w_N^{(B)} \to 2\sigma\sqrt{\frac{2}{k}} + \frac{4}{3k}$ (constant), thus $\mu_L^{\pi}(\nu) = \mu - \sigma\sqrt{\frac{2}{k}} - \frac{2}{3k}$ and $\mu_R^{\pi}(\nu) = \mu + \sigma\sqrt{\frac{2}{k}} + \frac{2}{3k}$.
- If $k = 0$: $w_N^{(B)} \to \infty$, thus $\mu_L^{\pi}(\nu) = -\infty$ and $\mu_R^{\pi}(\nu) = +\infty$.

Maurer & Pontil's Empirical Bernstein (MP-EB) CI. $C_N^{(\text{MP-EB})} = \left[\widehat{\mu}_N \pm \frac{w_N^{(\text{MP-EB})}}{2}\right]$, where $w_N^{(\text{MP-EB})} = 2\widehat{\sigma}_N\sqrt{\frac{2\log(4/\delta)}{N}} + \frac{14\log(4/\delta)}{3(N-1)}$, and $\widehat{\sigma}^2{}_N = \frac{\sum_{i=1}^N (X_i - \widehat{\mu}_N)^2}{N-1}$. For the MP-EB CI, note that $\widehat{\sigma}_{N_\delta} \xrightarrow{a.s.} \sigma$ as $N_\delta \to \infty$. Hence,

- If $k = \infty$: $w_N^{(\text{MP-EB})} \to 0$, thus $\mu_L^{\pi}(\nu) = \mu_R^{\pi}(\nu) = \mu$.
- If $k \in (0,\infty)$: $w_N^{(\text{MP-EB})} \to 2\sigma\sqrt{\frac{2}{k}} + \frac{14}{3k}$ (constant), thus $\mu_L^{\pi}(\nu) = \mu - \sigma\sqrt{\frac{2}{k}} - \frac{7}{3k}$ and $\mu_R^{\pi}(\nu) = \mu + \sigma\sqrt{\frac{2}{k}} + \frac{7}{3k}$.
- If $k = 0$: $w_N^{(\text{MP-EB})} \to \infty$, thus $\mu_L^{\pi}(\nu) = -\infty$ and $\mu_R^{\pi}(\nu) = +\infty$.

In all these cases, the CI boundaries converge almost surely to constants as $N_\delta \to \infty$ and $\delta \to 0$, confirming that these CI-based policies satisfy the stability assumption.

A Proofs of results in Section 4.

Proof of Theorem 1. Consider any $\pi \in \Pi^s_{\text{CI}}$. As the policy π is stable, it follows that for a given distribution $\nu \in \mathbf{S}$ with mean μ, we have,

$$\lim_{\delta\to 0} \hat{\mu}^\pi_L(N_\delta, \delta) \xrightarrow{p} \mu^\pi_L(\mu) \text{ and } \lim_{\delta\to 0} \hat{\mu}^\pi_R(N_\delta, \delta) \xrightarrow{p} \mu^\pi_R(\mu). \tag{12}$$

We now define the set of alternate environments $K(\mu^\pi_L(\mu), \mu^\pi_R(\mu)) = K_1(\mu^\pi_L(\mu)) \cup K_2(\mu^\pi_R(\mu))$, $K_1(\mu^\pi_L(\mu)) = \{\tilde{\nu} : \tilde{\nu} \in \mathbf{S}, \tilde{\mu} < \mu^\pi_L(\mu)\}$ and $K_2(\mu^\pi_R(\mu)) = \{\tilde{\nu} : \tilde{\nu} \in \mathbf{S}, \tilde{\mu} > \mu^\pi_R(\mu)\}$.

Using the data processing inequality for a given $\delta \in (0,1)$ and any alternate environment $\tilde{\nu}$ with mean $\tilde{\mu}$, we have

$$N_\delta \cdot d(\mu, \tilde{\mu}) \geq \sup_{\mathcal{E}_\delta \in \mathcal{F}_{N_\delta}} \phi(\mathbb{P}_\nu(\mathcal{E}_\delta), \mathbb{P}_{\tilde{\nu}}(\mathcal{E}_\delta)), \tag{13}$$

where $\phi(p_1, p_2) \triangleq p_1 \log \frac{p_1}{p_2} + (1-p_1)\log\left(\frac{1-p_1}{1-p_2}\right)$ for $p_1, p_2 \in (0,1)$. One can use Lemma 0.1 in Kaufmann (2020) to get (13) from the data processing inequality.

We first utilize (13) for $\tilde{\nu} \in K_1(\mu^\pi_L(\mu))$ and $\mathcal{E}_\delta = \{\tilde{\mu} \notin [\widehat{\mu}^\pi_L(N_\delta, \delta), \widehat{\mu}^\pi_R(N_\delta, \delta)]\}$. A similar approach would follow for $\tilde{\nu} \in K_2(\mu^\pi_R(\mu))$.

Since $\pi \in \Pi^s_{\text{CI}}$, using the definition of CI, we get, $\mathbb{P}_{\tilde{\nu}}(\mathcal{E}_\delta) \leq \delta$. Now observe that,

$$\mathbb{P}_\nu(\mathcal{E}_\delta) \geq \mathbb{P}_\nu\{\tilde{\mu} < \widehat{\mu}^\pi_L(N_\delta, \delta)\}.$$

Taking the limit of $\delta \to 0$ on both sides and using (12), we get,

$$\lim_{\delta\to 0} \mathbb{P}_\nu(\mathcal{E}_\delta) = 1.$$

Hence using the definition of $\phi(\cdot)$, we get,

$$\liminf_{\delta\to 0} \frac{\phi(\mathbb{P}_\nu(\mathcal{E}_\delta), \mathbb{P}_{\tilde{\nu}}(\mathcal{E}_\delta))}{\log(1/\delta)} \geq 1, \tag{14}$$

Hence using (13) and (14) it follows that for all $\tilde{\nu} \in K_1(\mu^\pi_L(\mu))$ with mean $\tilde{\mu}$, we have $\liminf_{\delta\to 0} \frac{N_\delta}{\log(1/\delta)} \cdot d(\mu, \tilde{\mu}) \geq 1$. Using a similar analysis for $\tilde{\nu} \in K_2(\mu^\pi_R(\mu))$ with mean $\tilde{\mu}$, we have a similar result, i.e., $\liminf_{\delta\to 0} \frac{N_\delta}{\log(1/\delta)} \cdot d(\mu, \tilde{\mu}) \geq 1$. Hence combining both, we get that for all $\tilde{\nu} \in K(\mu^\pi_L(\mu), \mu^\pi_R(\mu))$ with mean $\tilde{\mu}$ following holds,

$$\liminf_{\delta\to 0} \frac{N_\delta}{\log(1/\delta)} d(\mu, \tilde{\mu}) \geq 1.$$

It follows that,

$$\liminf_{\delta\to 0} \frac{N_\delta}{\log(1/\delta)} d(\mu, \mu^\pi_L(\mu) - \eta) \geq 1 \text{ and } \liminf_{\delta\to 0} \frac{N_\delta}{\log(1/\delta)} d(\mu, \mu^\pi_R(\mu) + \eta) \geq 1, \tag{15}$$

where $\eta > 0$ is any small positive number. Now we consider the case when $\liminf_{\delta\to 0} \frac{N_\delta}{\log(1/\delta)} = 0$. In this case, (15) can not hold for any $\mu^\pi_L(\mu) \in \mathbf{O}$ or $\mu^\pi_R(\mu) \in \mathbf{O}$. Hence, we can define $\mu^\pi_L(\mu) = \underline{\mu}$ and $\mu^\pi_R(\mu) = \overline{\mu}$ for any policy $\pi \in \Pi^s_{\text{CI}}$. This completes the proof of part (a).

Now we consider the case when $\liminf_{\delta\to 0} \frac{N_\delta}{\log(1/\delta)} = k$ for $k \in (0, \infty)$. In this case using (15) and taking $\eta \to 0$, we get,

$$d(\mu, \mu^\pi_L(\mu)) \geq \frac{1}{k} \text{ and } d(\mu, \mu^\pi_R(\mu)) \geq \frac{1}{k}.$$

Using the strict quasi-convexity of $d(\mu, \cdot)$, we get the desired result. This completes the proof.

□

B Proofs of results in Section 5.

Proof of Theorem 2.

