A Distribution Optimization Framework for Confidence Bounds of Risk Measures

Hao Liang [1 2] **Zhi-quan Luo** [1 2]

Abstract

We present a distribution optimization framework that significantly improves confidence bounds for various risk measures compared to previous methods. Our framework encompasses popular risk measures such as the entropic risk measure, conditional value at risk (CVaR), spectral risk measure, distortion risk measure, equivalent certainty, and rank-dependent expected utility, which are well established in risk-sensitive decision-making literature. To achieve this, we introduce two estimation schemes based on concentration bounds derived from the empirical distribution, specifically using either the Wasserstein distance or the supremum distance. Unlike traditional approaches that add or subtract a confidence radius from the empirical risk measures, our proposed schemes evaluate a specific transformation of the empirical distribution based on the distance. Consequently, our confidence bounds consistently yield tighter results compared to previous methods. We further verify the efficacy of the proposed framework by providing tighter problem-dependent regret bound for the CVaR bandit.

1. Introduction

The conventional machine learning literature primarily relies on the expected value or mean of a random variable as the performance metric for a given algorithm. However, in certain critical applications such as finance or medical treatment, the decision-maker's focus extends beyond the expected value and emphasizes other characteristics of the distribution. For instance, a risk-averse portfolio manager may place greater importance on tail behavior than expected value. To capture this risk-aware perspective, the decision-maker selects a risk measure (RM) as an alternative to the expected value, effectively representing their specific attitude towards risk.

In practice, however, it is often infeasible to directly evaluate the risk measure of the unknown underlying distribution. Instead, we must rely on constructing the point estimator based on finite samples. Consequently, the confidence interval that quantifies the coverage of the true risk measure becomes crucial in the risk-sensitive setting, as it certifies a trustworthy range for the decision maker.

In this paper, we aim to derive confidence bounds for several classes of risk measures: the Conditional Value at Risk (CVaR), the spectral risk measure (SRM), the distortion risk measure (DRM), the entropic risk measure (ERM), the certainty equivalent (CE), and the rank-dependent expected utility (RDEU). In safety-critical applications, such as medical treatment, CVaR is widely used, which represents the expected value within a fraction of the worst outcomes. Despite its practical utility, CVaR exhibits limitations in terms of expressing various risk preferences, as it assigns equal weight to all losses beyond a certain threshold. To address this, the SRM offers a notable generalization by incorporating a non-constant weighting function, enhancing flexibility in risk assessment. DRM came from insurance problems and later applied to investment risks. It encompasses CVaR as a special case and has gained attention in various fields. ERM is a well-known risk measure in mathematical finance and Markovian decision processes. Furthermore, CE serves as a generalization of ERM by replacing the exponential utility function with a more flexible function. This adaptation enhances the model's capability to capture a broader range of risk preferences. RDEU contributes to understanding decision-making under uncertainty and has been widely applied in diverse domains such as finance, psychology, and health economics [1].

In the existing literature, the confidence interval is commonly obtained through the concentration inequality, which bounds the deviation between the point estimator and the true risk with high probability. This deviation, referred to as the confidence radius, depends on the sample size and confi-

[1]School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen [2]Shenzhen Research Institute of Big Data. Correspondence to: Hao Liang <haoliang1@link.cuhk.edu.cn>.

Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

[1]For more descriptions about these risk measures, please refer to Section 2 and Appendix B.

dence level. Conventionally, the upper or lower confidence bound is determined by adding or subtracting the confidence radius from the point estimator. In this paper, we present two innovative approaches that construct confidence bounds for risk measures without relying on concentration inequalities. Our main contribution is summarized as follows.

(1) We propose a unified framework to obtain refined confidence bounds for several classes of risk measures, specifically for bounded distributions. We recast the problem of determining the confidence bound for risk measures based on finite samples as a constrained optimization problem. In particular, we optimize the value of risk measure over a confidence ball of distributions centered around the empirical distribution function (EDF). Furthermore, we obtain the closed-form solution that can be viewed as a transformation of the EDF. We set the confidence bound as the optimal solution's risk measure value. Notably, the computational overhead increases only marginally.

(2) We introduce a new baseline approach that leverages the *local* Lipschitz constant of a risk measure over the confidence ball, which may be of independent interest. In contrast, the previous bounds rely on the *global* Lipschitz constant over the entire space of bounded distributions. In addition, we suggest a systematic way to compute the local Lipschitz constant and show that our bounds outperform the new baseline approach in certain scenarios.

(3) As a minor contribution, we propose a meta-algorithm that handles generic risk measures. Specifically, the meta-algorithm specializes in the `CVaR-UCB` algorithm (Tamkin et al., 2019) for CVaR bandit problems. Interestingly, Tamkin et al. (2019) empirically observes that `CVaR-UCB` outperforms the global Lipschitz constant-based algorithm `U-UCB` (Cassel et al., 2018) with an order of magnitude improvement. Still, they only provide a regret bound that matches that of `U-UCB`. We fill this gap by providing an improved regret upper bound, quantifying the magnitude of improvement.

1.1. Related Work

Confidence bounds of risk measures The concentration of CVaR has been extensively explored in the literature, cf. Brown (2007); Wang & Gao (2010); Thomas & Learned-Miller (2019); Kolla et al. (2019); Prashanth et al. (2020); LA & Bhat (2022). The first three references primarily focus on the bounded distributions, while the remaining references consider unbounded distributions, including the sub-Gaussian, sub-exponential and heavy tail distributions. Pandey et al. (2019); LA & Bhat (2022) provide tail bounds for bounded, sub-Gaussian, or sub-exponential distributions. The concentration bounds for DRM, CE, and RDEU are presented in LA & Bhat (2022).

Lipschitz constant-based methods Kock et al. (2021); LA & Bhat (2022) relate the estimation error to the Wasserstein distance between the true and empirical distributions and then use concentration bounds for the latter. LA & Bhat (2022) establishes concentration bounds for empirical estimates for a broad class of risk measures, including CVaR, SRM, DRM, RDEU, etc. They derive the concentration bounds via the global Lipschitz constant of the risk measure over the Wasserstein distance for bounded, sub-Gaussian, and sub-exponential distributions. Our bounds only apply to bounded distributions, but we demonstrate that our bounds are tighter than their results whenever they are valid. The computation of the global Lipschitz constant can be challenging, particularly for highly nonlinear risk measures. In many cases, one may only obtain its upper bound as a surrogate, which further loosens the resulting bounds. In contrast, our framework does not require knowledge of the Lipschitz constant. Kock et al. (2021) obtain the concentration bounds for general functionals using the supremum distance instead of the Wasserstein distance. While their work primarily focuses on inequality, poverty, and welfare measures, their methodology can be extended to encompass the risk measures mentioned above. The resulting bounds apply to bounded distributions and are looser than ours. In addition, Liang & Luo (2023) focuses on risk-sensitive reinforcement learning with dynamic risk measures and leverages the Lipschitz property of risk measures to derive regret upper bounds. By quantifying the Lipschitz constants, Liang & Luo (2023) provide regret bounds that depend on these constants.

Off-policy risk evaluation Chandak et al. (2021); Huang et al. (2021) study the off-policy evaluation of functionals of reward or return distribution in bandit or RL setting. Chandak et al. (2021) formulates the problem of interval estimation for various functionals as a constrained optimization problem over a confidence band, which bears similarity to Formulation 6 in our paper. Meanwhile, our work differs from Chandak et al. (2021) in two aspects. First, Chandak et al. (2021) focuses on various functionals and derives the optimal solution for different functionals by a *case-by-case geometric* analysis. In particular, their method applies to the mean, variance, quantiles, inter-quantile range, CVaR, and entropy. In contrast, our framework focuses on general risk measures, including but not limited to ERM, CVaR, SRM, DRM, CE, and RDEU. We leverage the intrinsic property of risk measures, namely monotonicity, to derive closed-form optimal solutions that are common across different risk measures. In particular, our derivation for confidence bounds of CVaR differs from that in Chandak et al. (2021). Notably, our work is complementary to Chandak et al. (2021) in terms of the applicability of functionals. Our framework can handle arbitrary risk measures using a common optimal solution, while Chandak et al. (2021) provides confidence

bounds for CVaR and other functionals that are not risk measures, where the optimal solution depends on the specific functional. Huang et al. (2021) deal with the off-policy evaluation of Lipschitz risk measures based on their global Lipschitz constant with respect to the supremum distance.

The rest of the paper is organized as follows. We introduce some basic concepts and notations in section 2. We present our new framework under the Wasserstein distance and the supremum distance in section 3, and suggest the closed-form solution in Section 4. We then provide a new baseline method, which bridges our framework and the previous global Lipschitz constant-based method in Section 5. We validate the proposed framework by applying it to the risk-sensitive bandit problems in Section 6, and provide numerical experiments in Section 7. Finally, we provide the concluding remarks in Section 8.

2. Preliminaries

We introduce some notations here. Let $a < b$ be two real numbers. We denote by $\mathscr{D}([a,b])$ and $\mathscr{D}$ the space of all cumulative distribution functions (CDFs) supported on $[a,b]$ and the space of all CDFs on reals respectively. For a CDF $F \in \mathscr{D}$, let $X_1, X_2, \cdots, X_n$ be n i.i.d. samples from F. We denote by F_n the empirical distribution function corresponding to these samples:

$$F_n(\cdot) \triangleq \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\{X_i \le \cdot\} = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i},$$

where $\mathbb{I}$ is the indicator function and δ is the Dirac measure. We denote by $F^{-1} : (0,1] \mapsto \mathbb{R}$ is the inverse distribution function (IDF) of F, i.e., the quantile function $F^{-1}(y) \triangleq \inf\{x \in \mathbb{R} : F(x) \ge y\}$.

Supremum distance For two CDFs $F, G \in \mathscr{D}$, the supremum distance between them is defined as

$$\|F - G\|_\infty \triangleq \sup_{x\in\mathbb{R}} |F(x) - G(x)|.$$

The DKW inequality (Dvoretzky et al., 1956; Massart, 1990) bounds the deviation of the empirical distribution from the true distribution in terms of the supremum distance with high probability.

Fact 1 (Two-sided DKW inequality). Let $\delta \in (0,1]$, then the following holds with probability at least $1-\delta$

$$\|F - F_n\|_\infty \le c_n^\infty \triangleq \sqrt{\frac{\log(2/\delta)}{2n}}, \tag{1}$$

where c_n^∞ is the concentration radius.

The DKW inequality holds for any distribution, including discrete and unbounded distributions.

Wasserstein distance For CDFs $F, G \in \mathscr{D}$, the Wasserstein distance between them is defined as

$$W_1(F,G) \triangleq \int_{-\infty}^{\infty} |F(x) - G(x)|\, dx.$$

$W_1(F,G)$ can be expressed as the ℓ_1 norm between F and G. Therefore we also write $W_1(F,G) = \|F - G\|_1$. Fournier & Guillin (2015) establishes the concentration bounds on the Wasserstein distance between the EDF and the underlying one without explicit constants. LA & Bhat (2022) gives the concentration results for sub-Gaussian distributions with explicit constants. As a corollary, Fact 2 provides the concentration bound for bounded distributions.

Fact 2. Let $F \in \mathscr{D}([a,b])$. With probability at least $1-\delta$, for every $n \ge \log(1/\delta)$

$$\|F - F_n\|_1 \le c_n^1 \triangleq \frac{256(b-a)}{\sqrt{n}} + 8(b-a)\sqrt{\frac{e\log(1/\delta)}{n}} \tag{2}$$

where c_n^1 is the *concentration radius*.

Risk measure In this paper, we interpret the random variable as a loss instead of a reward. For two random variables $X \sim F$ and $Y \sim G$, we say that Y dominates X if $\forall x \in \mathbb{R}, F(x) \ge G(x)$, and we write $Y \succeq X$. A risk measure $\mathbf{T}$ is defined as a functional mapping from a set of r.v.s $\mathscr{X}$ to the reals that satisfy the following conditions (Föllmer & Schied, 2010; Weber, 2006)

- Monotonicity: $X \preceq Y \Rightarrow \mathbf{T}(X) \le \mathbf{T}(Y)$
- Translation-invariance: $\mathbf{T}(X+c) = \mathbf{T}(X) + c, c \in \mathbb{R}$

A risk measure T is said to be *distribution-invariant* if $\mathbf{T}(X) = \mathbf{T}(Y)$ when X and Y follow the same distribution (Acerbi, 2002; Weber, 2006). In this paper, we only consider distribution-invariant risk measures. We write $\mathbf{T}(F) = \mathbf{T}(X)$ for simplicity. We remark that there are other functionals mapping a r.v. to a real number, e.g., the inequality measures (Kock et al., 2021) that do not satisfy the monotonicity. In this paper, we derive the confidence bounds for several classes of risk measures. It turns out that the monotonicity of risk measures plays an essential role in our optimization framework.

Table 1 summarizes the relevant risk measures considered in this paper. These risk measures are grouped into classes, namely SRM, DRM, CE, and RDEU. CVaR and ERM belong to the SRM and CE classes, respectively, based on specific choices of the weighting function ϕ and the utility function u. The specific conditions related to the definitions of these risk measures are listed below. Please refer to Appendix B for detailed descriptions.

- SRM: $\phi : [0,1] \to [0,\infty)$ is increasing and satisfying $\int_0^1 \phi(y)dy = 1$.

Table 1: List of risk measures

RM	Notation	Definition
SRM	$\boldsymbol{M_\phi}(F)$	$\int_0^1 \phi(y)F^{-1}(y)dy$
DRM	$\boldsymbol{\rho_g}(F)$	$\int_0^\infty g(1-F(x))dx$
CE	$\boldsymbol{E_u}(F)$	$u^{-1}\left(\int_{\mathbb{R}} u(x)dF(x)\right)$
RDEU	$\boldsymbol{V}(F)$	$\int_a^b v(x)dw(F(x))$
CVaR	$\boldsymbol{C_\alpha}(F)$	$\inf_{\nu\in\mathbb{R}}\left\{\nu+\frac{1}{1-\alpha}\mathbb{E}_{X\sim F}[(X-\nu)^+]\right\}$
ERM	$\boldsymbol{U_\beta}(F)$	$\frac{1}{\beta}\log\left(\int_{\mathbb{R}}\exp(\beta x)dF(x)\right)$

- DRM: $g:[0,1]\to[0,1]$ is a continuous, concave and increasing function with $g(0)=0$ and $g(1)=1$.
- CE: u is a continuous, convex, and strictly increasing function.
- RDEU: $w:[0,1]\to[0,1]$ is an increasing weight function with $w(0)=0$ and $w(1)=1$; $v:\mathbb{R}\to\mathbb{R}$ is an (unbounded) increasing differentiable function with $u(0)=0$.
- CVaR: an instance of SRM with $\phi(y)=\frac{1}{\alpha}\mathbb{I}\{y\geq 1-\alpha\}$
- ERM: an instance of CE with $u(x)=\exp(\beta x)$.