We first show that our policy $\pi_1 \in \Pi^s_{\text{CI}}$. To prove this, we first show that,

$$\forall n \in \mathbb{N} : \mathbb{P}_\nu(\mu \in [\widehat{\mu}_L^{\pi_1}(n,\delta), \widehat{\mu}_R^{\pi_1}(n,\delta)]) \geq 1-\delta. \tag{16}$$

Observe that, for any $n \in \mathbb{N}$, it suffices to show that

$$\mathbb{P}_\nu(\mu \notin [\widehat{\mu}_L^{\pi_1}(n,\delta), \widehat{\mu}_R^{\pi_1}(n,\delta)]) \leq \mathbb{P}_\nu(nd(\hat{\mu}_n,\mu) \geq \beta(\delta)) = \mathbb{P}_\nu(nd(\hat{\mu}_n,\mu) \geq \log(2/\delta)) \leq \delta.$$

Using the result from Garivier and Cappé (2011), we know that for $\nu \in \mathbf{S}$,

$$d(\hat{\mu}_n,\mu) = I(\hat{\mu}_n) = \sup_{\lambda \in \mathbb{R}} \lambda\hat{\mu}_n - \log \mathbb{E}_\nu[e^{\lambda X}]. \tag{17}$$

Here $I(\cdot)$ is the good rate function used in large deviation. In the above, X is distributed according to ν.

Observe that we need to show (this can be derived from Lemma 4 in Ménard and Garivier (2017) or Theorem 4 in Busa-Fekete et al. (2019) as well),

$$\mathbb{P}_\nu(nd(\hat{\mu}_n,\mu) \geq \log(2/\delta)) \leq \delta.$$

It suffices to show that,

$$\mathbb{P}_\nu(nd(\hat{\mu}_n,\mu) \geq \log(2/\delta) \cap \hat{\mu}_n < \mu) \leq \frac{\delta}{2} \text{ and } \mathbb{P}_\nu(nd(\hat{\mu}_n,\mu) \geq \log(2/\delta) \cap \hat{\mu}_n \geq \mu) \leq \frac{\delta}{2}.$$

Now we show that

$$\mathbb{P}_\nu(nd(\hat{\mu}_n,\mu) \geq \log(2/\delta) \cap \hat{\mu}_n < \mu) \leq \frac{\delta}{2}. \tag{18}$$

Let $z \in (\underline{\mu}, \mu)$ such that, $I(z) = d(z,\mu) = \frac{\log(2/\delta)}{n}$. Using the property of good rate function we know that, for $z < \mu$, there exists a $\lambda(z) < 0$ such that the following holds

$$I(z) = d(z,\mu) = \lambda(z)z - \log \mathbb{E}_\nu[e^{\lambda(z)X}].$$

Now under the event $\{nd(\hat{\mu}_n,\mu) \geq \log(2/\delta) \cap \hat{\mu}_n < \mu\}$, we have,

$$\hat{\mu}_n \leq z.$$

Hence using the fact that $\lambda(z) < 0$, we get,

$$\{nd(\hat{\mu}_n,\mu) \geq \log(2/\delta) \cap \hat{\mu}_n < \mu\} \implies \left\{\lambda(z)\hat{\mu}_n - \log \mathbb{E}_\nu[e^{\lambda(z)X}] \geq d(z,\mu) = \frac{\log(2/\delta)}{n}\right\}.$$

Observe that, $\hat{\mu}_n = \frac{\sum_{i=1}^n X_i}{n}$, to prove (18), it suffices to show that,

$$\mathbb{P}_\nu\left(\lambda(z)\sum_{i=1}^n X_i - n\log \mathbb{E}_\nu[e^{\lambda(z)X}] \geq \log(2/\delta)\right) \leq \frac{\delta}{2}.$$

Observe that $\mathbb{E}_\nu[e^{\lambda(z)\sum_{i=1}^n X_i - n\log \mathbb{E}_\nu[e^{\lambda(z)X}]}] = 1$, hence using Markov's inequality, we get the desired result.

Similarly, one can show that,

$$\mathbb{P}_\nu(nd(\hat{\mu}_n,\mu) \geq \log(2/\delta) \cap \hat{\mu}_n \geq \mu) \leq \frac{\delta}{2}.$$

This completes the proof of (16).

Observe that from the definitions of $\widehat{\mu}_L^{\pi_1}(n,\delta)$, $\widehat{\mu}_R^{\pi_1}(n,\delta)$ and the strict quasi-convexity of $d(\mu,\cdot)$, we get,

$$d(\hat{\mu}_{N_\delta}, \widehat{\mu}_L^{\pi_1}(N_\delta,\delta)) = d(\hat{\mu}_{N_\delta}, \widehat{\mu}_R^{\pi_1}(N_\delta,\delta)) = \frac{\log(2/\delta)}{N_\delta}.$$

Recall $N_\delta \to \infty$, $\lim_{\delta\to 0} \hat{\mu}_{N_\delta} \to \mu$ almost surely from strong law of large numbers. Taking $\delta \to 0$ and using the joint-continuity of $d(\mu,x)$ in (μ,x), we get that $\pi_1 \in \Pi_{\text{CI}}^s$ and part (b), (c) of this theorem holds. This completes the proof.

□

C Results and proofs for the rate of convergence analysis in complete learning regime.

We first introduce a set of policies, denoted by $\Pi_{\text{CI}}^{sr} \subseteq \Pi_{\text{CI}}^s$ for which the analysis is valid. For $\pi \in \Pi_{\text{CI}}^{sr}$ and a given $\nu \in \mathbf{S}$ with mean μ, we assume that the following holds as $\lim_{\delta\to 0} \frac{N_\delta}{\log(1/\delta)} = \infty$:

$$\lim_{\delta\to 0} \frac{\mu - \widehat{\mu}_L^\pi(N_\delta,\delta)}{\sqrt{\frac{\log(1/\delta)}{N_\delta}}} \xrightarrow{p} \theta_L^\pi(\mu) \text{ and } \lim_{\delta\to 0} \frac{\widehat{\mu}_R^\pi(N_\delta,\delta) - \mu}{\sqrt{\frac{\log(1/\delta)}{N_\delta}}} \xrightarrow{p} \theta_R^\pi(\mu),$$

where, $\theta_L^\pi(\mu)$ and $\theta_R^\pi(\mu)$ are the non-negative constants.

It is worth noting that the above assumption can be interpreted as rate stability of policies under the complete learning regime. The example of stable policies discussed at the start of the supplementary material also satisfies this rate stability assumption.

We now state a lower bound on the rate of convergence of the width to zero.

Theorem 6. *Fix a $\nu \in \mathbf{S}$ with mean μ. If $\lim_{\delta\to 0} \frac{N_\delta}{\log(1/\delta)} = \infty$, then for any $\pi \in \Pi_{\text{CI}}^{sr}$, the following holds:*

$$\lim_{\delta\to 0} \frac{\widehat{\mu}_R^\pi(N_\delta,\delta) - \widehat{\mu}_L^\pi(N_\delta,\delta)}{\sqrt{\frac{\log(1/\delta)}{N_\delta}}} \xrightarrow{p} q(\mu) \geq \sqrt{8}\cdot\sigma(\mu),$$

where $\sigma(\mu)$ is the standard deviation of the ν. Further, $q(\mu)$ is some non-negative function.