It is more convenient to represent some risk measures using IDF, e.g., SRM $M_\phi(F)=\int_0^1\phi(y)F^{-1}(y)dy$. For this reason, we overload notation and write $\mathbf{T}(F^{-1})=\mathbf{T}(F)$ whenever convenient for some $\mathbf{T}$.

3. Distribution Optimization Framework

3.1. Global Lipschitz Constant-based Approach

Fact 2 and Fact 1 present the concentration bound of the empirical distribution in terms of the Wasserstein distance and the supremum distance, respectively. They can be written in a unified way: with probability at least $1-\delta$, we have

$$\|F-F_n\|_p\leq c_n^p, \tag{3}$$

where $p=1$ indicates the Wasserstein distance and $p=\infty$ indicates the supremum distance. To relate the concentration bound of EDF to that of risk measure, Kock et al. (2021); Bhat & LA (2019) use the Lipschitz property of the risk measure, i.e., for any two CDFs $F,G\in\mathscr{D}([a,b])$, there exists $L_p(\mathbf{T})>0$ such that the risk measure $\mathbf{T}$ satisfies

$$|\mathbf{T}(F)-\mathbf{T}(G)|\leq L_p(\mathbf{T})\|F-G\|_p. \tag{4}$$

$L_p(\mathbf{T})$ is called the *global Lipschitz constant* (GLC) of $\mathbf{T}$ w.r.t. $\|\cdot\|_p$ since the inequality holds for all possible pairs of CDFs. Combining Equation 3 and Equation 4, Kock et al. (2021); Bhat & LA (2019) establish the concentration bounds of a class of Lipschitz functionals

$$\mathbf{T}(F_n)-L_p(\mathbf{T})c_n^p\leq\mathbf{T}(F)\leq\mathbf{T}(F_n)+L_p(\mathbf{T})c_n^p.$$

The quality of the above bounds relies on the tightness of $L_p(\mathbf{T})$, so the finest bounds one can get fall back on identifying the tightest GLC

$$L_p(\mathbf{T})\triangleq\sup_{G,G'\in\mathscr{D}([a,b])}\frac{\mathbf{T}(G)-\mathbf{T}(G')}{\|G-G'\|_p},$$

where we overload the notation of $L_p(\mathbf{T})$. The GLC-based approach suffers from several limitations. The GLC may not be easy to compute, especially for some highly nonlinear risk measures. In most cases, one may only obtain its upper bound as a surrogate. Meanwhile, the concentration bounds are far from optimal. The confidence bounds are set to be the product of the GLC and the confidence radius. However, the GLC is loose since it is evaluated over the whole space of bounded distributions.

3.2. Local Lipschitz Constant-based Approach

Before introducing our framework as a remedy, we propose a *new baseline* approach that improves the previous bounds. Observe that Equation 3 together with the boundedness of F can be written as the norm ball constraint

$$B_p(F_n,c_n^p)\triangleq\left\{F\mid\|F-F_n\|_p\leq c_n^p, F\in\mathscr{D}([a,b])\right\}.$$

Define the *local* Lipschitz constant (LLC) over $B_p(F_n,c_n^p)$

$$\begin{aligned}L_p(\mathbf{T};F_n,c_n^p)&\triangleq\sup_{G,G'\in B_p(F_n,c_n^p)}\frac{\mathbf{T}(G)-\mathbf{T}(G')}{\|G-G'\|_p}\\&\leq\sup_{G,G'\in\mathscr{D}([a,b])}\frac{\mathbf{T}(G)-\mathbf{T}(G')}{\|G-G'\|_p}=L_p(\mathbf{T}).\end{aligned}$$

For simplicity, we drop $\mathbf{T}$ from the Lipschitz constants. We thus obtain the tighter upper/lower confidence bound (UCB/LCB)

$$\mathbf{T}(F_n)+(-)L_p(F_n,c_n^p)c_n^p\leq(\geq)\mathbf{T}(F_n)+L_pc_n^p.$$

As sample size increases, the confidence radius c_n^p shrinks, leading to smaller LLC and sharper bounds. In contrast, the previous bounds do not adapt to the sample size.

3.3. Distribution Optimization Framework

We now propose our unified framework to derive confidence bounds for a broad range of risk measures. The idea is quite simple and intuitive. Given a risk measure, we maximize/minimize the risk measure value over the confidence ball and set the maximal/minimal value as the UCB/LCB. By recasting the problem of finding the confidence bounds

to a constrained optimization problem, we obtain the *optimal* bounds from Equation 3. Different choices of distances lead to two types of frameworks:

$$\begin{aligned} \max_{G\in\mathscr{D}([a,b])} \quad & \mathbf{T}(G) \\ \text{s.t.} \quad & \|G-F_n\|_1 \le c_n^1 \end{aligned} \tag{5}$$

and

$$\begin{aligned} \max_{G\in\mathscr{D}([a,b])} \quad & \mathbf{T}(G) \\ \text{s.t.} \quad & \|G-F_n\|_\infty \le c_n^\infty \end{aligned} \tag{6}$$

We can obtain the LCB by reverting the maximization formulation to a minimization formulation. Denote by $\overline{F_n^p}$ ($\underline{F_n^p}$) the optimal solution, then the UCB and LCB are set to be $\mathbf{T}\left(\overline{F_n^p}\right)$ and $\mathbf{T}\left(\underline{F_n^p}\right)$. In the sequel, we may drop p when the statement holds for either $p=1$ or $p=\infty$.

To demonstrate the optimality of our framework, observe that $\overline{F_n} \in B(F_n, c_n)$, therefore

$$\mathbf{T}\left(\overline{F_n}\right) \le \mathbf{T}(F_n) + L(F_n, c_n)c_n \le \mathbf{T}(F_n) + Lc_n. \tag{7}$$

$\mathbf{T}\left(\overline{F_n}\right)$ is tighter than the bound derived from the tightest LLC, our new baseline approach, which already improves the previous bounds.

One may wonder whether $\overline{F_n}$ and $\underline{F_n}$ are easy to obtain. Fortunately, we will show that they admit analytic form for almost all risk measures introduced in Section 2 in the next section. Moreover, we will use Equation 7 to quantify the tightness of our confidence bounds in Section 5. For ease of notation, we will omit $\mathscr{D}([a,b])$.

4. Closed-form Solution

The following theorems present the closed-form solutions to Formulation 5-6. The proofs are deferred to Appendix C.

Theorem 4.1. *For any risk measure satisfying the monotonicity, the optimal solution to Formulation 6 is given by*

$$\overline{F_n^\infty} = \mathbf{P}^\infty_{c_n^\infty} F_n,\ \underline{F_n^\infty} = \mathbf{N}^\infty_{c_n^\infty} F_n, \tag{8}$$

where $\mathbf{P}^\infty_c / \mathbf{N}^\infty_c : \mathscr{D}([a,b]) \to \mathscr{D}([a,b])$ *is the positive/negative operator with coefficient* $c>0$ *for the supremum distance, which is defined as follows*

$$\begin{aligned} \left(\mathbf{P}^\infty_c F\right)(x) &\triangleq \max\left\{F(x) - c\mathbb{I}\{x<b\}, 0\right\}, \\ \left(\mathbf{N}^\infty_c F\right)(x) &\triangleq \min\left\{F(x) + c\mathbb{I}\{x\ge a\}, 1\right\}. \end{aligned}$$

The supremum ball $B_\infty(F_n, c_n^\infty)$ consists of the CDFs within the area sandwiched by $\mathbf{P}^\infty_{c_n^\infty} F_n$ and $\mathbf{N}^\infty_{c_n^\infty} F_n$ (see Figure 1). Since any risk measure $\mathbf{T}$ is monotonic, and

$$\mathbf{P}^\infty_{c_n^\infty} F_n(x) \le G(x) \le \mathbf{N}^\infty_{c_n^\infty} F_n, \forall x\in\mathbb{R}, \forall G \in B_\infty(F_n, c_n^\infty)$$

then $\mathbf{P}^\infty_{c_n^\infty} F_n$ and $\mathbf{N}^\infty_{c_n^\infty} F_n$ are the maximizer and the minimizer respectively. Another interpretation is that $\mathbf{P}^\infty_{c_n^\infty}$ transports the leftmost atoms of F_n with total mass of c_n^∞ to the maximally possible atom b, while $\mathbf{P}^\infty_{c_n^\infty}$ transports the rightmost atoms of F_n with total mass of c_n^∞ to the minimally possible atom a. Although we can explicitly represent the optimal solutions in the PMF form, it is more convenient to work with the CDF form. Please refer to Appendix F for more details.

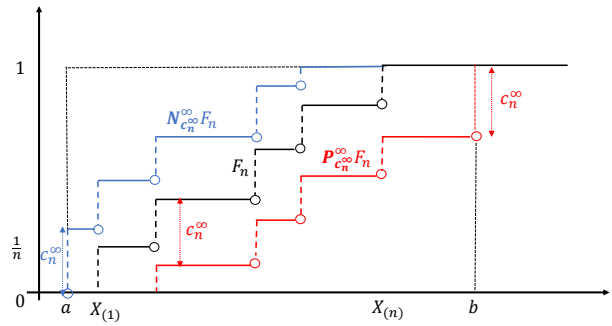


Figure 1: F_n (black), $\mathbf{P}^\infty_{c_n^\infty} F_n$ (blue) and $\mathbf{N}^\infty_{c_n^\infty} F_n$ (red).

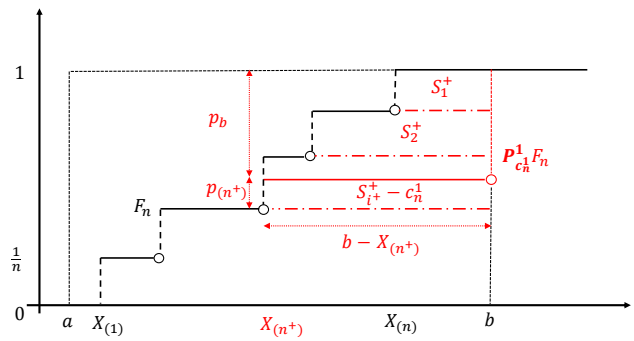


Figure 2: F_n (black) and $\mathbf{P}^1_{c_n^1} F_n$ (red). $\mathbf{P}^1_{c_n^1} F_n$ overlaps F_n for $x < X_{(n^+)}$, and it has only two jumps at $X_{(n^+)}$ and b for $x \ge X_{(n^+)}$.

Remark 4.2. The positive operator in Equation 8 reduces to the optimistic operator introduced in the CVaR bandit/RL (Tamkin et al., 2019; Keramati et al., 2020). However, they only consider the case of CVaR, and we generalize it to arbitrary risk measures.

Theorem 4.3. *For the risk measures except RDEU in Section 2, the optimal solution to Formulation 5 is given by*

$$\overline{F_n^1} = \mathbf{O}^1_{c_n^1} F_n,\ \underline{F_n^1} = \mathbf{P}^1_{c_n^1} F_n, \tag{9}$$

where $\mathbf{P}^1_c / \mathbf{N}^1_c : \mathscr{D}([a,b]) \to \mathscr{D}([a,b])$ *is called the positive/negative operator for CDF with coefficient* c *for the Wasserstein distance, which is defined as follows.*

Fix F_n and $c_n^1 > 0$. Let $X_{(1)} \le X_{(2)} \cdots \le X_{(n)}$ be the order statistic of $\{X_i\}$. For $i\in[n]$, we recursively define

$$S_1^+ \triangleq \frac{1}{n}\left(b - X_{(n)}\right),\ S_i^+ \triangleq S_{i-1}^+ + \frac{1}{n}(b - X_{(n+1-i)}).$$

The geometric interpretation of S_i^+ is the area sandwiched between F_n and the horizontal line $1-\frac{i}{n}$ (see Figure 2). Define $i^+ \triangleq \min\{i : S_i^+ \geq c_n^1\}$ as the first index that S_i^+ exceeds c_n^1. Let $n^+ \triangleq n+1-i^+$. Then $\mathbf{P}^{\mathbf{1}}_{\boldsymbol{c}_n^1} F_n$ is a categorical distribution with atoms $\{X_{(i)}\}_{i\in[n^+]} \cup \{b\}$. The probability mass of $X_{(n^+)}$ and b are assigned to be

$$p_{n^+} \triangleq \frac{1}{b-X_{n^+}}(S_{i^+}^+ - c_n^1),\ p_b \triangleq \frac{i^+}{n} - p_{n^+},$$

meanwhile the probability mass of the first n^+-1 atoms $\{X_{(i)}\}_{i\in[n^+-1]}$ remains $\frac{1}{n}$. To be more precise, $\mathbf{P}^{\mathbf{1}}_{\boldsymbol{c}_n^1} F_n$ is described by the following probability mass function (PMF)

$$\frac{1}{n}\sum_{i=1}^{n^+-1} \delta_{X_{(i)}} + p_{n^+}\cdot\delta_{X_{(n^+)}} + p_b \cdot b.$$

The way of transforming F_n into $\mathbf{P}^{\mathbf{1}}_{\boldsymbol{c}_n^1} F_n$ resembles the well-known *water-filling algorithm* (Telatar, 1999) in wireless commutations in the opposite direction. Imagine that the gravity is reversed to the upward direction, and we fill the water of amount c_n^1 to a tank enclosed by F_n and the vertical line b. The water is sequentially filled in the bins from right to left, in which the i-th bin corresponds to the $X_{(n+1-i)}$ until the water is filled up at the i^+ bin. By a volume argument, the water level is $\frac{n^+-1}{n} + p_{n^+}$. We then recover the analytic form via the shape of $\mathbf{P}^{\mathbf{1}}_{\boldsymbol{c}_n^1} F_n$.