Proof of Theorem 6. Recall that from (13), for a given $\delta \in (0,1)$ and any alternate environment $\tilde{\nu}$ with mean $\tilde{\mu}$, we have

$$N_\delta \cdot d(\mu,\tilde{\mu}) \geq \sup_{\mathcal{E}_\delta \in \mathcal{F}_{N_\delta}} \phi(\mathbb{P}_\nu(\mathcal{E}_\delta), \mathbb{P}_{\tilde{\nu}}(\mathcal{E}_\delta)).$$

Choose $\tilde{\mu} = \mu - (\theta_L^\pi(\mu) + \eta)\sqrt{\frac{\log(1/\delta)}{N_\delta}}$. Since we know that $d(\mu,x)$ is twice continuously differentiable in x and $\frac{\partial d(\mu,x)}{\partial x}|_{x=\mu} = 0$. Hence, using the Taylor series expansion of $d(\mu,\cdot)$ in the above equation, we get

$$N_\delta \frac{I(c_1)(\theta_L^\pi(\mu)+\eta)^2 \log(1/\delta)}{2N_\delta} \geq \sup_{\mathcal{E}_\delta \in \mathcal{F}_{N_\delta}} \phi(\mathbb{P}_\nu(\mathcal{E}_\delta), \mathbb{P}_{\tilde{\nu}}(\mathcal{E}_\delta)).$$

Here $I(c) = \frac{\partial^2 d(\mu,x)}{\partial x^2}|_{x=c}$ and $c_1 \in (\tilde{\mu},\mu)$ is a constant. Now we choose $\mathcal{E}_\delta = \{\tilde{\mu} \notin [\widehat{\mu}_L^\pi(N_\delta,\delta), \widehat{\mu}_R^\pi(N_\delta,\delta)]\}$ and take $\delta \to 0$, we get

$$\frac{I(\mu)(\theta_L^\pi(\mu)+\eta)^2}{2} \geq \lim_{\delta\to 0} \frac{\phi(\mathbb{P}_\nu(\mathcal{E}_\delta), \mathbb{P}_{\tilde{\nu}}(\mathcal{E}_\delta))}{\log(1/\delta)}. \tag{19}$$

Now consider the following:

$$\mathbb{P}_\nu(\mathcal{E}_\delta) = \mathbb{P}_\nu(\tilde{\mu} \notin [\widehat{\mu}_L^\pi(N_\delta,\delta), \widehat{\mu}_R^\pi(N_\delta,\delta)]).$$

It follows that,

$$\mathbb{P}_\nu(\mathcal{E}_\delta) \geq \mathbb{P}_\nu(\tilde{\mu} < \widehat{\mu}_L^\pi(N_\delta, \delta)).$$

Substituting the definition of $\tilde{\mu}$, we get,

$$\mathbb{P}_\nu(\mathcal{E}_\delta) \geq \mathbb{P}_\nu\left(\mu - \widehat{\mu}_L^\pi(N_\delta, \delta) < (\theta_L^\pi(\mu) + \eta)\sqrt{\frac{\log(1/\delta)}{N_\delta}}\right).$$

It can be re-written as,

$$\lim_{\delta\to 0} \mathbb{P}_\nu(\mathcal{E}_\delta) \geq \lim_{\delta\to 0} \mathbb{P}_\nu\left(\frac{\mu - \widehat{\mu}_L^\pi(N_\delta, \delta)}{\sqrt{\frac{\log(1/\delta)}{N_\delta}}} < (\theta_L^\pi(\mu) + \eta)\right).$$

Now using the fact that $\pi \in \Pi_{\text{CI}}^{sr}$, we know that, $\lim_{\delta\to 0} \frac{\mu - \widehat{\mu}_L^\pi(N_\delta,\delta)}{\sqrt{\frac{\log(1/\delta)}{N_\delta}}} \xrightarrow{p} \theta_L^\pi(\mu)$. Hence it follows that, for any $\eta > 0$, we have,

$$\lim_{\delta\to 0} \mathbb{P}_\nu(\mathcal{E}_\delta) \geq \lim_{\delta\to 0} \mathbb{P}_\nu\left(\frac{\mu - \widehat{\mu}_L^\pi(N_\delta, \delta)}{\sqrt{\frac{\log(1/\delta)}{N_\delta}}} < (\theta_L^\pi(\mu) + \eta)\right) = 1.$$

Observe that we get trivially, $\mathbb{P}_{\tilde{\nu}}(\mathcal{E}_\delta) \leq \delta$. Substituting the value of $\mathbb{P}_{\tilde{\nu}}(\mathcal{E}_\delta)$ and $\lim_{\delta\to 0} \mathbb{P}_\nu(\mathcal{E}_\delta)$ in (19), we get that,

$$\frac{I(\mu)(\theta_L^\pi(\mu) + \eta)^2}{2} \geq \lim_{\delta\to 0} \frac{\phi(\mathbb{P}_\nu(\mathcal{E}_\delta), \mathbb{P}_{\tilde{\nu}}(\mathcal{E}_\delta))}{\log(1/\delta)} \geq 1.$$

Taking $\eta \to 0$ and the fact that $I(\mu) = \frac{1}{\sigma^2(\mu)}$, we get,

$$\theta_L^\pi(\mu) \geq \sqrt{2} \cdot \sigma(\mu).$$

Similarly, we get,

$$\theta_R^\pi(\mu) \geq \sqrt{2} \cdot \sigma(\mu).$$

Now observe that,

$$\lim_{\delta\to 0} \frac{\widehat{\mu}_R^\pi(N_\delta, \delta) - \widehat{\mu}_L^\pi(N_\delta, \delta)}{\sqrt{\frac{\log(1/\delta)}{N_\delta}}} = \lim_{\delta\to 0} \frac{\widehat{\mu}_R^\pi(N_\delta, \delta) - \mu}{\sqrt{\frac{\log(1/\delta)}{N_\delta}}} + \lim_{\delta\to 0} \frac{\mu - \widehat{\mu}_L^\pi(N_\delta, \delta)}{\sqrt{\frac{\log(1/\delta)}{N_\delta}}} \xrightarrow{p} \theta_L^\pi(\mu) + \theta_R^\pi(\mu).$$

Using the fact shown above that, $\theta_L^\pi(\mu) \geq \sqrt{2} \cdot \sigma(\mu)$ and $\theta_R^\pi(\mu) \geq \sqrt{2} \cdot \sigma(\mu)$, we get the desired result.

□

Now we state the result which shows that our policy π_1 has the fastest rate of convergence of width to zero in the complete learning regime.

Theorem 7. *The policy π_1 has the following properties:*

a) $\pi_1 \in \Pi_{\text{CI}}^{sr}$. b) If $\lim_{\delta\to 0} \frac{N_\delta}{\log(1/\delta)} = \infty$, then we have,

$$\lim_{\delta\to 0} \frac{\widehat{\mu}_R^{\pi_1}(N_\delta, \delta) - \widehat{\mu}_L^{\pi_1}(N_\delta, \delta)}{\sqrt{\frac{\log(1/\delta)}{N_\delta}}} \xrightarrow{p} \sqrt{8} \cdot \sigma(\mu).$$

Proof of the Theorem 7.

Observe that from the definition of $\widehat{\mu}_L^{\pi_1}(n,\delta)$, $\widehat{\mu}_R^{\pi_1}(n,\delta)$ and the strict quasi-convexity of $d(\mu,\cdot)$, we get,

$$d(\hat{\mu}_{N_\delta}^{\pi_1}, \widehat{\mu}_L^{\pi_1}(N_\delta,\delta)) = d(\hat{\mu}_{N_\delta}^{\pi_1}, \widehat{\mu}_R^{\pi_1}(N_\delta,\delta)) = \frac{\log(2/\delta)}{N_\delta}. \tag{20}$$

First, note that $\hat{\mu}_{N_\delta}^{\pi_1}$ is the sample average based on N_δ observations. Since $N_\delta \to \infty$ as $\delta \to 0$, the central limit theorem applies. In particular,

$$\sqrt{N_\delta}\,(\hat{\mu}_{N_\delta}^{\pi_1} - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2(\mu)),$$

This implies that

$$\sqrt{N_\delta}\,(\hat{\mu}_{N_\delta}^{\pi_1} - \mu) = O_p(1).$$

Now consider

$$\frac{\sqrt{N_\delta}(\mu - \hat{\mu}_{N_\delta}^{\pi_1})}{\sqrt{\log(1/\delta)}} = \left(\sqrt{N_\delta}(\mu - \hat{\mu}_{N_\delta}^{\pi_1})\right) \cdot \frac{1}{\sqrt{\log(1/\delta)}}.$$

Since $\delta \to 0$, we have $\log(1/\delta) \to \infty$. Therefore, using Slutsky's theorem (or equivalently, the fact that $O_p(1) \cdot o(1) \xrightarrow{p} 0$), we obtain

$$\lim_{\delta\to 0} \frac{\sqrt{N_\delta}(\mu - \hat{\mu}_{N_\delta}^{\pi_1})}{\sqrt{\log(1/\delta)}} \xrightarrow{p} 0. \tag{21}$$

Now, to prove that $\pi_1 \in \Pi_{\text{CI}}^{sr}$ and the (b) part of the Theorem, we perform a Taylor series expansion of $d(\hat{\mu}_{N_\delta}^{\pi_1}, \cdot)$ in (20), we get,

$$\lim_{\delta\to 0} \frac{\hat{\mu}_{N_\delta}^{\pi_1} - \widehat{\mu}_L^{\pi_1}(N_\delta,\delta)}{\sqrt{\frac{\log(1/\delta)}{N_\delta}}} \xrightarrow{p} \sqrt{2}\cdot\sigma(\mu) \text{ and } \lim_{\delta\to 0} \frac{\widehat{\mu}_R^{\pi_1}(N_\delta,\delta) - \hat{\mu}_{N_\delta}^{\pi_1}}{\sqrt{\frac{\log(1/\delta)}{N_\delta}}} \xrightarrow{p} \sqrt{2}\cdot\sigma(\mu).$$

Using (21) and the above, we get,

$$\lim_{\delta\to 0} \frac{\mu - \widehat{\mu}_L^{\pi_1}(N_\delta,\delta)}{\sqrt{\frac{\log(1/\delta)}{N_\delta}}} \xrightarrow{p} \sqrt{2}\cdot\sigma(\mu) \text{ and } \lim_{\delta\to 0} \frac{\widehat{\mu}_R^{\pi_1}(N_\delta,\delta) - \mu}{\sqrt{\frac{\log(1/\delta)}{N_\delta}}} \xrightarrow{p} \sqrt{2}\cdot\sigma(\mu). \tag{22}$$

Using Theorem 2, we know that $\pi_1 \in \Pi_{\text{CI}}^{s}$. Hence, using (22), we get that $\pi_1 \in \Pi_{\text{CI}}^{sr}$. Further, it follows that,

$$\lim_{\delta\to 0} \frac{\widehat{\mu}_R^{\pi_1}(N_\delta,\delta) - \widehat{\mu}_L^{\pi_1}(N_\delta,\delta)}{\sqrt{\frac{\log(1/\delta)}{N_\delta}}} = \lim_{\delta\to 0} \frac{\widehat{\mu}_R^{\pi_1}(N_\delta,\delta) - \mu}{\sqrt{\frac{\log(1/\delta)}{N_\delta}}} + \lim_{\delta\to 0} \frac{\mu - \widehat{\mu}_L^{\pi_1}(N_\delta,\delta)}{\sqrt{\frac{\log(1/\delta)}{N_\delta}}} \xrightarrow{p} \sqrt{8}\cdot\sigma(\mu).$$

This completes the proof.

□

D Proofs of results in Section 6.

Proof of Theorem 3.

Recall that,

$$\tau_\delta = \sup\{n \in \mathbb{Z}^+ : \sum_{i=1}^{n} c_i \le C_\delta\}.$$

Using elementary renewal theorem for the renewal process, as $C_\delta \to \infty$, we get,

$$\lim_{\delta \to 0} \frac{\mathbb{E}[\tau_\delta]}{C_\delta} = \frac{1}{\bar{c}} \text{ almost surely.} \tag{23}$$

Consider any $\pi \in \hat{\Pi}^s_{\text{CI}}$. As the policy π is stable, it follows that for a given distribution $\nu \in \mathbf{S}$ with mean μ and a cost distribution $\mathcal{C}$ with mean $\bar{c}$, we have,

$$\lim_{\delta \to 0} \hat{\mu}^\pi_L(\tau_\delta, \delta) \xrightarrow{p} \mu^\pi_L(\mu) \text{ and } \lim_{\delta \to 0} \hat{\mu}^\pi_R(\tau_\delta, \delta) \xrightarrow{p} \mu^\pi_R(\mu).$$

We now define the set of alternate environments as $K(\mu^\pi_L(\mu), \mu^\pi_R(\mu)) = K_1(\mu^\pi_L(\mu)) \cup K_2(\mu^\pi_R(\mu))$, where

$$K_1(\mu^\pi_L(\mu)) = \{\tilde{\nu} : \tilde{\nu} \in \mathbf{S}, \tilde{\mu} < \mu^\pi_L(\mu)\}$$

and

$$K_2(\mu^\pi_R(\mu)) = \{\tilde{\nu} : \tilde{\nu} \in \mathbf{S}, \tilde{\mu} > \mu^\pi_R(\mu)\}.$$

It is worth noting that to define alternate environments, we also need to choose the cost distribution. We choose the cost distribution to be the same as $\mathcal{C}$, with mean $\bar{c}$, in the alternate environments.

Observe that $\tau_\delta + 1$ is a stopping time. Using the data processing inequality and Wald's lemma (see Lemma 0.1 in Kaufmann (2020)), for a given $\delta \in (0, 1)$ and any alternate environment $\tilde{\nu}$ with mean $\tilde{\mu}$, we have, using the independence of ν and $\tilde{\nu}$ with respect to $\mathcal{C}$, that:

$$\mathbb{E}[\tau_\delta + 1] \cdot d(\mu, \tilde{\mu}) \geq \sup_{\mathcal{E}_\delta \in \mathcal{F}_{\tau_\delta + 1}} \phi(\mathbb{P}_\nu(\mathcal{E}_\delta), \mathbb{P}_{\tilde{\nu}}(\mathcal{E}_\delta).$$

It can be re-written as,

$$\frac{C_\delta}{\log(1/\delta)} \frac{\mathbb{E}[\tau_\delta + 1]}{C_\delta} \cdot d(\mu, \tilde{\mu}) \geq \frac{\sup_{\mathcal{E}_\delta \in \mathcal{F}_{\tau_\delta + 1}} \phi(\mathbb{P}_\nu(\mathcal{E}_\delta), \mathbb{P}_{\tilde{\nu}}(\mathcal{E}_\delta))}{\log(1/\delta)}.$$

Choose $\mathcal{E}_\delta = \{\tilde{\mu} \notin [\widehat{\mu}^\pi_L(\tau_\delta, \delta), \widehat{\mu}^\pi_R(\tau_\delta, \delta)]\}$. Observe that, $\mathcal{E}_\delta$ is a measurable event in $\mathcal{F}_{\tau_\delta} \subseteq \mathcal{F}_{\tau_\delta + 1}$.

Using (23) and the arguments similar to the proof of Theorem 1, we get the desired result.

□

Proof of Theorem 4.

First, we show that

$$\mathbb{P}_\nu(\mu \notin [\widehat{\mu}^{\hat{\pi}_1}_L(\tau_\delta, \delta), \widehat{\mu}^{\hat{\pi}_1}_R(\tau_\delta, \delta)]) \leq \delta. \tag{24}$$

Using Kaufmann and Koolen (2021), we get,

$$\mathbb{P}_\nu(\exists n \in \mathbb{N} : \, n d(\hat{\mu}_n, \mu) \geq \beta(n, \delta)) \leq \delta, \tag{25}$$

where $\beta(n, \delta)$ is defined in Section 6. Observe that,

$$\{\mu \notin [\widehat{\mu}^{\hat{\pi}_1}_L(\tau_\delta, \delta), \widehat{\mu}^{\hat{\pi}_1}_R(\tau_\delta, \delta)]\} \subseteq \{\tau_\delta d(\hat{\mu}_{\tau_\delta}, \mu) \geq \beta(\tau_\delta, \delta)\}.$$

Further, the following holds as well,

$$\{\tau_\delta d(\hat{\mu}_{\tau_\delta}, \mu) \geq \beta(\tau_\delta, \delta)\} \subseteq \{\exists n \in \mathbb{N} : \, n d(\hat{\mu}_n, \mu) \geq \beta(n, \delta)\}.$$

Using (25), we get that (24) holds.

Observe that from the definition of $\widehat{\mu}^{\hat{\pi}_1}_L(\tau_\delta, \delta)$, $\widehat{\mu}^{\hat{\pi}_1}_R(n, \delta)$ and the strict quasi-convexity of $d(\mu, \cdot)$, we get,

$$d(\hat{\mu}_{\tau_\delta}, \widehat{\mu}^{\hat{\pi}_1}_L(\tau_\delta, \delta)) = d(\hat{\mu}_{\tau_\delta}, \widehat{\mu}^{\hat{\pi}_1}_R(\tau_\delta, \delta)) = \frac{\log(2/\delta)}{\tau_\delta}.$$

It can be re-written as,

$$d(\hat{\mu}_{\tau_\delta}, \widehat{\mu}_L^{\hat{\pi}_1}(\tau_\delta, \delta)) = d(\hat{\mu}_{\tau_\delta}, \widehat{\mu}_R^{\hat{\pi}_1}(\tau_\delta, \delta)) = \frac{\log(2/\delta)}{C_\delta} \frac{C_\delta}{\tau_\delta}.$$

Taking $\delta \to 0$ and using the joint-continuity of $d(\mu, x)$ in (μ, x) and (23), we get that $\pi_1 \in \hat{\Pi}^s_{\text{CI}}$ and part (b), (c) of this theorem holds. This completes the proof.