Another interpretation is that $\mathbf{P}^{\mathbf{1}}_{\boldsymbol{c}_n^1}$ replaces the probability mass $\frac{1}{n} - p_{n^+}$ of $X_{(n^+)}$ and all the atoms to its right $\{X_{(i)}\}_{n^+<i\leq n}$ by the upper bound b. For convenience, we let $X_{(0)} = a$. We recursively define for $i \in [n]$

$$S_1^- \triangleq \frac{X_{(n)} - X_{(n-1)}}{n}, S_i^- \triangleq S_{i-1}^- + \frac{i(X_{(n+1-i)} - X_{(n-i)})}{n}.$$

Now S_i^- represents the area sandwiched between F_n and the vertical line $X_{(n-i)}$ (see Figure 3). Define $i^- \triangleq \min\{i : S_i^- \geq c_n^1\}$ as the first index that S_i^- exceeds c_n^1. Let $n^- \triangleq n+1-i^-$. Then $\mathbf{N}^{\mathbf{1}}_{\boldsymbol{c}_n^1} F_n$ is a categorical distribution with atoms $\{X_{(i)}\}_{i\in[n^--1]} \cup \{b^-\}$, where b^- is given by

$$b^- \triangleq X_{(n^--1)} + \frac{n}{i^-}(S_n^- - c_n^1).$$

$\mathbf{N}^{\mathbf{1}}_{\boldsymbol{c}_n^1} F_n$ is given by the PMF

$$\frac{1}{n}\sum_{i=1}^{n^--1} \delta_{X_{(i)}} + \frac{i^-}{n}\cdot b^-.$$

$\mathbf{N}^{\mathbf{1}}_{\boldsymbol{c}_n^1} F_n$ mirrors the water-filling, but we now fill the water rightward until it is filled out with water level b^-. It can also be interpreted as replacing the atoms to the right of $X_{(n^-)}$ with a total mass of $\frac{i^-}{n}$ by a single atom $b^- < X_{(n^-)}$.

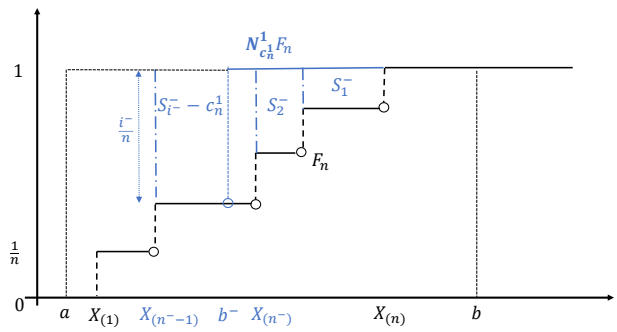


Figure 3: F_n (black) and $\mathbf{N}^{\mathbf{1}}_{\boldsymbol{c}_n^1} F_n$ (blue). $\mathbf{N}^{\mathbf{1}}_{\boldsymbol{c}_n^1} F_n$ overlaps F_n for $x < b^-$ and has a single jump at b^- with height $\frac{i^-}{n}$.

Computational issue. We present the algorithms to actualize Equation 5-6 in practice in Appendix F. We demonstrate that their computational complexity increases slightly more than the LC-based methods.

Remark 4.4. For both distances, we only require F to be bounded above by a known constant b to perform $\mathbf{P}_{\boldsymbol{c}_n} F_n$, and require F to be bounded below by a known constant a to perform $\mathbf{N}_{\boldsymbol{c}_n} F_n$. Thus we only require F to be bounded on one side to obtain the one-sided confidence bound.

5. Improvement of Confidence Bounds

5.1. Derivation of the LLC

We present a systematic way of computing the LLC over the confidence ball $B(F_n, c_n)$, which bridges our framework and the GLC-based method. Define $\boldsymbol{\psi}(t; F, G) \triangleq \mathrm{T}((1-t)F + tG)$ for $F, G \in \mathscr{D}([a,b])$ and $t \in [0,1]$. For simplicity, we may drop F, G and write $\boldsymbol{\psi}(t)$ if it is clear from the context. Note that $\boldsymbol{\psi}(0) = \mathrm{T}(F)$, $\boldsymbol{\psi}(1) = \mathrm{T}(G)$ and $(1-t)F + tG \in \mathscr{D}([a,b])$ for all $t \in [0,1]$. It can be shown that $\boldsymbol{\psi}$ is continuously differentiable under some mild conditions on $\mathbf{T}$. Observe that

$$\begin{aligned}
L(\mathbf{T}; F_n, c_n) &= \sup_{F,G\in B(F_n,c_n)} \frac{\mathbf{T}(F) - \mathbf{T}(G)}{\|F-G\|} \\
&= \sup_{F,G\in B_p(F_n,c_n)} \frac{\boldsymbol{\psi}(1; F, G) - \boldsymbol{\psi}(0; F, G)}{\|F-G\|} \\
&\leq \sup_{F,G\in B(F_n,c_n), t\in[0,1]} \frac{\boldsymbol{\psi}'(t; F, G)}{\|F-G\|} \\
&\leq \sup_{F,G\in B(F_n,c_n), t\in[0,1]} \boldsymbol{\upsilon}((1-t)F + tG) \\
&= \sup_{F\in B(F_n,c_n)} \boldsymbol{\upsilon}(F),
\end{aligned}$$

where $\boldsymbol{\upsilon}$ is a functional that satisfies for any F, G

$$\boldsymbol{\psi}'(t; F, G)) \leq \boldsymbol{\upsilon}((1-t)F + tG)\,\|F-G\|.$$

Note that $\boldsymbol{\upsilon}$ implicitly depends on p and the risk measure $\mathbf{T}$. Consequently, we can obtain an upper bound on the LLC

by bounding the last term. We can obtain the upper bound on the GLC by removing the ball constraint. For interested readers, please refer to in Table 4 in Appendix D for the functional υ for different risk measures. We will use SRM as an example to illustrate the procedure.

5.1.1. An Example: SRM

Here we consider an alternative form of SRM $\boldsymbol{M_\phi}(F) = \int_a^b \phi(F(x))xdF(x)$. ψ is continuously differentiable with derivative

$$\begin{aligned}\boldsymbol{\psi}'(t) &= \frac{d}{dt}\int_a^b \phi((1-t)F+tG)(x))xd((1-t)F+tG)(x)\\ &= -\int_a^b (G-F)(x)\phi((1-t)F+tG)(x))dx\\ &\le \|G-F\|_p \|\phi(F+tH)\|_q .\end{aligned}$$

We omit the details in the second equality and leave the full derivations to Appendix D. Since

$$\frac{\boldsymbol{\psi}'(t;F,G)}{\|F-G\|_p} \le \|\phi((1-t)F+tG)\|_q ,$$

where $\|\cdot\|_q$ is the dual norm of $\|\cdot\|_p$, we obtain $\boldsymbol{\upsilon}(F) = \|\phi(F)\|_q$. Consider the case $p=\infty$. Since $\underline{F_n^\infty} \preceq F$ for $F \in B_\infty(F_n, c_n^\infty)$, then $\phi\left(\underline{F_n^\infty}(x)\right) \ge \phi(F(x))$, $\forall F \in B_\infty(F_n, c_n^\infty)$, $\forall x \in [a,b]$. It holds that

$$\begin{aligned}\max_{F\in B_\infty(F_n,c_n^\infty)} \|\phi(F)\|_1 &= \max_{F\in B_\infty(F_n,c_n^\infty)} \int_a^b \phi(F(x))dx\\ &= \int_a^b \phi(\underline{F_n^\infty}(x))dx = \left\|\phi\left(\underline{F_n^\infty}\right)\right\|_1 .\end{aligned}$$

In contrast, the GLC can be bounded by choosing $F=\delta_a$

$$\max_{F\in\mathscr{D}([a,b])} \int_a^b \phi(F(x))dx = (b-a)\|\phi\|_\infty = (b-a)\phi(1).$$

Following such principle, we obtain the LLCs for other risk measures (cf. Table 2).

5.2. Improvement of Distribution Optimization Framework

Equation 7 qualitatively establishes that the confidence bounds derived from our framework are tighter than that based on the LLC

$$\mathbf{T}\left(\overline{F_n}\right) - \mathbf{T}(F_n) \le L(F_n, c_n)c_n.$$

Furthermore, we can quantitatively show the improvement

$$\begin{aligned}\mathbf{T}\left(\overline{F_n}\right) - \mathbf{T}(F_n) &= \boldsymbol{\psi}\left(1; F_n, \overline{F_n}\right) - \boldsymbol{\psi}\left(0; F_n, \overline{F_n}\right)\\ &\le \max_{t\in[0,1]} \boldsymbol{\psi}'\left(t; F_n, \overline{F_n}\right)\\ &\le \max_{t\in[0,1]} \boldsymbol{\upsilon}\left((1-t)F_n + t\cdot\overline{F_n}\right)c_n,\end{aligned}$$

where the second to the last inequality follows from the definition of $\boldsymbol{\upsilon}$. The following also holds

$$\mathbf{T}(F_n) - \mathbf{T}\left(\underline{F_n}\right) \le \max_{t\in[0,1]} \boldsymbol{\upsilon}\left((1-t)\underline{F_n} + tF_n\right)c_n.$$

Therefore, it is more convenient to compare $\mathbf{T}\left(\overline{F_n}\right) - \mathbf{T}(F_n)$ and $L(F_n, c_n)c_n$, which is shown for the supremum distance in Table 3. For convenience, we normalize them by c_n and state the results for general CDF F. Our UCBs are strictly and consistently tighter than the LLC-based bounds for the supremum distance. Due to the space limit, the results for the Wasserstein distance are shown in Appendix D.

5.3. Illustrating example: CVaR

We use CVaR to illustrate Table 2 and Table 3. Table 2 compares the LLC with GLC for different risk measures. The second row in Table 2 shows that the LLC and GLC of CVaR for the Wasserstein distance are identical. In addition, the GLC of CVaR for supremum distance is $\frac{b-a}{\alpha}$, which is larger than its LLC $\frac{b-F^{-1}((1-\alpha-c)^+)}{\alpha}$.

Table 3 presents the improvement of our bound for the supremum distance. The second column in Table 3 implies

$$\begin{aligned}\frac{\boldsymbol{C_\alpha}\left(\overline{F^\infty}\right) - \boldsymbol{C_\alpha}(F)}{c} &= \frac{b-F^{-1}(1-\alpha)}{1-\alpha}\\ < L_\infty(\boldsymbol{C_\alpha}; F, c) &= \frac{b-F^{-1}(1-\alpha-c)}{\alpha}\\ < L_\infty(\boldsymbol{C_\alpha}) &= \frac{b-a}{\alpha}.\end{aligned}$$

Our upper bound is strictly tighter than the LLC-based and GLC-based bound. The improvement depends on the distribution F and α. For small α and distribution F with non-fat upper tail, $b-F^{-1}(1-\alpha) \ll b-a$, leading to a much finer bound. In particular, consider a uniform distribution

$$\begin{aligned}\frac{\boldsymbol{C_\alpha}\left(\overline{F^\infty}\right) - \boldsymbol{C_\alpha}(F)}{c} &= \frac{\alpha(b-a)}{\alpha}\\ < L_\infty(\boldsymbol{C_\alpha}; F, c) &= \frac{(\alpha+c)(b-a)}{\alpha}\\ < L_\infty(\boldsymbol{C_\alpha}) &= \frac{b-a}{\alpha}.\end{aligned}$$

Our bound for uniform distribution considerably improves by a factor of $1/\alpha$ and $(\alpha+c)/\alpha$ compared to the GLC-based and LLC-based bound, respectively.

6. Application to Risk-sensitive Bandits

Now we consider the risk-sensitive multi-armed bandit (MAB) problems. The quality of each arm is measured by the risk measure value of its loss distribution. The loss distribution of the i-th arm is denoted by F_i, and the risk value associated with $\mathbf{T}$ is $\mathbf{T}(F_i)$. The algorithm interacts

Table 2: Comparison between the LLC and the GLC

RM	Local ($p=1$)	Global ($p=1$)	Improvement	Local ($p=\infty$)	Global ($p=\infty$)	Improvement
CVaR	$\frac{1}{\alpha}$	$\frac{1}{\alpha}$	✗	$\frac{b-F_n^{-1}((1-\alpha-c)^+)}{\alpha}$	$\frac{b-a}{\alpha}$	✓
SRM	$\phi(1)$	$\phi(1)$	✗	$\left\lVert\phi(\underline{F_n^\infty})\right\rVert_1$	$(b-a)\phi(1)$	✓
DRM	$\lVert g'\rVert_\infty$	$\lVert g'\rVert_\infty$	✗	$\left\lVert g'(1-\underline{F_n^\infty})\right\rVert_1$	$(b-a)\lVert g'\rVert_\infty$	✓
ERM	$\frac{\exp(\beta b)}{\int_a^b \exp(\beta x) d\underline{F_n^1}(x))}$	$\exp(\beta(b-a))$	✓	$\frac{\exp(\beta b)-\exp(\beta a)}{\beta\int_a^b \exp(\beta x) d\underline{F_n^\infty}(x))}$	$\frac{\exp(\beta(b-a))-1}{\beta}$	✓
RDEU	N/A[2]	$\lVert w'\rVert_\infty \lVert u'\rVert_\infty$	N/A	$\left\lVert w'(\underline{F_n^\infty})u'\right\rVert_1$	$\lVert w'\rVert_\infty \lVert u'\rVert_1$	✓

Table 3: Improvement of confidence bounds for supremum distance over LLC

RM	CVaR	SRM	DRM	ERM	RDEU
$L_\infty(\mathbf{T}; F, c)$	$\frac{b-F^{-1}(1-\alpha-c)}{1-\alpha}$	$\lVert\phi(\underline{F^\infty})\rVert_1$	$\lVert g'(1-\underline{F^\infty})\rVert_1$	$\frac{\exp(\beta b)-\exp(\beta a)}{\int_a^b \exp(\beta x) d\underline{F^\infty}(x)}$	$\lVert w'(\underline{F^\infty})u'\rVert_1$
$\frac{\mathbf{T}(\overline{F^\infty})-\mathbf{T}(F)}{c}$	$\frac{b-F^{-1}(1-\alpha)}{1-\alpha}$	$\lVert\phi(F)\rVert_1$	$\lVert g'(1-F)\rVert_1$	$\frac{\exp(\beta b)-\exp(\beta a)}{\beta\int_a^b \exp(\beta x) dF(x)}$	$\lVert w'(F)u'\rVert_1$
Improvement	✓	✓	✓	✓	✓

with a bandit instance $\nu = (F_i)_{i\in[K]}$ for N rounds. In each round $t \in [N]$, the algorithm π chooses an arm $I_t \in [K]$ and observes the loss $X_t \sim F_{I_t}$. The performance of an algorithm π is measured by the cumulative regret

$$\text{Regret}(\pi, \nu, N) \triangleq \mathbb{E}\left[\sum_{t\in[N]} \mathbf{T}(F_{I_t}) - \min_{i\in[K]} \mathbf{T}(F_i)\right].$$

While the UCB-type algorithms (Auer et al., 2002) are widely applied to the risk-neutral MAB problems, they rely on the concentration bound of the mean. We propose a *meta-algorithm* (cf. Algorithm 1) to deal with generic risk measures, where the LCB is derived from our framework. The algorithm maintains the EDF $\hat{F}_{i,t}$ for each arm i

$$\hat{F}_{i,t} \triangleq \frac{1}{s_i(t)} \sum_{t'=1}^{t-1} \mathbb{I}\{X_{t'} \le \cdot, I_{t'} = i\}, \tag{10}$$

where $s_i(t) \triangleq \sum_{t'=1}^{t-1} \mathbb{I}\{I_{t'} = i\}$ is the number of times of pulling arm i up to time t. For convenience, we assume the first arm is the optimal arm, i.e., $\mathbf{T}(F_1) < \mathbf{T}(F_i), i \neq 1$. When specializing the risk measure to the CVaR, we obtain a distribution-dependent regret upper bound.

Proposition 6.1. *The expected regret of Algorithm 1 on a instance $\nu \in \mathscr{D}([a,\infty])$ with $\mathbf{T} = C_\alpha$ is bounded as*

$$\textit{Regret}(\textit{LCB}, \nu, N) \le \frac{4\log(\sqrt{2}N)}{\alpha^2} \sum_{i>1}^{K} \frac{\left(b - F_i^{-1}(1-\alpha-2c_i^*)\right)^2}{\Delta_i} + 3\sum_{i=1}^{K} \Delta_i,$$

where the sub-optimality gap $\Delta_i \triangleq C_\alpha(F_i) - C_\alpha(F_1) < b-a$, and c_i^* is a F_i-dependent constant that solves the equation $2\frac{b-F_i^{-1}(1-\alpha-2c)}{\alpha}c = \Delta_i$.

Algorithm 1 `Lower Confidence Bound`

Input: N, a
for round $t \in [K]$ **do**
  Pull arm $I_t \leftarrow t$
end for
for round $t = K+1, K+2, \cdots, N$ **do**
  Compute $\hat{F}_{i,t}$ via Equation 10
  Set $c_i(t) \leftarrow \sqrt{\frac{\log(2KN^2)}{s_i(t)}}$ for all $i \in [K]$
  $\underline{F_{i,t}} \leftarrow \mathbf{N}^{\infty}_{\boldsymbol{c_i(t)}} \hat{F}_{i,t}$
  $I_t \leftarrow \arg\min_{i\in[K]} \mathbf{T}\left(\underline{F_{i,t}}\right)$
end for

The proof is deferred to Appendix E.

Remark 6.2. The meta-algorithm for CVaR reduces to the `CVaR-UCB` algorithm (Tamkin et al., 2019). Interestingly, Tamkin et al. (2019) empirically observes that `CVaR-UCB` outperforms the GLC-based algorithm `U-UCB` (Cassel et al., 2018) with an order of magnitude improvement, but they only provide a regret bound of

$$\frac{4\log(\sqrt{2}N)}{\alpha^2} \sum_{i>1}^{K} \frac{(b-a)^2}{\Delta_i} + 3\sum_{i=1}^{K} \Delta_i,$$

which matches that of `U-UCB`. We fill this gap by quantify-

ing the improvement in the magnitude

$$\sum_{i>1} \frac{\left(b - F_i^{-1}(1-\alpha-2c_i^*)\right)^2}{\Delta_i^2} \Big/ \sum_{i>1} \frac{(b-a)^2}{\Delta_i^2} < 1.$$

Remark 6.3. Baudry et al. (2021) introduces a Thompson Sampling algorithm `B-CVTS` for CVaR bandit with bounded rewards, which is the first asymptotic optimal CVaR bandit algorithm. Notably, our main contribution in this paper is the framework to improve confidence bounds rather than designing optimal CVaR bandit algorithms. Additionaly, `CVaR-UCB` has several advantages. `CVaR-UCB` can perform incremental updates to compute the CVaR values, but `B-CVTS` needs to maintain and sample from the posterior of distributions. Thus the time and space complexity of `CVaR-UCB` is quite low. Moreover, `CVaR-UCB` can be applied to semi-unbounded distributions, e.g., log-normal distribution, while `B-CVTS` assumes bounded rewards.

7. Numerical Experiments

To better visualize the benefits of our framework relative to those of LLC and GLC, we conducted a series of empirical comparisons. Details and complete figures are deferred to Appendix G.

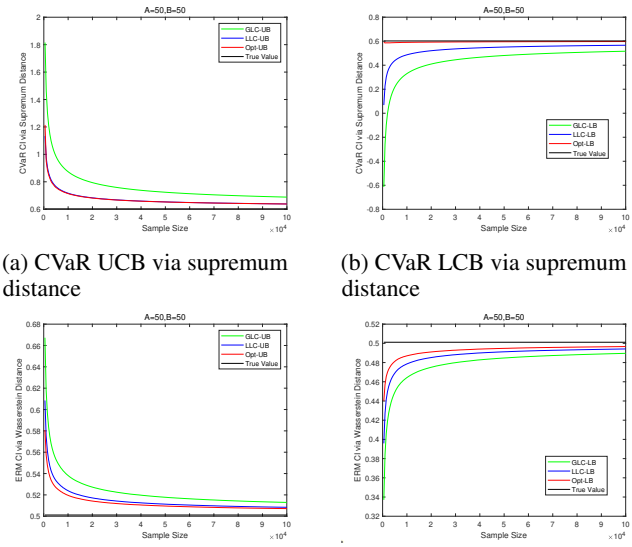
Confidence bounds. We consider five different beta distributions and two risk measures: CVaR and ERM. Due to space limitations, we provide the results for one typical beta distribution and one particular distance in Figure 4. Our bounds are consistently tighter than LC-based ones for various risk measures and varying sample sizes.

CVaR bandits. We compare `CVaR-UCB` with the UCB algorithm using GLC (`GLC-UCB`) and using LLC (`LLC-UCB`) in Figure 5. It shows that `CVaR-UCB` outperforms `LLC-UCB` and `GLC-UCB`.

8. Conclusion

We propose a distribution optimization framework to obtain improved confidence bounds of several risk measures. By viewing the solutions as certain transformations of the EDF, we design efficient algorithms to compute the confidence bounds. The tightness of our bounds is further illustrated via comparisons with the new baseline method.

The major limitation is that our framework only deals with bounded distribution in general. However, it is applicable to semi-unbounded distributions for CVaR, SRM, DRM, and RDEU. It would be interesting to study the distribution optimization framework under more general assumptions, e.g., sub-Gaussian or sub-exponential distributions. Another promising future direction is to generalize the framework to the multivariate setting. One may apply the multivariate DKW inequality to the multivariate risk measures.



(a) CVaR UCB via supremum distance

(b) CVaR LCB via supremum distance

(c) ERM UCB via Wasserstein distance

(d) ERM LCB via Wasserstein distance

Figure 4: Comparisons of CIs for CVaR and ERM with varying sample sizes.

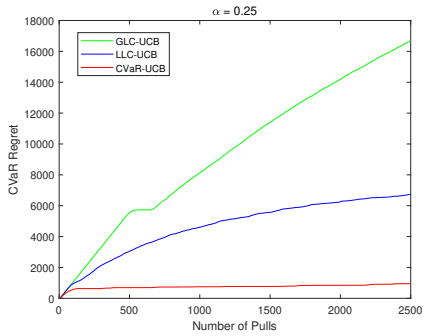


Figure 5: Cumulative CVaR-regret of `CVaR-UCB` (red), `LLC-UCB` (blue), and `GLC-UCB` (green).

Acknowledgements

We thank all the anonymous reviewers for their helpful comments and suggestions. The work of Zhi-quan Luo was supported in part by the National Key Research and Development Project under grant 2022YFA1003900 and in part by the Guangdong Provincial Key Laboratory of Big Data Computing.

References

Acerbi, C. Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking & Finance*, 26(7):1505–1518, 2002.

Acerbi, C. and Tasche, D. On the coherence of expected shortfall. *Journal of Banking & Finance*, 26(7):1487–1503, 2002.

Auer, P., Cesa-Bianchi, N., and Fischer, P. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2):235–256, 2002.

Baudry, D., Gautron, R., Kaufmann, E., and Maillard, O. Optimal thompson sampling strategies for support-aware cvar bandits. In *International Conference on Machine Learning*, pp. 716–726. PMLR, 2021.

Bäuerle, N. and Rieder, U. More risk-sensitive markov decision processes. *Mathematics of Operations Research*, 39(1):105–120, 2014.

Bhat, S. P. and LA, P. Concentration of risk measures: A wasserstein distance approach. *Advances in Neural Information Processing Systems*, 32, 2019.

Brown, D. B. Large deviations bounds for estimating conditional value-at-risk. *Operations Research Letters*, 35(6): 722–730, 2007.

Cassel, A., Mannor, S., and Zeevi, A. A general approach to multi-armed bandits under risk criteria. In *Conference On Learning Theory*, pp. 1295–1306. PMLR, 2018.

Chandak, Y., Niekum, S., da Silva, B., Learned-Miller, E., Brunskill, E., and Thomas, P. S. Universal off-policy evaluation. *Advances in Neural Information Processing Systems*, 34:27475–27490, 2021.

Dvoretzky, A., Kiefer, J., and Wolfowitz, J. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *The Annals of Mathematical Statistics*, pp. 642–669, 1956.

Föllmer, H. and Schied, A. Convex and coherent risk measures. *Encyclopedia of Quantitative Finance*, pp. 355–363, 2010.

Föllmer, H. and Schied, A. Stochastic finance. In *Stochastic Finance*. de Gruyter, 2016.

Fournier, N. and Guillin, A. On the rate of convergence in wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3):707–738, 2015.

Huang, A., Leqi, L., Lipton, Z., and Azizzadenesheli, K. Off-policy risk assessment in contextual bandits. *Advances in Neural Information Processing Systems*, 34: 23714–23726, 2021.

Keramati, R., Dann, C., Tamkin, A., and Brunskill, E. Being optimistic to be conservative: Quickly learning a cvar policy. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 4436–4443, 2020.

Kock, A. B., Preinerstorfer, D., and Veliyev, B. Functional sequential treatment allocation. *Journal of the American Statistical Association*, pp. 1–13, 2021.

Kolla, R. K., Prashanth, L., Bhat, S. P., and Jagannathan, K. Concentration bounds for empirical conditional value-at-risk: The unbounded case. *Operations Research Letters*, 47(1):16–20, 2019.

Krokhmal, P., Palmquist, J., and Uryasev, S. Portfolio optimization with conditional value-at-risk objective and constraints. *Journal of risk*, 4:43–68, 2002.

LA, P. and Bhat, S. P. A wasserstein distance approach for concentration of empirical risk estimates. *J. Machine Learn. Res*, 23(238):1–61, 2022.

Liang, H. and Luo, Z.-Q. Regret bounds for risk-sensitive reinforcement learning with lipschitz dynamic risk measures. *arXiv preprint arXiv:2306.02399*, 2023.

Massart, P. The tight constant in the dvoretzky-kiefer-wolfowitz inequality. *The annals of Probability*, pp. 1269–1283, 1990.

Pandey, A. K., Prashanth, L., and Bhat, S. P. Estimation of spectral risk measures. *arXiv preprint arXiv:1912.10398*, 2019.

Prashanth, L., Jagannathan, K., and Kolla, R. K. Concentration bounds for cvar estimation: The cases of light-tailed and heavy-tailed distributions. In *Proceedings of the 37th International Conference on Machine Learning*, pp. 5577–5586, 2020.

Quiggin, J. *Generalized expected utility theory: The rank-dependent model*. Springer Science & Business Media, 2012.

Rockafellar, R. T., Uryasev, S., et al. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000.

Tamkin, A., Keramati, R., Dann, C., and Brunskill, E. Distributionally-aware exploration for cvar bandits. In *NeurIPS 2019 Workshop on Safety and Robustness on Decision Making*, 2019.

Telatar, E. Capacity of multi-antenna gaussian channels. *European transactions on telecommunications*, 10(6):585–595, 1999.

Thomas, P. and Learned-Miller, E. Concentration inequalities for conditional value at risk. In *International Conference on Machine Learning*, pp. 6225–6233. PMLR, 2019.

Wang, S. Premium calculation by transforming the layer premium density. *ASTIN Bulletin: The Journal of the IAA*, 26(1):71–92, 1996.

Wang, S. S. Cat bond pricing using probability transforms. *Geneva Papers: Etudes et Dossiers*, 278:19–29, 2004.

Wang, Y. and Gao, F. Deviation inequalities for an estimator of the conditional value-at-risk. *Operations Research Letters*, 38(3):236–239, 2010.

Weber, S. Distribution-invariant risk measures, information, and dynamic consistency. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 16(2):419–441, 2006.

Zhu, S. and Fukushima, M. Worst-case conditional value-at-risk with application to robust portfolio management. *Operations research*, 57(5):1155–1168, 2009.