□

E Proofs of results in Section 7.

Proof of Theorem 5.

We prove the result for $\nu \in \mathbf{B}$. A similar proof holds for the case when $\nu \in \mathbf{H}$.

Recall that,

$$\tau_\delta = \sup\{n \in \mathbb{Z}^+ : \sum_{i=1}^{n} c_i \leq C_\delta\}.$$

Again, using elementary renewal theorem for the renewal process, as $C_\delta \to \infty$, we get,

$$\lim_{\delta \to 0} \frac{\mathbb{E}[\tau_\delta]}{C_\delta} = \frac{1}{\overline{c}} \text{ almost surely.} \tag{26}$$

Consider any $\pi \in \hat{\Pi}^s_{\text{CI}}$. As the policy π is stable, it follows that for a given distribution $\nu \in \mathbf{B}$ with mean μ and a cost distribution $\mathcal{C}$ with mean $\overline{c}$, we have,

$$\lim_{\delta \to 0} \hat{\mu}_L^\pi(\tau_\delta, \delta) \xrightarrow{p} \mu_L^\pi(\nu) \text{ and } \lim_{\delta \to 0} \hat{\mu}_R^\pi(\tau_\delta, \delta) \xrightarrow{p} \mu_R^\pi(\nu).$$

We now define the set of alternate environments as $K(\mu_L^\pi(\nu), \mu_R^\pi(\nu)) = K_1(\mu_L^\pi(\nu)) \cup K_2(\mu_R^\pi(\nu))$, where

$$K_1(\mu_L^\pi(\nu)) = \{\tilde{\nu} : \tilde{\nu} \in \mathbf{B}, m(\tilde{\nu}) < \mu_L^\pi(\nu)\}$$

and

$$K_2(\mu_R^\pi(\nu)) = \{\tilde{\nu} : \tilde{\nu} \in \mathbf{B}, m(\tilde{\nu}) > \mu_R^\pi(\nu)\}.$$

It is worth noting that to define alternate environments, we also need to choose the cost distribution. We choose the cost distribution to be the same as $\mathcal{C}$, with mean $\overline{c}$, in the alternate environments.

Observe that $\tau_\delta + 1$ is a stopping time. Using the data processing inequality and Wald's lemma, for a given $\delta \in (0, 1)$ and any alternate environment $\tilde{\nu}$ with mean $\tilde{\mu}$, we have, using the independence of ν and $\tilde{\nu}$ with respect to $\mathcal{C}$:

$$\mathbb{E}[\tau_\delta + 1] \cdot KL(\nu, \tilde{\nu}) \geq \sup_{\mathcal{E} \in \mathcal{F}_{\tau_\delta + 1}} \phi(\mathbb{P}_\nu(\mathcal{E}), \mathbb{P}_{\tilde{\nu}}(\mathcal{E})).$$

First we choose $\tilde{\nu} \in K_1(\mu_L^\pi(\nu))$. Observe that $\inf_{\tilde{\nu} \in K_1(\mu_L^\pi(\nu))} KL(\nu, \tilde{\nu}) = KL_{\inf}(\nu, \mathbf{B}, \mu_L^\pi(\nu))$. Hence, it can be re-written as,

$$\frac{C_\delta}{\log(1/\delta)} \frac{\mathbb{E}[\tau_\delta + 1]}{C_\delta} \cdot KL_{\inf}(\nu, \mathbf{B}, \mu_L^\pi(\nu)) \geq \frac{\sup_{\mathcal{E}_\delta \in \mathcal{F}_{\tau_\delta + 1}} \phi(\mathbb{P}_\nu(\mathcal{E}_\delta), \mathbb{P}_{\tilde{\nu}}(\mathcal{E}_\delta))}{\log(1/\delta)}.$$

Choose $\mathcal{E}_\delta = \{\tilde{\mu} \notin [\widehat{\mu}_L^\pi(\tau_\delta, \delta), \widehat{\mu}_R^\pi(\tau_\delta, \delta)]\}$. Observe that, $\mathcal{E}_\delta$ is a measurable event in $\mathcal{F}_{\tau_\delta} \subseteq \mathcal{F}_{\tau_\delta + 1}$.

Using (26) and similar to those in the proof of Theorem 1, we get the desired result.

□

Remark 4. *Now we give an equivalent version of Theorem 4 that holds for the policy* π_1^{b}.

Theorem 8. *For* $\nu \in \mathbf{B}$ *and a cost distribution* $\mathcal{C}$ *with mean* $0 < \overline{c} < \infty$*, the policy* π_1^{b} *has following properties:*

a) $\pi_1^{\mathrm{b}} \in \hat{\Pi}^s_{\mathrm{CI}}$.

b) If $\lim_{\delta\to 0} \frac{C_\delta}{\log(1/\delta)} \to k$ *for* $k \in (0,\infty)$*, then we have,* $\mu_R^{\pi_1^{\mathrm{b}}}(\nu) = \mu_R^*(\nu, k, \overline{c})$*, and* $\mu_L^{\pi_1^{\mathrm{b}}}(\nu) = \mu_L^*(\nu, k, \overline{c})$ *where,* $\mu_L^*(\nu, k, \overline{c}) < \mu$ *and* $\mu_R^*(\nu, k, \overline{c}) > \mu$ *uniquely solve* (10).

c) If $\lim_{\delta\to 0} \frac{C_\delta}{\log(1/\delta)} \to 0$*, then we have,* $\mu_R^{\pi_1^{\mathrm{b}}}(\nu) - \mu_L^{\pi_1^{\mathrm{b}}}(\nu) = 0$.

Proof of Theorem 8.

Recall that in the proof of Theorem 4, we need two properties of $d(\mu, x)$ function. First is that, $d(\mu, x)$ is a strictly quasi-convex function in x. The second one is the joint continuity of $d(\mu, x)$ function in (μ, x). Furthermore, we require an equivalent concentration bound in this setting, as in (4). At last, we also needed the convergence of $\hat{\mu}_{N_\delta}$ to μ.

Observe that from the definitions of $\mu_L^{\pi_1^{\mathrm{b}}}(n,\delta)$, $\mu_R^{\pi_1^{\mathrm{b}}}(n,\delta)$ and the strict quasi-convexity of $KL_{\inf}(\hat{\nu}_n, \mathbf{B}, \cdot)$, we get,

$$KL_{\inf}(\hat{\nu}_{N_\delta}, \mathbf{B}, \mu_L^{\pi_1^{\mathrm{b}}}(n,\delta)) = KL_{\inf}(\hat{\nu}_{N_\delta}, \mathbf{B}, \mu_R^{\pi_1^{\mathrm{b}}}(n,\delta)) = \frac{\beta(N_\delta, \delta)}{N_\delta}.$$

Using results in Appendix F below, we get that, $KL_{\inf}(\nu, \mathbf{B}, x)$ is a strictly convex function in $x \in (0,1)$ and $KL_{\inf}(\nu, \mathbf{B}, x)$ is a jointly continuous function in (ν, x) for $\nu \in \mathbf{B}$ and $x \in (0,1)$. We also know that, empirical distribution $\hat{\nu}_{N_\delta}$ weakly converges to ν and $\hat{\nu}_{N_\delta} \in \mathbf{B}$. Further, using (11), we know that we get a valid CI in both fixed sample size (N) and fixed cost budget (C_δ) setting.

Now, using the arguments given in the proof of Theorem 4, we get the desired result. It is worth noting that one can prove an equivalent version of Theorem 2 as well.