A. Table of Notation

Symbol	Explanation
$\mathscr{D}$	The space of all CDFs
$\mathscr{D}([a,b])$	The space of all CDFs supported on $[a,b]$
$B_p(F,c)$	The $\|\cdot\|_p$ norm ball centered at F with radius c
F_n	The empirical distribution function corresponding to n samples from F
c_n^p	The confidence radius w.r.t. $\|\cdot\|_p$ for n samples
$\mathbf{T}$	Risk measure
$L_p(\mathbf{T})$	The global Lipschitz constant of $\mathbf{T}$ w.r.t. $\|\cdot\|_p$
$L_p(\mathbf{T};F,c)$	The local Lipschitz constant of $\mathbf{T}$ w.r.t. $\|\cdot\|_p$ over $B_p(F_n,c_n^p)$
$\overline{F_n^p}$	The maximizer of Formulation 5 or 6
$\underline{F_n^p}$	The minimizer of Formulation 5 or 6
$\mathbf{P}_{\boldsymbol{c}}^{\mathbf{1}}$	The positive operator w.r.t. $\|\cdot\|_1$ with coefficient c
$\mathbf{N}_{\boldsymbol{c}}^{\mathbf{1}}$	The negative operator w.r.t. $\|\cdot\|_1$ with coefficient c
$\mathbf{P}_{\boldsymbol{c}}^{\boldsymbol{\infty}}$	The positive operator w.r.t. $\|\cdot\|_\infty$ with coefficient c
$\mathbf{N}_{\boldsymbol{c}}^{\boldsymbol{\infty}}$	The negative operator w.r.t. $\|\cdot\|_\infty$ with coefficient c
ν	Bandit instance
N	Number of total rounds
K	Number of total arms
π	Bandit algorithm
$s_i(t)$	The number of times of pulling arm i up to time t

B. Risk Measures

Conditional Value at Risk (CVaR) CVaR (Rockafellar et al., 2000) is a popular risk measure in financial portfolio optimization (Krokhmal et al., 2002; Zhu & Fukushima, 2009). Formally, the CVaR value at level $\alpha \in (0,1)$ for a distribution F is defined as

$$\boldsymbol{C_\alpha}(F) \triangleq \inf_{\nu\in\mathbb{R}} \left\{ \nu + \frac{1}{1-\alpha}\mathbb{E}_{X\sim F}[(X-\nu)^+] \right\}.$$

Acerbi & Tasche (2002) showed that when F is a continuous distribution, $C_\alpha(F) = \mathbb{E}_{X\sim F}[X|X \geq F^{-1}(1-\alpha)]$.

Spectral risk measure (SRM) SRM is a generalization of CVaR that adopts a non-constant weighting function (Acerbi, 2002). The SRM of F is defined as

$$\boldsymbol{S_\phi}(F) \triangleq \int_0^1 \phi(y)F^{-1}(y)dy,$$

where $\phi : [0,1] \to [0,\infty)$ is weighting function. ϕ is said to be admissible if it is increasing and satisfies that $\int_0^1 \phi(y)dy = 1$. Acerbi (2002) showed that $S_\phi(F)$ is a coherent risk measure if ϕ is admissible. SRM can be viewed as a weighted average of the quantiles F^{-1}, with weight specified by $\phi(y)$. In fact, $\boldsymbol{S_\phi}(F)$ specializes in $C_\alpha(F)$ for $\phi(y) = \frac{1}{1-\alpha}\mathbb{I}\{y \geq 1-\alpha\}$.

Distortion risk measure (DRM) DRM is originally from the insurance problems and later applied to investment risks (Wang, 1996; 2004). For a distribution $F \in \mathscr{D}([0,\infty))$, the DRM $\rho_g(F)$ is defined as

$$\boldsymbol{\rho_g}(F) \triangleq \int_0^\infty g(1-F(x))dx,$$

where $g : [0,1] \to [0,1]$ is a continuous increasing function with $g(0) = 0$ and $g(1) = 1$. We refer to g as the distortion function. Similar to SRM, DRM can also recover CVaR by setting $g(y) = \min\left(\frac{x}{1-\alpha}, 1\right)$.

Entropic risk measure (ERM) ERM is a well-known risk measure in risk-sensitive decision-making, including mathematical finance (Föllmer & Schied, 2016), Markovian decision processes (Bäuerle & Rieder, 2014). The ERM value of F

with coefficient $\beta \neq 0$ is defined as

$$\boldsymbol{U_\beta}(F) \triangleq \frac{1}{\beta}\log(\mathbb{E}_{X\sim F}[\exp(\beta X)]) = \frac{1}{\beta}\log\left(\int_{\mathbb{R}}\exp(\beta x)dF(x)\right).$$

Certainty equivalent Certainty equivalent can be viewed as a generalization of ERM, which replace the exponential utility function with a more general function. Let u be a continuous and strictly increasing function such that its inverse u^{-1} exists, then the certainty equivalent $C_u(F)$ of F associated with a function u is given by

$$\boldsymbol{C_u}(F) \triangleq u^{-1}(\mathbb{E}_{X\sim F}[u(X)]) = u^{-1}\left(\int_{\mathbb{R}}u(x)dF(x)\right).$$

It is trivial that the certainty equivalent $C_u(F)$ reduces to the ERM when $u(x) = \exp(\beta x)$.

Rank dependent expected utility (RDEU) RDEU value (Quiggin, 2012) of $F \in \mathscr{D}([a,b])$ is defined as

$$\boldsymbol{V}(F) \triangleq \int_a^b v(x)dw(F(x)),$$

where $w : [0,1] \to [0,1]$ is an increasing weight function such that $w(0) = 0$ and $w(1) = 1$, and $v : \mathbb{R} \to \mathbb{R}$ be an (unbounded) increasing differentiable function with $v(0) = 0$.

C. Proof of Theorems

For CDFs $F, G \in \mathscr{D}$, the Wasserstein distance between them can be represented by their IDFs

$$W_1(F,G) = \int_{-\infty}^{\infty}|F(x) - G(x)|\,dx = \int_0^1 \left|F^{-1}(y) - G^{-1}(y)\right|dy,$$

With slight abuse of notation, we write $W_1(F,G) = \left\|F^{-1} - G^{-1}\right\|_1$. We will prove the theorems for the more general formulations in the following.

$$\begin{aligned} \max_{G\in\mathscr{D}([a,b])} \quad & \mathbf{T}(G) \\ \text{s.t.} \quad & \|G - F\|_1 \leq c \end{aligned} \tag{11}$$

and

$$\begin{aligned} \max_{G\in\mathscr{D}([a,b])} \quad & \mathbf{T}(G) \\ \text{s.t.} \quad & \|G - F\|_\infty \leq c \end{aligned} \tag{12}$$

Observe that $\|G - F\|_1 = \left\|G^{-1} - F^{-1}\right\|_1$, thus we can recast Formulation 5 as

$$\begin{aligned} \max_{G\in\mathscr{D}([a,b])} \quad & \mathbf{T}(G^{-1}) \\ \text{s.t.} \quad & \left\|G^{-1} - F^{-1}\right\|_1 \leq c \end{aligned} \tag{13}$$

Proposition C.1. *For risk measures ERM and CE defined in Section 2, the optimal solution to Formulation 11 is given by*

$$\overline{F^1} = \mathbf{P_c^1}F,\ \underline{F}^1 = \mathbf{N_c^1}F, \tag{14}$$

where $\mathbf{P_c^1}/\mathbf{N_c^1} : \mathscr{D}([a,b]) \to \mathscr{D}([a,b])$ is the positive/negative operator for CDF for the Wasserstein distance, which defined as follows.

Define $S^+(F,x) \triangleq \int_x^b (F(z) - F(x))\,dz$. Its geometric meaning is the area sandwiched between F and a constant $F(x)$ from x (see Figure 4 (a)). Notice that S^+ may be discontinuous w.r.t. x since F can be discontinuous w.r.t. x ($S^+(x) < S^+(x^-)$). For $c > 0$, we define $g^+(F,c) \triangleq \max\{x \geq a : S^+(F,x) \leq c\} \in [a,b)$. For simplicity, we drop F from the notations if it is clear from the context. Given $F \in \mathscr{D}([a,b])$ as input, O_c^1 outputs a CDF

$$\left(\mathbf{P_c^1}F\right)(x) \triangleq \begin{cases} F(g^+(c)) - \frac{c - S^+(g^+(c))}{b - g^+(c)}, x \in [g^+(c), b), \\ F(x), \text{otherwise.} \end{cases}$$

Analogously, we define $S^-(F,x) \triangleq \int_x^b (1-F(z))\,dz$ and $g^-(F,c) \triangleq \max\{x \geq a : S^-(F,x) \leq c\} \in [a,b)$. We omit F for simplicity. Note that S^- is continuous w.r.t. x. Hence $g^-(c) = \{x : S^-(x) = c\}$. For $F \in \mathscr{D}([a,b])$, $\mathbf{N_c^1}$ outputs a CDF

$$\left(\mathbf{N_c^1} F\right)(x) \triangleq \begin{cases} 1, x \in [g^-(c), b), \\ F(x), \text{otherwise}. \end{cases}$$

Proposition C.2. *For risk measures CVaR, SRM and DRM defined in Section 2, the optimal solution to 13 is given by*

$$\left(\overline{F^1}^{-1}\right) = \mathbf{P_c^1} F^{-1},\ \left(\underline{F}^1\right)^{-1} = \mathbf{N_c^1} F^{-1}, \tag{15}$$

where we overload notations to denote by $\mathbf{P_c^1}/\mathbf{N_c^1} : (\mathscr{D}([a,b]))^{-1} \to (\mathscr{D}([a,b]))^{-1}$ the positive/negative operator for IDF for the Wasserstein distance.

We overload notations and define $S^+(F^{-1},y) \triangleq \int_y^1 \left(b - F^{-1}(z)\right) dz$. For $c > 0$, we define $g^+(F^{-1},c) \triangleq \min\{y : S^+(F^{-1},y) \geq c\} \in (0,1)$. For simplicity, we drop F from the notations if it is clear from the context. Given $F^{-1} \in \mathscr{D}([a,b])^{-1}$ as input, $\mathbf{P_c^1}$ outputs a IDF

$$\left(\mathbf{P_c^1} F^{-1}\right)(y) \triangleq \begin{cases} 1, y \in (g^+(c), 1], \\ F^{-1}(y), \text{otherwise}. \end{cases}$$

Analogously, we define $S^-(F^{-1},y) \triangleq \int_y^1 \left(F^{-1}(z) - F^{-1}(y)\right) dz$ and $g^-(F^{-1},c) \triangleq \min\{y : S^-(F^{-1},y) \geq c\} \in (0,1)$ for $c > 0$. Notice that S^- may be discontinuous w.r.t. y ($S^-(y+) < S^-(y)$). Given $F^{-1} \in \mathscr{D}([a,b])^{-1}$ as input, $\mathbf{N_c^1}$ outputs a IDF

$$\left(\mathbf{N_c^1} F^{-1}\right)(y) \triangleq \begin{cases} F^{-1}(g^-(c)) + \frac{S^{-1}(g^{-1}(c)) - c}{1 - g^-(c)}, y \in (g^-(c), 1], \\ F^{-1}(y), \text{otherwise}. \end{cases}$$

Proposition C.3. *For any risk measure, the optimal solution to Formulation 12 is given by*

$$\overline{F^\infty} = \mathbf{P_c^\infty} F,\ \underline{F}^\infty = \mathbf{N_c^\infty} F, \tag{16}$$

where $\mathbf{P_c^\infty}/\mathbf{N_c^\infty} : \mathscr{D}([a,b]) \to \mathscr{D}([a,b])$ is the positive/negative operator with coefficient $c > 0$ for the supremum distance, which is defined as follows

$$\left(\mathbf{P_c^\infty} F\right)(x) \triangleq \max\left\{F(x) - c\mathbb{I}\{x \in [a,b), 0\}\right.,$$
$$\left(\mathbf{N_c^\infty} F\right)(x) \triangleq \min\left\{F(x) + c\mathbb{I}\{x \in [a,b), 1\}\right..$$

Figure 4 illustrates how the operators defined in Proposition C.1-C.3 transforms a typical continuous CDF $F \in \mathscr{D}([a,b])$.

C.1. Proof of Proposition C.1

We only provide the proof for CE because ERM is a special case of CE by choosing $u(x) = \exp(\beta x)$. For simplicity, we write the optimization problem as

$$\begin{aligned} \max_{G \in \mathscr{D}([a,b])} \quad & C_u(G) \\ \text{s.t.} \quad & \|G - F\|_1 \leq c \end{aligned}$$

Proof. We consider the maximization problem first. The objective function can be written as

$$C_u(G) = u^{-1}\left(\int_a^b u(x) dG(x)\right) = u^{-1}\left(u(b) - \int_a^b G(x) u'(x) dx\right).$$

Since u^{-1} is monotonically increasing, we can reformulate the original optimization problem as

$$\begin{aligned} \min_G \quad & \int_a^b G(x) u'(x) dx \\ \text{s.t.} \quad & G \in B_1(F, c) \end{aligned}$$

The fact that the ball constraint $G \in B_1(F, c)$ is symmetric about F implies $G^*(x) \leq F(x), \forall x \in [a, b]$. Suppose $G^*(y) > F(y)$ for y in a set that is a union of disjoint intervals $\cup_i I_i = \cup_i (a_i, b_i)$, and $G^*(y) \leq F(y)$ otherwise. We can choose

$$H(y) = \begin{cases} \max\{2F(y) - G^*(y), F(a_i)\}, y \in I_i, \\ G^*(y), \text{otherwise}, \end{cases}$$

which satisfies the ball constraint. However, $C_u(H) > C_u(G^*)$ since $\int_{I_i} H(x)u'(x)dx \leq \int_{I_i} F(x)u'(x)dx < \int_{I_i} G^*(x)\exp(\beta x)dx$, which leads to a contradiction. Hence we have $G^*(x) \leq F(x), \forall x \in [a, b]$. Define

$$\tilde{G}(x) \triangleq \left(\mathbf{P}_{\boldsymbol{c}}^{\mathbf{1}} F\right)(x) = \begin{cases} F(g^+(c)) - \frac{c - S^+(g^+(c))}{b - g^+(c)}, x \in [g^+(c), b), \\ F(x), \text{otherwise}. \end{cases}$$

Notice that the (new) minimization problem has a linear objective and convex ball constraint. Thus, the optimal solution exists in the boundary of the ball constraint. It suffices to consider the following optimization problem

$$\begin{aligned} \min_{G} \quad & \int_a^b G(x)u'(x)dx \\ \text{s.t.} \quad & \|F - G\|_1 = c, \\ & G \succeq F. \end{aligned}$$