□

E.1 Asymptotic optimality of KLinf-based construction of CI for $\nu \in \mathbf{H}$

We first define $\mathbf{H}$ formally. For a given $\varepsilon > 0$ and $\Gamma > 0$. Let

$$\mathbf{H} := \Big\{\kappa \in \mathbf{P}(\mathbb{R}) :\ \mathbb{E}_{X\sim\kappa}\big[\,|X|^{1+\varepsilon}\big] \le \Gamma\Big\}.$$

Similar to the case of $\nu \in \mathbf{B}$, we utilize the CI constructed in Section 3.3.4 of Agrawal (2022). Let $\hat{\nu}_n$ denotes the empirical distribution after n samples and $\beta(n,\delta) = 1 + \log\left(\frac{2(1+n)^2}{\delta}\right)$. We denote this method as π_1^{h}. Let, $\mu_R^{\pi_1^{\mathrm{h}}}(n,\delta) \triangleq \max\{q > m(\hat{\nu}_n) :\quad n\,KL_{\inf}(\hat{\nu}_n, \mathbf{H}, q) \le \beta(n,\delta)\}$, and $\mu_L^{\pi_1^{\mathrm{h}}}(n,\delta) \triangleq \min\{q < m(\hat{\nu}_n) : n\,KL_{\inf}(\hat{\nu}_n, \mathbf{H}, q) \le \beta(n,\delta)\}$. It follows that the reported CI is $[\mu_L^{\pi_1^{\mathrm{h}}}(n,\delta), \mu_R^{\pi_1^{\mathrm{h}}}(n,\delta)]$.

The underlying concentration bound for the above CI construction (see Section 3.3.4 of Agrawal (2022)) is

$$\mathbb{P}_\nu(\exists\, n \in \mathbb{N} :\ n\,KL_{\inf}(\hat{\nu}_n, \mathbf{H}, \mu) \ge \beta(n,\delta)) \le \delta. \tag{27}$$

Remark 5. *Now we give an equivalent version of Theorem 4 that holds for the policy* π_1^{h}.

Theorem 9. *For* $\nu \in \mathbf{H}$ *with* $\mathbb{E}_{X\sim\nu}[|X|^{1+\varepsilon}] < \Gamma$*, the policy* π_1^{h} *has following properties:*

a) $\pi_1^{\mathrm{h}} \in \hat{\Pi}^s_{\mathrm{CI}}$.

b) If $\lim_{\delta\to 0} \frac{C_\delta}{\log(1/\delta)} \to k$ *for* $k \in (0,\infty)$*, then we have,* $\mu_R^{\pi_1^{\mathrm{h}}}(\nu) = \mu_R^*(\nu, k, \overline{c})$*, and* $\mu_L^{\pi_1^{\mathrm{h}}}(\nu) = \mu_L^*(\nu, k, \overline{c})$ *where,* $\mu_L^*(\nu, k, \overline{c}) < \mu$ *and* $\mu_R^*(\nu, k, \overline{c}) > \mu$ *uniquely solve* (10).

c) If $\lim_{\delta\to 0} \frac{C_\delta}{\log(1/\delta)} \to 0$*, then we have,* $\mu_R^{\pi_1^{\mathrm{h}}}(\nu) - \mu_L^{\pi_1^{\mathrm{h}}}(\nu) = 0$.

Proof of Theorem 9.

Using results in Appendix G, we get that $KL_{\inf}(\nu, \mathbf{H}, x)$ is a strictly convex function in $x \in (0,1)$ and $KL_{\inf}(\nu, \mathbf{H}, x)$ is a jointly continuously function in (ν, x) for $\nu \in \mathbf{H}$ and $x \in (0,1)$. We also know that, empirical distribution $\hat{\nu}_n$ weakly converges to ν. It is worth noting that the empirical distribution, denoted by $\hat{\nu}_n$, may not always belong to $\mathbf{H}$. However, since the theorem assumes that $\mathbb{E}_{X\sim\nu}[|X|^{1+\varepsilon}] < \Gamma$, along each sample path, $\hat{\nu}_n \in \mathbf{H}$ will eventually hold.

To see this, observe that

$$\mathbb{E}_{X\sim\hat{\nu}_n}[|X|^{1+\varepsilon}] = \frac{1}{n}\sum_{i=1}^{n} |X_i|^{1+\varepsilon}.$$

By the strong law of large numbers, we have that, for almost every sample path,

$$\frac{1}{n}\sum_{i=1}^{n} |X_i|^{1+\varepsilon} \longrightarrow \mathbb{E}_{X\sim\nu}[|X|^{1+\varepsilon}] < \Gamma.$$

Hence, for each sample path, there exists an N such that for all $n \geq N$,

$$\mathbb{E}_{X\sim\hat{\nu}_n}[|X|^{1+\varepsilon}] \leq \Gamma.$$

Therefore, along every sample path, the empirical distribution $\hat{\nu}_n$ eventually lies in $\mathbf{H}$. It is worth noting that the above argument is needed because we only have continuity of $KL_{\inf}(\nu, \mathbf{H}, x)$ in ν for $\nu \in \mathbf{H}$. Further, using (27), we know that we get a valid CI in both fixed sample size (N) and fixed cost budget (C_δ) setting.

Now using the arguments given in the proof of Theorem 4, we get the desired result. It is worth noting that one can prove an equivalent version of Theorem 2 as well.

F Properties of $KL_{\inf}(\nu, \mathbf{B}, x)$

For $\nu \in \mathbf{B}$, we define

$$KL^U_{\inf}(\nu, \mathbf{B}, x) = \inf_{\kappa\in\mathbf{B}: m(\kappa)\geq x} KL(\nu, \kappa), \quad \text{and}$$
$$KL^L_{\inf}(\nu, \mathbf{B}, x) = \inf_{\kappa\in\mathbf{B}: m(\kappa)\leq x} KL(\nu, \kappa).$$

Recall that we had defined $KL_{\inf}(\nu, \mathbf{B}, x)$ in Section 7. Since $KL^U_{\inf}(\nu, \mathbf{B}, x) = 0$ if $m(\nu) \geq x$ and $KL^L_{\inf}(\nu, \mathbf{B}, x) = 0$ if $m(\nu) \leq x$, it follows that

$$KL_{\inf}(\nu, \mathbf{B}, x) = \max\{KL^L_{\inf}(\nu, \mathbf{B}, x), KL^U_{\inf}(\nu, \mathbf{B}, x)\}. \tag{28}$$

We now state some properties of $KL^U_{\inf}(\nu, \mathbf{B}, x)$ and $KL^L_{\inf}(\nu, \mathbf{B}, x)$.

Dual Representation of $KL^U_{\inf}(\nu, \mathbf{B}, x)$ and $KL^L_{\inf}(\nu, \mathbf{B}, x)$: In the literature, it is well established that dual representations of $KL^U_{\inf}(\nu, \mathbf{B}, x)$ and $KL^L_{\inf}(\nu, \mathbf{B}, x)$ are much more tractable. We now rewrite Theorem 3 of Jourdan et al. (2022). For $(\lambda, \nu, x) \in \mathbb{R} \times \mathbf{B} \times [0,1]$, define

$$H^+(\lambda, \nu, x) = \mathbb{E}_\nu\big[\log(1 - \lambda(X - x))\big],$$

where we define $\log(z) = -\infty$ for $z \leq 0$. Similarly, define

$$H^-(\lambda, \nu, x) = \mathbb{E}_\nu\big[\log(1 + \lambda(X - x))\big].$$

Theorem 10. *For all $\nu \in \mathbf{B}$ and $x \in (0,1)$, we have*

$$KL^U_{\inf}(\nu, \mathbf{B}, x) = \sup_{\lambda\in[0,1/(1-x)]} H^+(\lambda, \nu, x), \text{ and}$$

$$KL^L_{\inf}(\nu, \mathbf{B}, x) = \sup_{\lambda\in[0,1/x]} H^-(\lambda, \nu, x).$$

It is well-known in the literature that in the above dual representation, it is a univariate convex optimization problem and can be computed efficiently by iterative methods such as Newton's method (see Honda and Takemura (2010)).

Now we re-write some properties of $KL^U_{\inf}(\nu, \mathbf{B}, x)$ and $KL^L_{\inf}(\nu, \mathbf{B}, x)$ functions which are proven in Honda and Takemura (2010) and Jourdan et al. (2022). It is worth noting that for continuity of $KL^U_{\inf}(\nu, \mathbf{B}, x)$ and $KL^L_{\inf}(\nu, \mathbf{B}, x)$ in $\mathbf{B}$, we endow it with the topology of weak convergence.

F.1 Properties of $KL^U_{\inf}(\nu, \mathbf{B}, x)$ and $KL^L_{\inf}(\nu, \mathbf{B}, x)$:

1. The function $KL^U_{\inf}(\nu, \mathbf{B}, x)$ (resp. $KL^L_{\inf}(\nu, \mathbf{B}, x)$) is continuous on $\mathbf{B} \times [0, 1)$ (resp. $\mathbf{B} \times (0, 1]$).
2. The function $x \mapsto KL^U_{\inf}(\nu, \mathbf{B}, x)$ is strictly convex on $(m(\nu), 1]$. Further, the function $x \mapsto KL^L_{\inf}(\nu, \mathbf{B}, x)$ is strictly convex on $[0, m(\nu))$.

The continuity in the above properties is in this topology on $\mathbf{B}$.

G Properties of $KL_{\inf}(\nu, \mathbf{H}, x)$

Let

$$M := \big[-\Gamma^{\frac{1}{1+\varepsilon}},\ \Gamma^{\frac{1}{1+\varepsilon}} \big], \quad M^\circ = \mathrm{int}(M).$$

For $\nu \in \mathbf{H}$, we define

$$\begin{aligned} KL^U_{\inf}(\nu, \mathbf{H}, x) &= \inf_{\kappa \in \mathbf{H}: m(\kappa) \geq x} KL(\nu, \kappa), \quad \text{and} \\ KL^L_{\inf}(\nu, \mathbf{H}, x) &= \inf_{\kappa \in \mathbf{H}: m(\kappa) \leq x} KL(\nu, \kappa). \end{aligned}$$

Recall that we had defined $KL_{\inf}(\nu, \mathbf{H}, x)$ in Section 7. Since $KL^U_{\inf}(\nu, \mathbf{H}, x) = 0$ if $m(\nu) \geq x$ and $KL^L_{\inf}(\nu, \mathbf{H}, x) = 0$ if $m(\nu) \leq x$, it follows that

$$KL_{\inf}(\nu, \mathbf{H}, x) = \max\{KL^L_{\inf}(\nu, \mathbf{H}, x), KL^U_{\inf}(\nu, \mathbf{H}, x)\}. \tag{29}$$

Dual functions. For $x \in M^\circ$, $\lambda = (\lambda_1, \lambda_2) \in \mathbb{R}^2$, $\gamma = (\gamma_1, \gamma_2) \in \mathbb{R}^2$, and $X \in \mathbb{R}$, define

$$g^U(X, \lambda, x) \ := \ 1 \ - \ \lambda_1(X - x) \ - \ \lambda_2\big(\Gamma - |X|^{1+\varepsilon}\big),$$

$$g^L(X, \gamma, x) \ := \ 1 \ + \ \gamma_1(X - x) \ - \ \gamma_2\big(\Gamma - |X|^{1+\varepsilon}\big).$$

Feasible dual sets.

$$S^U_\gamma(x) \ := \ \Big\{\lambda_1 \geq 0,\ \lambda_2 \geq 0 :\ 1 + \lambda_1 x - \lambda_2 \gamma \ - \ \frac{\varepsilon\, \lambda_1^{1+\frac{1}{\varepsilon}}}{(1+\varepsilon)^{1+\frac{1}{\varepsilon}}\, \lambda_2^{\frac{1}{\varepsilon}}} \ \geq \ 0\Big\},$$

$$S^L_\gamma(x) \ := \ \Big\{\gamma_1 \geq 0,\ \gamma_2 \geq 0 :\ 1 - \gamma_1 x - \gamma_2 \gamma \ - \ \frac{\varepsilon\, \gamma_1^{1+\frac{1}{\varepsilon}}}{(1+\varepsilon)^{1+\frac{1}{\varepsilon}}\, \gamma_2^{\frac{1}{\varepsilon}}} \ \geq \ 0\Big\}.$$

We now re-write the Theorem 4.5 of Agrawal (2022).

Theorem 11. *Let $\nu \in \mathbf{H}$ and $x \in M^\circ$. Then*

$$KL^U_{\inf}(\nu, \mathbf{H}, x) \ = \ \max_{\lambda \in S^U_\gamma(x)} \ \mathbb{E}_\nu\big[\log\big(g^U(X, \lambda, x)\big)\big]\,,$$

$$KL^L_{\inf}(\nu, \mathbf{H}, x) \ = \ \max_{\gamma \in S^L_\gamma(x)} \ \mathbb{E}_\nu\big[\log\big(g^L(X, \gamma, x)\big)\big]\,.$$

Table 2: Average CI widths over 1000 i.i.d. datasets.

Cost Budget C	Avg. CI width
500	0.492
1000	0.355
2000	0.255
3000	0.199

The above dual representation is a bivariate convex optimization problem and can be computed efficiently by iterative methods. We now restate the relevant parts of Lemma 4.9, Lemma 4.10 and Lemma 4.11 of Agrawal (2022).

Lemma 1. *For $\eta \in \mathbf{H}$ and $x \in M^o$ such that $x > m(\eta)$, $KL^U_{\inf}(\eta, \mathbf{H}, x)$ is a strictly convex function of x.*

Lemma 2. *For $\eta \in \mathbf{H}$ and $x \in M^o$ such that $x < m(\eta)$, $KL^L_{\inf}(\eta, \mathbf{H}, x)$ is a strictly convex function of x.*

Lemma 3. *When restricted to $\mathcal{H} \times M^o$, $KL^U_{\inf}$ and $KL^L_{\inf}$ are jointly continuous in their arguments.*

The continuity in the above Lemma is in the topology of weak convergence on $\mathbf{H}$.

H Numerical study for the case when $\nu \in \mathbf{H}$

We now simulate our policy π^h_1. The experiment is conducted on Pareto distributions with the scale parameter $x_m = 1$ and the shape parameter $\alpha = 3$. For the description of $\mathbf{H}$, we have chosen $\epsilon = 1$ and $\Gamma = 4$. We run the experiment in the set-up where we have costly samples with cost budget $C \in \{500, 1000, 2000, 3000\}$. We assume the cost distribution to be of uniform distributon in $[0, 2]$, i.e., $\mathcal{C} = Unif[0, 2]$ and δ is set to be 5%.

We generated 1000 i.i.d. datasets of size $C \in \{2000, 3000\}$, computed CI widths, and report the average width (max 95% CI width: 0.01) in Table 2.

I One-sided CI setting

Let $\Pi^s_{\text{CI}_\text{L}}$ denote the collection of stable policies for constructing one-sided CIs of the form $[\widehat{\mu}^\pi_L(N, \delta), \overline{\mu})$ for the mean after observing $X_1, X_2, \ldots, X_N$ for any $\delta \in (0, 1)$. For any policy $\pi \in \Pi^s_{\text{CI}_\text{L}}$ and a given distribution ν with mean μ, the following condition must hold for any $\delta \in (0, 1)$: $\forall n \in \mathbb{N} : \mathbb{P}_\nu(\mu \in [\widehat{\mu}^\pi_L(n, \delta), \overline{\mu})) \geq 1 - \delta$, and $\lim_{\delta \to 0} \widehat{\mu}^\pi_L(N_\delta, \delta) \xrightarrow{p} \mu^\pi_L(\nu)$, where $\mu^\pi_L(\nu) \leq \mu$ is a constant. Here, we define $\mu - \mu^\pi_L(\nu)$ as the "half-width" of the one-sided CI of the mean. Now we state the result that characterizes the three learning regimes for this setting.

Corollary 1. *For a given $\nu \in \mathbf{S}$ with mean μ, and any $\pi \in \Pi^s_{\text{CI}_\text{L}}$, the following holds:*

1. ***No learning regime:*** *If $\lim_{\delta \to 0} \frac{N_\delta}{\log(1/\delta)} \to 0$, then $\lim_{\delta \to 0}[\mu - \widehat{\mu}^\pi_L(N_\delta, \delta)] \xrightarrow{p} \mu - \underline{\mu}$.*

2. ***Sufficient learning regime:*** *If $\lim_{\delta \to 0} \frac{N_\delta}{\log(1/\delta)} \to k$ for $k \in (0, \infty)$, then $\mu - \mu^\pi_L(\mu) \geq \mu - \mu^*_L(\mu, k)$, where $\mu^*_L(\mu, k) < \mu$ uniquely solves $d(\mu, \mu^*_L(\mu, k)) = \frac{1}{k}$.*

The above results can similarly be applied to one-sided CIs where the goal is to ensure the upper bound of the interval is below a threshold by replacing all occurrences of subscripts L with R in the definitions and results.