It is easy to check that $\tilde{G}$ satisfies the constraints. It remains to show that $\int_a^b \tilde{G}(x)u'(x)dx \leq \int_a^b G(x)u'(x)dx$ for any feasible G. Consider a feasible $G \neq \tilde{G}$. It is obvious that $G(b^-) \geq \tilde{G}(g^+(c))$, otherwise $\|F - G\|_1 = \int_a^b F(x) - G(x)dx > \int_a^b F(x) - \tilde{G}(x)dx = c$, which contradicts with $G \in B_1(F, c)$. If $G(b^-) = \tilde{G}(g^+(c))$, then we again have $G = \tilde{G}$. Otherwise $\|F - G\|_1 > c$. Therefore it holds that $G(b^-) > \tilde{G}(g^+(c))$. We have $G(g^+(c)) < \tilde{G}(g^+(c))$, since otherwise $\|F - G\|_1 < \left\|F - \tilde{G}\right\|_1 = c$. Since G is monotonically increasing and right continuous, there exists $g'(c) \in (g^+(c), b)$ such that $G(x) < \tilde{G}(x) = \tilde{G}(g^+(c))$ for $x \in [g^+(c), g'(c))$ and $G(x) > \tilde{G}(x) = \tilde{G}(g^+(c))$ for $x \in (g'(c), b)$. Moreover, $G(x) \leq \tilde{G}(x) = F(x)$ for $x \in [a, g^+(c))$. It follows that

$$\begin{aligned} & \int_a^b G(x)u'(x)dx - \int_a^b \tilde{G}(x)u'(x)dx \\ & = \int_a^{g'(c)} (G(x) - \tilde{G}(x))u'(x)dx + \int_{g'(c)}^b (G(x) - \tilde{G}(x))u'(x)dx \\ & \geq u'(g'(c)) \int_{g'(c)}^b (G(x) - \tilde{G}(x))dx - u'(g'(c)) \int_a^{g'(c)} (\tilde{G}(x) - G(x))dx \\ & = 0. \end{aligned}$$

The last equality follows from that $\|F - G\|_1 - \left\|F - \tilde{G}\right\|_1 = \int_a^b F(x) - G^(x)dx - \int_a^b F(x) - \tilde{G}(x)dx = \int_a^b \tilde{G}(x) - G(x)dx = 0$. □

Similarly, we can prove that the optimal solution to the minimization problem is given by

$$\left(\mathbf{N}_{\boldsymbol{c}}^{\mathbf{1}} F^{-1}\right)(y) \triangleq \begin{cases} F^{-1}(g^-(c)) + \frac{S^{-1}(g^{-1}(c)) - c}{1 - g^-(c)}, y \in (g^-(c), 1], \\ F^{-1}(y), \text{otherwise}. \end{cases}$$

C.2. Proof of Proposition C.2

We only provide proof for SRM and DRM because CVaR is a special case of SRM or DRM.

C.2.1. Proof for SRM

Proof. The objective function is given by

$$M_\phi(G) = \int_0^1 \phi(y)G^{-1}(y)dy,$$

Table 4: $\boldsymbol{\upsilon}(F)$

RM	$\boldsymbol{\upsilon}(F)$
CVaR	$\frac{1}{\alpha}\left\|\mathbb{I}\{F(\cdot)\geq 1-\alpha\}\right\|_q$
SRM	$\left\|\phi(F)\right\|_q$
DRM	$\left\|g'(1-F)\right\|_q$
ERM	$\frac{\left\|\exp(\beta\cdot)\right\|_q}{\int_a^b u(x)dF(x)}$
CE	$\left\|u'\right\|_q (u^{-1})'(\int_a^b u(x)dF(x))$
RDEU	$\left\|w'(F)u'\right\|_q$

which is linear in G^{-1}. Meanwhile, the constraint $\left\|G^{-1}-F^{-1}\right\|_1 \leq c$ is also a convex ball constraint w.r.t. G^{-1}. Furthermore, $\phi(y)$ is an increasing function. Using analogous arguments to the proof of Theorem 4.3 completes the proof. □

C.2.2. Proof for DRM

Proof. By a change of variable $y=F(x)$, the DRM can be represented as

$$\begin{aligned}\rho_g(G)=\int_0^1 g(1-y)dG^{-1}(y)&=g(1-y)F^{-1}(y)|_0^1-\int_0^1 G^{-1}(y)dg(1-y)\\&=-a+\int_0^1 G^{-1}(y)g'(1-y)dy.\end{aligned}$$

Again, the objective function is linear in G^{-1}, and the constraint is also a convex ball constraint. Besides, $g'(1-y)$ is increasing in y since g is concave. Using analogous arguments to the proof of Theorem 4.3 completes the proof. □

C.3. Proof of Proposition C.3

It is easy to verify that $\mathbf{P}_{\boldsymbol{c}}^{\boldsymbol{\infty}}F\in B_\infty(F,c)$. Consider an arbitrary $F\in\mathscr{D}([a,b])$ and $c>0$. For $x\in[a,b)$, we have

$$\left(\mathbf{P}_{\boldsymbol{c}}^{\boldsymbol{\infty}}F\right)(x)=\max\{F(x)-c,0\}\leq G(x),\ \forall G\in B_\infty(F,c).$$

Besides, $\left(\mathbf{P}_{\boldsymbol{c}}^{\boldsymbol{\infty}}F\right)(b)=G(b)=1$ for any $G\in B_\infty(F,c)$. Therefore $G\preceq\mathbf{P}_{\boldsymbol{c}}^{\boldsymbol{\infty}}F$ for any $G\in B_\infty(F,c)$. The result follows from the monotonicity of $\mathbf{T}$.

D. Derivations of Results in Section 5

D.1. Identification of $\boldsymbol{\nu}$

We list functional $\boldsymbol{\nu}$ for different risk measures in Table 4. The functional $\boldsymbol{\nu}$ is crucial to compute the LLC and to show the tightness of our method. We provide detailed derivations of ν in the following. Since the CVaR (ERM) is a special case of SRM (CE), we omit the derivation of CVaR and ERM. Recall that $\upsilon((1-t)F+tG,p)$ is a functional satisfying that for any F,G

$$\psi'\left(t;F,G\right))\leq\upsilon((1-t)F+tG,p)\left\|F-G\right\|_p.$$

D.1.1. SRM

Consider M_ϕ in the form of

$$M_\phi(F)=\int_a^b\phi(F(x))xdF(x),$$

where ϕ is increasing and integrates to 1. ψ is continuously differentiable with derivative

$$\begin{aligned}\psi'(t;F,G) &= \frac{d}{dt}\int_a^b \phi((1-t)F+tG)(x))xd((1-t)F+tG)(x)\\ &= \frac{d}{dt}\left[\int_a^b \phi((1-t)F+tG)(x))xdF(x) + t\int_a^b \phi((1-t)F+tG)(x))xd(G-F)(x)\right]\\ &= \underbrace{\int_a^b \phi'((1-t)F+tG)(x))(G(x)-F(x))xdF(x)}_{(a)} + \underbrace{\int_a^b \phi((1-t)F+tG)(x))xd(G-F)(x)}_{(b)}\\ &\quad + \underbrace{t\int_a^b \phi'((1-t)F+tG)(x))(G(x)-F(x))xd(G-F)(x)}_{(c)}.\end{aligned}$$

Since

$$\begin{aligned}(b) &= \phi((1-t)F+tG)(x))x(G-F)(x)|_a^b - \int_a^b (G-F)(x)d\left[\phi((1-t)F+tG)(x))x\right]\\ &= -\int_a^b (G-F)(x)d\left[\phi((1-t)F+tG)(x))x\right]\end{aligned}$$

and

$$(c) = \int_a^b (G-F)(x)xd\phi((1-t)F+tG)(x)) - \int_a^b \phi'((1-t)F+tG)(x))(G(x)-F(x))xdF(x)$$

We have

$$\begin{aligned}\psi'(t) = (a)+(b)+(c) = -\int_a^b (G-F)(x)\phi((1-t)F+tG)(x))dx &= \langle G-F, -\phi((1-t)F+tG)\rangle\\ &\le \|G-F\|_p \|\phi((1-t)F+tG)\|_q.\end{aligned}$$

Hence we can choose $\upsilon(F,p) = \|\phi(F)\|_q$.

D.1.2. DRM

A distortion risk measure associated with distortion function g for a distribution F is

$$\rho_g(F) = \int_a^b g(1-F(x))dx,$$

where $g : [0,1] \to [0,1]$ is a non-decreasing function with $g(0) = 0$ and $g(1) = 1$. Thus g' is non-negative. ψ is continuously differentiable with derivative

$$\begin{aligned}\psi'(t;F,G) = \frac{d}{dt}\int_a^b g(1-(1-t)F(x)-tG(x))dx &= \int_a^b g'(1-(1-t)F(x)-tG(x))(F(x)-G(x))dx\\ &= \langle F-G, g'(1-(1-t)F-tG)\rangle\\ &\le \|F-G\|_p \|g'(1-(1-t)F-tG)\|_q.\end{aligned}$$

Hence we can choose $\upsilon(F,p) = \|g'(1-F)\|_q$.

D.1.3. CE

For a CE C_u, we define $E_u(F) \triangleq \int u(x)dF(x)$. Notice that E_u is a *linear* functional, i.e.,

$$E_u((1-t)F+tG) = (1-t)E_u(F) + tE_u(G)$$

for any $F, G \in D([a,b])$. ψ for E_u is continuously differentiable with derivative

$$\begin{aligned}
\psi'(t;F,G) = \frac{d}{dt}E_u((1-t)F+tG) &= \frac{d}{dt}(1-t)E_u(F) + tE_u(G) \\
&= E_u(G) - E_u(F) \\
&= \int_a^b u(x)dG(x) - \int_a^b u(x)dF(x) \\
&= u(x)G(x)|_a^b - \int_a^b G(x)du(x) - u(x)F(x)|_a^b + \int_a^b F(x)du(x) \\
&= \int_a^b F(x) - G(x)du(x) \\
&= \langle F - G, u' \rangle
\end{aligned}$$

where the last equality follows from that $F(b) = G(b) = 1$ and $F(a) = G(a) = 0$. The certainty equivalent of a distribution F with utility function u is defined as $C_u(F) = u^{-1}(E_u(F))$. It follows that

$$\begin{aligned}
\psi'(t;F,G) &= \left(u^{-1}\right)'(E_u((1-t)F+tG)) \cdot \langle F - G, u' \rangle \\
&\leq \left(u^{-1}\right)'(E_u((1-t)F+tG)) \cdot \|F-G\|_p, \|u'\|_q
\end{aligned}$$

Hence we can choose $\upsilon(F,p) = \left(u^{-1}\right)'(E_u(F))\|u'\|_q$.

D.1.4. RDEU

Let $w : [0,1] \to [0,1]$ be an *increasing* weight function such that $w(0) = 0$ and $w(1) = 1$. Let $u : \mathbb{R} \to \mathbb{R}$ be an (unbounded) *increasing differentiable* function with $u(0) = 0$. The RDEU value of $F \in \mathscr{D}([a,b])$ is given by

$$\begin{aligned}
V(F) &= \int_a^b u(x)dw(F(x)) \\
&= u(x)w(F(x))|_a^b - \int_a^b w(F(x))du(x) \\
&= u(b) - \int_a^b w(F(x))u'(x)dx.
\end{aligned}$$

We have

$$\begin{aligned}
\psi'(t;F,G) &= \frac{d}{dt}\left[u(b) - \int_a^b w((1-t)F(x)+tG(x))u'(x)dx\right] \\
&= -\int_a^b w'((1-t)F(x)+tG(x))(G(x)-F(x))u'(x)dx \\
&= \langle F - G, w'((1-t)F+tG)u' \rangle \\
&\leq \|F-G\|_p \|w'((1-t)F+tG)u'\|_q
\end{aligned}$$

Hence we can choose $\upsilon(F,p) = \|w'(F)u'\|_q$.

D.2. Derivation of the LLC

In Section 5.1, we have shown that

$$\begin{aligned} L_p(\mathbf{T}; F_n, c_n^p) &= \sup_{F,G\in B_p(F_n,c_n^p)} \frac{\mathbf{T}(F)-\mathbf{T}(G)}{\|F-G\|_p} \\ &= \sup_{F,G\in B_p(F_n,c_n^p)} \frac{\psi(1;F,G)-\psi(0;F,G)}{\|F-G\|_p} \\ &\le \sup_{F,G\in B_p(F_n,c_n^p),t\in[0,1]} \frac{\psi'(t;F,G)}{\|F-G\|_p} \\ &\le \sup_{F,G\in B_p(F_n,c_n^p),t\in[0,1]} \upsilon((1-t)F+tG,p) \\ &= \sup_{F\in B_p(F_n,c_n^p)} \upsilon(F,p). \end{aligned}$$

We have derived the υ function in the previous subsection, which enables us to obtain an upper bound on $L_p(\mathbf{T}; F, c)$. We only provide the derivations for DRM, CE, and RDEU here because SRM is presented as an illustrating example in Section 5.1. We first give a useful fact to deal with the Wasserstein distance.

Fact 3. Fix real numbers $a < b$. For any $G \in \mathscr{D}([a,b])$ and any c, there exists a continuous and monotonically increasing distribution $F \in \mathscr{D}([a,b])$ such that $\|F - F_n\|_1 \le c$.

D.2.1. DRM

For DRM, $\upsilon(F,p) = \|g'(1-F)\|_q$. Since g is concave, g' is monotonically decreasing.

Case $p = 1$**.** By Fact 3, there exists a continuous CDF $F \in B_1(F_n, r_n^1)$. Such F attains all possible value in $[0,1]$, hence

$$\max_{F\in B_1(F_n,c_n^1)} \|g'(1-F)\|_\infty = \max_{F\in B_1(F_n,c_n^1)} \max_{x\in\mathbb{R}} g'(1-F(x)) = \max_{y\in[0,1]} g'(y) = \|g'\|_\infty .$$

Case $p = \infty$**.** Since

$$\max_{F\in B_\infty(F_n,c_n^\infty)} \|g'(1-F)\|_1 = \max_{F\in B_\infty(F_n,c_n^\infty)} \int_a^b g'(1-F(x))dx,$$

it follows that

$$\max_{F\in B_\infty(F_n,c_n^\infty)} \int_a^b g'(1-F(x))dx = \left\|g'(1-\mathrm{N}_{c_n^\infty}^\infty F_n)\right\|_1$$

D.2.2. CE

For CE, $\upsilon(F,p) = \left(u^{-1}\right)'(E_u(F))\|u'\|_q$. Since u is convex, $\left(u^{-1}\right)'$ is a decreasing function. It follows that

$$\sup_{F\in B_p(F_n,c_n^p)} \left(u^{-1}\right)'(E_u(F))\|u'\|_q = (u^{-1})'(E_u(\mathrm{N}_{c_n^p}^p F_n))\|u'\|_q .$$

D.2.3. RDEU

For RDEU, $\upsilon(F,p) = \|w'(F)u'\|_q$. Recall that $w : [0,1] \to [0,1]$ is an increasing weight function with $w(0) = 0$ and $w(1) = 1$, and $v : \mathbb{R} \to \mathbb{R}$ is an (unbounded) increasing differentiable function with $u(0) = 0$. We further assume that w is convex so that w' is an increasing function.