Proof of Corollary 1.

Proof of this corollary follows similar to the proof of Theorem 1.

Remark 6. *It is worth noting that π_1 can be used to construct an asymptotically optimal one-sided CI. Observe that the one-sided CI using π_1 is given by $[\hat{\mu}^{\pi_1}_L(N, \delta), \overline{\mu})$, where $\hat{\mu}^{\pi_1}_L(N, \delta)$ is defined in (5).*

Remark 7. *The results for one-sided CI can be extended for the non-parametric and costly sample cases as well trivially.*

J Non-asymptotic analysis of CI Width for our policies

We provide a non-asymptotic analysis of CI width for our policies for both the canonical exponential-family case and the non-parametric (bounded) distributional case.

Exponential family case. Fix $\nu \in \mathbf{S}$ and let μ be its mean. Let the variance function be $\sigma^2(\mu)$.

Recall, under policy π_1, we have

$$d(\hat{\mu}_n,\, \hat{\mu}_R^{\pi_1}(n,\delta)) = d(\hat{\mu}_n,\, \hat{\mu}_L^{\pi_1}(n,\delta)) = \frac{\log(2/\delta)}{n}.$$

Using a Taylor-series expansion of $d(\mu, x)$ around $x = \mu$, we get

$$\frac{(\hat{\mu}_R^{\pi_1}(n,\delta) - \hat{\mu}_n)^2}{2\,\sigma^2(c_{n,R})} = \frac{(\hat{\mu}_n - \hat{\mu}_L^{\pi_1}(n,\delta))^2}{2\,\sigma^2(c_{n,L})} = \frac{\log(2/\delta)}{n},$$

for some $c_{n,R} \in \big[\hat{\mu}_n,\, \hat{\mu}_R^{\pi_1}(n,\delta)\big]$ and $c_{n,L} \in \big[\hat{\mu}_L^{\pi_1}(n,\delta),\, \hat{\mu}_n\big]$.

Hence, we obtain the width

$$\hat{\mu}_R^{\pi_1}(n,\delta) - \hat{\mu}_L^{\pi_1}(n,\delta) = \sqrt{\frac{2\log(2/\delta)}{n}}\,\Big(\sigma(c_{n,R}) + \sigma(c_{n,L})\Big).$$

Assume the mean space contains a compact interval I such that $\hat{\mu}_L^{\pi_1}(n,\delta),\, \hat{\mu}_n,\, \hat{\mu}_R^{\pi_1}(n,\delta) \in I$. Define $\sigma_{\max} = \sup_{\mu\in I}\sigma(\mu) < \infty$. Then the width is bounded by

$$\hat{\mu}_R^{\pi_1}(n,\delta) - \hat{\mu}_L^{\pi_1}(n,\delta) \;\le\; 2\,\sigma_{\max}\sqrt{\frac{2\log(2/\delta)}{n}}.$$

Non-parametric (bounded) case. Now, we perform a similar analysis for fixed $\nu \in \mathbf{B}$. Let $m(\nu)$ denote the mean of ν. Recall, under policy π_1^b, we have

$$n\,KL_{\inf}\Big(\hat{\nu}_n,\, \mathbf{B},\, \hat{\mu}_R^{\pi_1^b}(n,\delta)\Big) = n\,KL_{\inf}\Big(\hat{\nu}_n,\, \mathbf{B},\, \hat{\mu}_L^{\pi_1^b}(n,\delta)\Big) = \beta(n,\delta).$$

Using Proposition 1 in Orabona and Jun (2023), we have

$$KL_{\inf}(\rho,\, \mathbf{B},\, x) \;\ge\; \phi(m(\rho),\, x),$$

where $\phi(p,x) = p\log\frac{p}{x} + (1-p)\log\frac{1-p}{1-x}$. Using Pinsker's inequality, we have

$$KL_{\inf}(\rho,\, \mathbf{B},\, x) \;\ge\; \phi(m(\rho),\, x) \;\ge\; 2\,(m(\rho) - x)^2, \qquad \forall\,\rho \in \mathbf{B},\; x \in [0,1].$$

Applying this with $\rho = \hat{\nu}_n$ and $x = \hat{\mu}_R^{\pi_1^b}(n,\delta)$ yields

$$\frac{1}{n}\,\beta(n,\delta) = KL_{\inf}\Big(\hat{\nu}_n,\, \mathbf{B},\, \hat{\mu}_R^{\pi_1^b}(n,\delta)\Big) \;\ge\; 2\,\Big(\hat{\mu}_R^{\pi_1^b}(n,\delta) - \hat{\mu}_n\Big)^2.$$

A similar argument holds for $x = \hat{\mu}_L^{\pi_1^b}(n,\delta)$. Hence,

$$\hat{\mu}_R^{\pi_1^b}(n,\delta) - \hat{\mu}_L^{\pi_1^b}(n,\delta) = \Big(\hat{\mu}_R^{\pi_1^b}(n,\delta) - \hat{\mu}_n\Big) + \Big(\hat{\mu}_n - \hat{\mu}_L^{\pi_1^b}(n,\delta)\Big) \;\le\; \sqrt{\frac{2\,\beta(n,\delta)}{n}}.$$

Using $\beta(n,\delta) = 1 + \log\Big(\frac{2(1+n)}{\delta}\Big)$, we obtain the explicit bound

$$\hat{\mu}_R^{\pi_1^b}(n,\delta) - \hat{\mu}_L^{\pi_1^b}(n,\delta) \;\le\; \sqrt{\frac{2\Big(1 + \log\Big(\dfrac{2(1+n)}{\delta}\Big)\Big)}{n}}.$$

K Heuristic connections of Cramér–Rao lower bound to our lower bound

In this section, we draw a heuristic connection between the Cramér–Rao lower bound and our lower bound in the canonical one-parameter exponential family setting. In a canonical one-parameter exponential family, i.e., **S** parameterized by its mean μ, the Fisher information equals

$$I(\mu) = \frac{1}{\sigma^2(\mu)},$$

where $\sigma^2(\mu)$ denotes the variance of the true distribution ν with mean μ. Accordingly, the KL divergence admits the local expansion

$$d(\mu, \mu') = \frac{1}{2} I(\mu)(\mu' - \mu)^2 + o\big((\mu' - \mu)^2\big)\,.$$

It follows that

$$d(\mu, \mu') = \frac{1}{2\sigma^2(\mu)}(\mu' - \mu)^2 + o\big((\mu' - \mu)^2\big)\,.$$

In the sufficient learning regime, the optimal endpoints satisfy

$$d(\mu,\, \mu_L^*(\mu, k)) = d(\mu,\, \mu_R^*(\mu, k)) = \frac{1}{k},$$

and the quadratic approximation yields

$$|\mu_R^*(\mu, k) - \mu| \approx |\mu - \mu_L^*(\mu, k)| \approx \sqrt{\frac{2\sigma^2(\mu)}{k}},$$

so that

$$\mu_R^*(\mu, k) - \mu_L^*(\mu, k) \approx 2\sqrt{\frac{2\sigma^2(\mu)}{k}}.$$

A similar expression arises by combining the Cramér–Rao lower bound with the classical CLT. Let $\hat{\mu}_N$ be any regular estimator for which

$$\sqrt{N}\,(\hat{\mu}_N - \mu) \xrightarrow{d} \mathcal{N}(0,\, v(\mu)).$$

By the Cramér–Rao lower bound, $v(\mu) \geq 1/I(\mu) = \sigma^2(\mu)$. In the sufficient learning regime with $N = k\log(1/\delta)$, a CLT-based CI of half-width $z_{1-\delta/2}\sqrt{v(\mu)/N}$ has limiting width

$$2z_{1-\delta/2}\sqrt{\frac{v(\mu)}{N}} \;\approx\; 2\sqrt{\frac{2\log(1/\delta)\, v(\mu)}{k\log(1/\delta)}} = 2\sqrt{\frac{2\, v(\mu)}{k}} \;\geq\; 2\sqrt{\frac{2\sigma^2(\mu)}{k}},$$

with equality when $v(\mu) = \sigma^2(\mu)$, i.e., when the estimator is efficient. This heuristically confirms that the KL-based CI attains the Cramér–Rao optimal limiting width in the sufficient learning regime.