Case $p = 1$**.** By Fact 3, there exists a continuous CDF $F \in B_1(F_n, r_n^1)$. Such F attains all possible value in $[0,1]$, hence

$$\max_{F\in B_1(F_n,c_n^1)} \|w'(F)u'\|_\infty = \max_{F\in B_1(F_n,c_n^1)} \max_{x\in\mathbb{R}} w'(F(x))u'(x),$$

which might not admit closed form in general. Meanwhile, the GLC (holds without monotonicity)

$$L_1 = \max_{F\in\mathscr{D}([a,b])} \|w'(F)u'\|_\infty = \|w'\|_\infty \|u'\|_\infty .$$

Case $p = \infty$. The following holds

$$\max_{F \in B_\infty(F_n, c_n^\infty)} \|w'(F)u'\|_1 = \max_{F \in B_\infty(F_n, c_n^\infty)} \int_a^b w'(F(x))u'(x)dx$$
$$= \left\| w'(\mathrm{N}_{c_n^\infty}^\infty F)u' \right\|_1 .$$

The global Lipschitz constant (holds without monotonicity)

$$L_\infty = \max_{F \in \mathscr{D}([a,b])} \|w'(F)u'\|_1 = \|w'\|_\infty \|u'\|_1 .$$

The last equality uses that for any $d \in [0,1]$, $F(x) = d$ for any $x \in (a,b)$ if $F = d\psi_a + (1-d)\psi_b$ (shifted Bernoulli).

D.3. Improved Confidence Bound

In Section 5.2, we established that

$$\begin{aligned}
\mathbf{T}\left(\mathrm{P}_{c_n^p}^p F_n\right) - \mathbf{T}(F_n) &= \psi\left(1; F_n, \mathrm{P}_{c_n^p}^p F_n\right) - \psi\left(0; F_n, \mathrm{P}_{c_n^p}^p F_n\right) \\
&\leq \max_{t \in [0,1]} \psi'\left(t; F_n, \mathrm{P}_{c_n^p}^p F_n\right) \\
&\leq \max_{t \in [0,1]} \upsilon\left((1-t)F_n + t\mathrm{P}_{c_n^p}^p F_n, p\right) c_n^p
\end{aligned}$$

as well as

$$\mathbf{T}(F_n) - \mathbf{T}\left(\mathrm{N}_{c_n^p}^p F_n\right) \leq \max_{t \in [0,1]} \upsilon\left((1-t)\mathrm{N}_{c_n^p}^p F_n + tF_n, p\right) c_n^p.$$

The above inequalities lead to the following bounds.

D.3.1. Supremum distance

Proposition D.1 (CE). *For $F \in \mathscr{D}([a,b])$, it holds that (assume $\beta > 0$)*

$$\begin{aligned}
C_u(\mathrm{P}_c^\infty F) - C_u(F) &\leq \|u'\|_1 (u^{-1})' \left(\int_a^b u(x)dF(x) \right) \cdot c \\
&\leq \|u'\|_1 (u^{-1})' \left(\int_a^b u(x)d\mathrm{N}_c^\infty F(x) \right) \cdot c, \\
C_u(F) - C_u(\mathrm{N}_c^\infty F) &\leq \|u'\|_1 (u^{-1})' \left(\int_a^b u(x)d\mathrm{N}_c^\infty F(x) \right) \cdot c.
\end{aligned}$$

Corollary D.2 (ERM). *For $F \in \mathscr{D}([a,b])$, it holds that (assume u convex)*

$$\begin{aligned}
U_\beta(\mathrm{P}_c^\infty F) - U_\beta(F) &\leq \frac{\exp(\beta b) - \exp(\beta a)}{\beta \int_a^b \exp(\beta x)dF(x)} \cdot c \leq \frac{L_\infty(U_\beta)c}{\int_a^b \exp(\beta x)d\mathrm{N}_c^\infty F(x)}, \\
U_\beta(F) - U_\beta(\mathrm{P}_c^\infty F) &\leq \frac{L_\infty(U_\beta)c}{\int_a^b \exp(\beta x)d\mathrm{N}_c^\infty F(x)}.
\end{aligned}$$

Proposition D.3 (SRM). *For $F \in \mathscr{D}([a,b])$, it holds that (assume ϕ increasing)*

$$\begin{aligned}
M_\phi(\mathrm{P}_c^\infty F) - M_\phi(F) &\leq \int_a^b \phi(F(x))dx \cdot c \leq \int_a^b \phi(\mathrm{N}_c^\infty F(x))dx \cdot c = L_\infty(M_\phi; F, c)c \\
M_\phi(F) - M_\phi(\mathrm{N}_c^\infty F) &\leq \int_a^b \phi(\mathrm{N}_c^\infty F(x))dx \cdot c = L_\infty(M_\phi; F, c)c.
\end{aligned}$$

Proposition D.4 (DRM). *For $F \in \mathscr{D}([a,b])$, it holds that (assume g concave)*

$$\rho_g(\mathrm{P}_c^\infty F) - \rho_g(F) \le \int_a^b g'(1-F(x))dx \cdot c \le \int_a^b g'(1-\mathrm{N}_c^\infty F(x))dx \cdot c = L_\infty(\rho_g; F, c)c,$$
$$\rho_g(F) - \rho_g(\mathrm{N}_c^\infty F) \le \int_a^b g'(1-\mathrm{N}_c^\infty F(x))dx \cdot c = L_\infty(\rho_g; F, c)c.$$

Corollary D.5 (CVaR). *For $F \in \mathscr{D}([a,b])$, it holds that*

$$C_\alpha(\mathrm{P}_c^\infty F) - C_\alpha(F) \le \frac{b - F^{-1}(1-\alpha)}{\alpha}c \le \frac{b - F^{-1}(1-\alpha-c)}{\alpha}c = L_\infty(C_\alpha; F, c)c,$$
$$C_\alpha(F) - C_\alpha(\mathrm{N}_c^\infty F) \le \frac{b - F^{-1}(1-\alpha-c)}{\alpha}c = L_\infty(C_\alpha; F, c)c.$$

Proposition D.6 (RDEU). *For $F \in \mathscr{D}([a,b])$, it holds that (assume w convex)*

$$V(\mathrm{P}_c^\infty F) - V(F) \le \int_a^b w'(F(x))u'(x)dx \cdot c \le \int_a^b w'(\mathrm{N}_c^\infty F(x))u'(x)dx \cdot c = L_\infty(V; F, c)c,$$
$$V(F) - V(\mathrm{N}_c^\infty F) \le \int_a^b w'(\mathrm{N}_c^\infty F(x))u'(x)dx \cdot c.$$

D.3.2. WASSERSTEIN DISTANCE

Proposition D.7 (SRM). *For $F \in \mathscr{D}([a,b])$, it holds that*

$$M_\phi(\mathrm{P}_c^1 F) - M_\phi(F) = \int_{g^+(c)}^1 (\mathrm{P}_c^1 F^{-1}(y) - F^{-1}(y))\phi(y)dy \le \phi(1)c = L_1(M_\phi; F, c)c$$
$$M_\phi(F) - M_\phi(\mathrm{N}_c^1 F) = \int_{g^-(c)}^1 (F^{-1}(y) - \mathrm{N}_c^1 F^{-1}(y))\phi(y)dy \le \phi(1)c.$$

Proposition D.8 (DRM). *For $F \in \mathscr{D}([a,b])$, it holds that*

$$\rho_g(\mathrm{P}_c^1 F) - \rho_g(F) = \int_{g^+(c)}^1 (\mathrm{P}_c^1 F^{-1}(y) - F^{-1}(y))g'(1-y)dy \le g'(0)c = L_1(\rho_g; F, c)c$$
$$\rho_g(F) - \rho_g(\mathrm{N}_c^1 F) = \int_{g^-(c)}^1 (F^{-1}(y) - \mathrm{N}_c^1 F^{-1}(y))g'(1-y)dy \le g'(0)c.$$

Corollary D.9 (CVaR). *For $F \in \mathscr{D}([a,b])$, it holds that*

$$C_\alpha(\mathrm{P}_c^1 F) - C_\alpha(F) = \frac{c}{\alpha} = L_1(C_\alpha; F, c)c,$$
$$C_\alpha(F) - C_\alpha(\mathrm{N}_c^1 F) = \frac{c}{\alpha} = L_1(C_\alpha; F, c)c.$$

Proposition D.10 (CE). *For $F \in \mathscr{D}([a,b])$, it holds that*

$$\begin{aligned} C_u(\mathrm{P}_c^1 F) - C_u(F) &\le \|u'\|_\infty (u^{-1})' \left(\int_a^b u(x)dF(x)\right) \cdot c \\ &\le \|u'\|_\infty (u^{-1})' \left(\int_a^b u(x)d\mathrm{N}_c^1 F(x)\right) \cdot c, \\ C_u(F) - C_u(\mathrm{N}_c^1 F) &\le \|u'\|_\infty (u^{-1})' \left(\int_a^b u(x)d\mathrm{N}_c^1 F(x)\right) \cdot c. \end{aligned}$$

Corollary D.11 (ERM). *For $F \in \mathscr{D}([a,b])$, it holds that (assume $\beta > 0$)*

$$U_\beta(\mathrm{P}_c^1 F) - U_\beta(F) \le \frac{\exp(\beta b)}{\int_a^b \exp(\beta x) dF(x)} \cdot c \le \frac{L_\infty(U_\beta)c}{\int_a^b \exp(\beta x) d\mathrm{N}_c^1 F(x)},$$
$$U_\beta(F) - U_\beta(\mathrm{N}_c^1 F) \le \frac{L_\infty(U_\beta)c}{\int_a^b \exp(\beta x) d\mathrm{N}_c^1 F(x)}.$$

E. Proof of Proposition 6.1

Proof. Observe that the regret decomposes as $\mathrm{Regret}(\mathrm{LCB}, \nu, T) = \sum_{i=1}^K \Delta_i \mathbb{E}[N_i(T)]$. We will bound $\mathbb{E}[N_i(T)]$ for each suboptimal arm $i \neq 1$. Recall that $\underline{F}_{i,t} = \mathrm{N}^\infty_{c_i(t)} \hat{F}_{i,t}$. Denote by $\hat{F}_i^s$ the empirical CDF corresponding to arm i after observing s samples, therefore we have $\hat{F}_{i,t} = \hat{F}_i^{N_i(t)}$. Without loss of generality, we assume the first arm is the optimal arm, i.e.,

$$C_\alpha(F_1) \le C_\alpha(F_i),\ \forall i \in [K].$$

Define the good event for arm i as

$$G_i = \{C_\alpha(F_1) > \max_{t\in[T]} C_\alpha(\underline{F}_{1,t})\} \cap \{C_\alpha(\mathrm{N}^\infty_{\sqrt{\frac{\log(2/\delta)}{2u_i}}} \hat{F}_i^{u_i}) > C_\alpha(F_1)\}.$$

We claim that if G_i occurs, then $N_i(T) \le u_i$. The proof follows from a contradiction. Suppose $N_i(T) > u_i$, then there exists some round $t \in [T]$ such that $N_i(t) = u_i$ and $I_t = i$. It follows that

$$\begin{aligned} C_\alpha(\underline{F}_{i,t}) &= C_\alpha(\mathrm{N}^\infty_{c_i(t)} \hat{F}_{i,t}) \\ &= C_\alpha(\mathrm{N}^\infty_{\sqrt{\frac{\log(2/\delta)}{2u_i}}} \hat{F}_i^{u_i}) \\ &> C_\alpha(F_1) \\ &> C_\alpha(\underline{F}_{1,t}), \end{aligned}$$

where the inequalities come from the definition of G_i. Hence $I_t = \arg\min_{i\in[K]} C_\alpha(\underline{F}_{i,t}) \neq i$, which leads to a contradiction. Using the tower property,

$$\mathbb{E}[N_i(T)] = \mathbb{E}[N_i(T)|G_i]\mathbb{P}(G_i) + \mathbb{E}[N_i(T)|G_i^c]\mathbb{P}(G_i^c) \le u_i + T\mathbb{P}(G_i^c).$$

Next, we show that $\mathbb{P}(G_i^c)$ is small. By using union bound, we have

$$\mathbb{P}(G_i^c) \le \mathbb{P}\left(C_\alpha(F_1) \le \max_{t\in[T]} C_\alpha(\tilde{F}_{1,t})\right) + \mathbb{P}\left(C_\alpha\left(\mathrm{N}^\infty_{\sqrt{\frac{\log(2/\delta)}{2u_i}}} \hat{F}_i^{u_i}\right) \le C_\alpha(F_1)\right).$$

By Theorem 4.1, for any $t \in [T]$, if $F_1 \in B(\hat{F}_{1,t}, c_1(t))$ then $\underline{F}_{1,t} = \mathrm{N}^\infty_{c_1(t)} \hat{F}_{1,t} \succeq F_1$, and $C_\alpha(\underline{F}_{1,t}) \le C_\alpha(F_1)$. Hence the first term on the r.h.s. can be bounded as

$$\begin{aligned} \mathbb{P}\left(C_\alpha(F_1) \le \max_{t\in[T]} C_\alpha(\tilde{F}_{1,t})\right) &= \mathbb{P}\left(\exists t \in [T] : C_\alpha(F_1) \le C_\alpha(\underline{F}_{1,t})\right) \\ &\le \mathbb{P}\left(\exists t \in [T] : \left\|F_1 - \hat{F}_{1,t}\right\| \ge \sqrt{\frac{\log(2/\delta)}{2N_i(t)}}\right) \\ &\le \mathbb{P}\left(\cup_{s\in[T]}\left\{\left\|F_1 - \hat{F}_1^s\right\| \ge \sqrt{\frac{\log(2/\delta)}{2s}}\right\}\right) \\ &\le T\delta, \end{aligned}$$

where the last inequality follows from a union bound and the DKW inequality. Denote by $c_i \triangleq \sqrt{\frac{\log(2/\delta)}{2u_i}}$. By Corollary D.5, we have that $C_\alpha(F) - C_\alpha(\mathrm{N}_c^\infty F) \le \frac{b - F^{-1}(1-\alpha-c)}{\alpha} c$. If the event $\left\{\left\|\hat{F}_i^{u_i} - F_i\right\|_\infty < c_i\right\}$ occurs, then

$$C_\alpha(\hat{F}_i^{u_i}) - C_\alpha\left(\mathrm{N}_{c_i}^\infty \hat{F}_i^{u_i}\right) \le \frac{1 - (\hat{F}_i^{u_i})^{-1}(1-\alpha-c_i)}{\alpha} c_i$$

and

$$C_\alpha(F_i) - C_\alpha(\hat{F}_i^{u_i}) \le C_\alpha(\mathrm{P}_{c_i}^\infty \hat{F}_i^{u_i}) - C_\alpha(\hat{F}_i^{u_i}) \le \frac{b - (\mathrm{P}_{c_i}^\infty \hat{F}_i^{u_i})^{-1}(1-\alpha-c_i)}{\alpha} c_i.$$

Combining these inequalities,

$$\begin{aligned} C_\alpha\left(\mathrm{P}_{c_i}\hat{F}_i^{u_i}\right) - C_\alpha(F_i) &\le \left(\frac{b - (\hat{F}_i^{u_i})^{-1}(1-\alpha-c_i)}{\alpha} + \frac{b - (\mathrm{N}_{c_i}\hat{F}_i^{u_i})^{-1}(1-\alpha-c_i)}{\alpha}\right) c_i \\ &\le 2\frac{b - (\hat{F}_i^{u_i})^{-1}(1-\alpha-c_i)}{\alpha} c_i \\ &\le 2\frac{b - F_i^{-1}(1-\alpha-2c_i)}{\alpha} c_i \end{aligned}$$

We choose u_i such that $\Delta_i = C_\alpha(F_i) - C_\alpha(F_1) \ge 2\frac{b-F_i^{-1}(1-\alpha-2c_i)}{\alpha}c_i$, then the second term on the r.h.s. can be bounded as

$$\begin{aligned} \mathbb{P}\left(C_\alpha\left(\mathrm{N}_{c_i}^\infty \hat{F}_i^{u_i}\right) \le C_\alpha(F_1)\right) &= \mathbb{P}\left(C_\alpha(F_i) - C_\alpha\left(\mathrm{N}_{c_i}^\infty \hat{F}_i^{u_i}\right) \ge \Delta_i\right) \\ &\le \mathbb{P}\left(C_\alpha(F_i) - C_\alpha\left(\mathrm{N}_{c_i}^\infty \hat{F}_i^{u_i}\right) \ge 2\frac{b - F_i^{-1}(1-\alpha-2c_i)}{\alpha} c_i\right) \\ &\le \mathbb{P}\left(\left\|\hat{F}_i^{u_i} - F_i\right\|_\infty \ge c_i\right) \\ &\le \delta. \end{aligned}$$

Hence, the probability of G_i^c is bounded as $\mathbb{P}(G_i^c) \le (T+1)\delta$. It follows that

$$\mathbb{E}[N_i(T)] \le u_i + T(T+1)\delta.$$

Define $h_i(c) \triangleq 2\frac{b-F_i^{-1}(1-\alpha-2c)}{\alpha}c$. Let $c_i^* \triangleq h_i^{-1}(\Delta_i)$ be the solution to the equation

$$h_i(c) = 2\frac{b - F_i^{-1}(1-\alpha-2c)}{\alpha} c = \Delta_i.$$

Note that c_i^* is a **distribution-dependent constant**. We let $u_i := \left\lceil \frac{\log(2/\delta)}{2(c_i^*)^2} \right\rceil$ and let $\delta = \frac{1}{T^2}$:

$$\mathbb{E}[N_i(T)] \le \left\lceil \frac{\log(2T^2)}{2(c_i^*)^2} \right\rceil + 2 \le \frac{\log(\sqrt{2}T)}{(c_i^*)^2} + 3.$$

Substituting it into the regret decomposition, we get

$$\begin{aligned} \sum_{i=1}^K \Delta_i \mathbb{E}[N_i(T)] &\le \log(\sqrt{2}T) \sum_{i=2}^K \frac{\Delta_i}{(c_i^*)^2} + 3\sum_{i=1}^K \Delta_i \\ &= \frac{4\log(\sqrt{2}T)}{\alpha^2} \sum_{i>1}^K \frac{\left(b - F_i^{-1}(1-\alpha-2c_i^*)\right)^2}{\Delta_i} + 3\sum_{i=1}^K \Delta_i. \end{aligned}$$

□

F. Algorithms

We present several comprehensive algorithms that output the confidence bounds for a given risk measure when given n i.i.d. samples and confidence radius as input. Algorithm 2 and Algorithm 3 compute the UCB and LCB of a risk measure via Wasserstein distance respectively. Meanwhile, Algorithm 4 and Algorithm 5 compute the UCB and LCB of a risk measure via supremum distance respectively.

Algorithm 2 `Wasserstein upper confidence bound`

Input: b, samples $\boldsymbol{X} = X_1, X_2, \cdots, X_n$, risk measure $\mathbf{T}$, $c > 0$
Sort the n samples in ascent order $X_{(1)} \leq X_{(2)} \cdots \leq X_{(n)}$
Initialize $S_1 = \frac{b - X_{(n)}}{n}$
for $i = 1 : n$ **do**
  if $S_i \leq c$ **then**
    $S_{i+1} = S_i + \frac{1}{n}(b - X_{(n-i)})$
  else
    $n' = n + 1 - i$
    break
  end if
end for
$p_{n'} = \frac{S_i - c}{b - X_{(n')}}$, $p_b = \frac{i}{n} - p_{n'}$
$\overline{F_n^1} = \frac{1}{n}\sum_i^{n'-1} \mathbb{I}\{X_{(i)} \leq \cdot\} + p_{n'}\mathbb{I}\{X_{(n')} \leq \cdot\} + p_b\mathbb{I}\{b \leq \cdot\}$
Output: $\mathbf{T}\left(\overline{F_n^1}\right)$

Algorithm 3 `Wasserstein lower confidence bound`

Input: a, samples $\boldsymbol{X} = X_1, X_2, \cdots, X_n$, risk measure $\mathbf{T}$, $c > 0$
Sort the n samples in ascent order $X_{(1)} \leq X_{(2)} \cdots \leq X_{(n)}$
Initialize $S_1 = \frac{X_{(n)} - X_{(n-1)}}{n}$
for $i = 1 : n - 1$ **do**
  if $S_i \leq c$ **then**
    $S_{i+1} = S_i + \frac{i+1}{n}(X_{(n-i)} - X_{(n-1-i)})$
  else
    $n' = n + 1 - i$
    break
  end if
end for
$b^- = X_{(n'-1)} + \frac{n(S_i - c)}{i}$
$\underline{F_n^1} = \frac{1}{n}\sum_i^{n'-1} \mathbb{I}\{X_i \leq \cdot\} + \frac{i}{n}\mathbb{I}\{b^- \leq \cdot\}$
Output: $\mathbf{T}\left(\underline{F_n^1}\right)$

F.1. Time Complexity

We start with Algorithm 2. The sorting of n samples incurs $\mathcal{O}(n \log n)$. The for-loop costs $\mathcal{O}(n)$ since the cost in each iteration is $\mathcal{O}(1)$. Therefore the total time complexity is $\mathcal{O}(n \log n + \log n)$. The time complexity of Algorithm 3-Algorithm 5 is $\mathcal{O}(n \log n + \log n)$.

F.2. Space Complexity

Consider Algorithm 2. The space complexity of storing the samples is $\mathcal{O}(n)$. In addition, storing S_i, $p_{n'}$, and p_b costs $\mathcal{O}(n)$. The total space complexity is $\mathcal{O}(n)$. It is easy to check that the time complexity of Algorithm 3-Algorithm 5 is $\mathcal{O}(n)$.

G. Experiments

G.1. Confidence Bounds

We consider five beta distributions with different parameters. The specific parameters (A, B) is shown above in each figure. Unless otherwise specified, we always use $N = 10^5$ samples, $\alpha = 0.05$, $\beta = 1$, and $\delta = 0.05$. For convenience, we use $c_n^1 = (b - a)c_n^\infty$. We plot the CIs for ERM and CVaR for varying sample size and varying risk parameter in Figure 6-13.

Algorithm 4 `Supremum upper confidence bound`

Input: b, samples $\boldsymbol{X} = X_1, X_2, \cdots, X_n$, risk measure $\mathbf{T}$, $c > 0$
Sort the n samples in ascent order $X_{(1)} \leq X_{(2)} \cdots \leq X_{(n)}$
Initialize $i = 1$
while $\frac{i}{n} \leq c$ **do**
  $i = i + 1$
  $l = i$
end while
$\overline{F_n^\infty} = (\frac{l}{n} - c)\mathbb{I}\{X_{(l)} \leq \cdot\} + \frac{1}{n}\sum_{i=l+1}^{n} \mathbb{I}\{X_{(i)} \leq \cdot\} + c\mathbb{I}\{b \leq \cdot\}$
Output: $\mathbf{T}\left(\overline{F_n^\infty}\right)$

Algorithm 5 `Supremum lower confidence bound`

Input: a, samples $\boldsymbol{X} = X_1, X_2, \cdots, X_n$, risk measure $\mathbf{T}$, $c > 0$
Sort the n samples in ascent order $X_{(1)} \leq X_{(2)} \cdots \leq X_{(n)}$
Initialize $i = n$
while $\frac{i}{n} + c \geq 1$ **do**
  $i = i - 1$
  $l = i$
end while
$\underline{F_n^\infty} = c\mathbb{I}\{a \leq \cdot\} + \frac{1}{n}\sum_{i=1}^{l} \mathbb{I}\{X_{(i)} \leq \cdot\} + (1 - \frac{l}{n} - c)\mathbb{I}\{X_{(l+1)} \leq \cdot\}$
Output: $\mathbf{T}\left(\underline{F_n^\infty}\right)$

G.2. CVaR Bandit

We adopt the same bandit instances as Tamkin et al. (2019). The parameters of these distributions are given in Table 1 in Tamkin et al. (2019). The left, middle, and right part of Figure 14 plots the results for easy bandit instance with $\alpha = 0.25$, hard bandit instance with $\alpha = 0.25$, and hard bandit instance with $\alpha = 0.05$. As expected, `CVaR-UCB` consistently outperforms `LLC-UCB` and `GLC-UCB`.

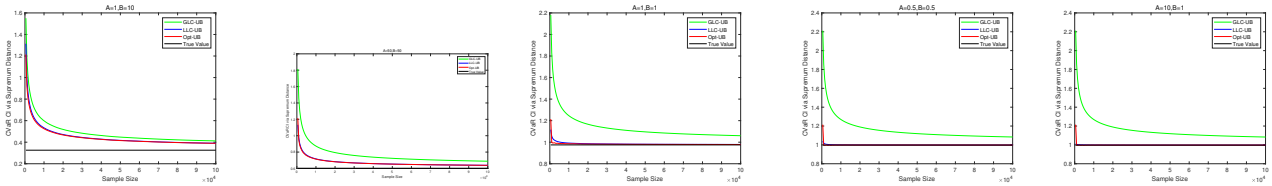


Figure 6: CVaR UCB with varying sample size

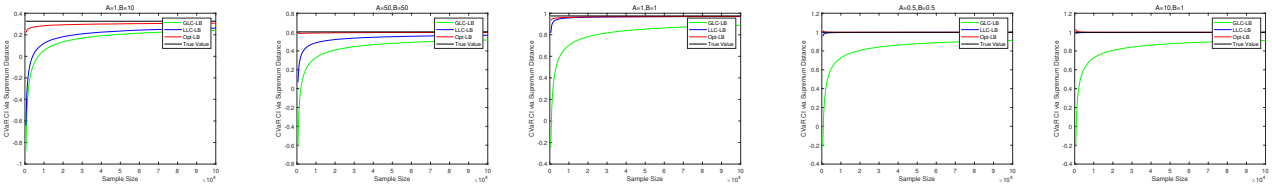


Figure 7: CVaR LCB with varying sample size

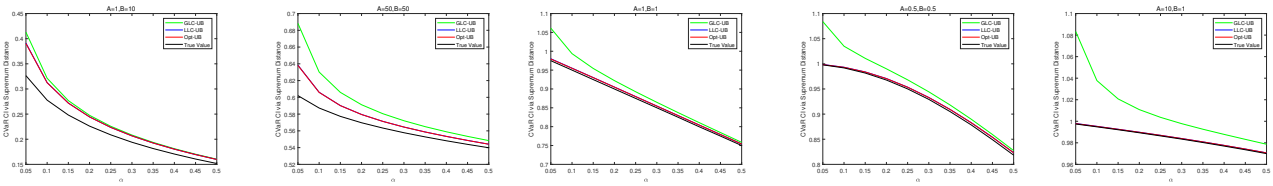


Figure 8: CVaR UCB with varying α

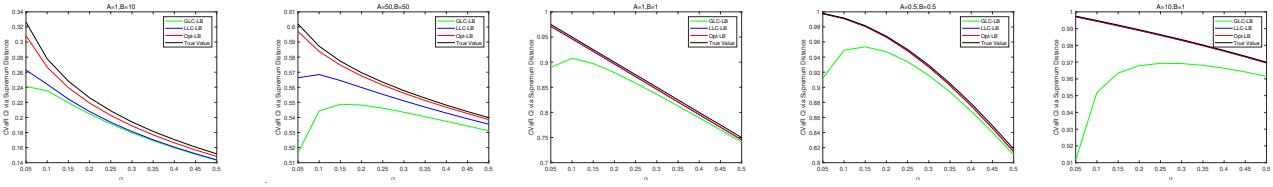


Figure 9: CVaR LCB with varying α

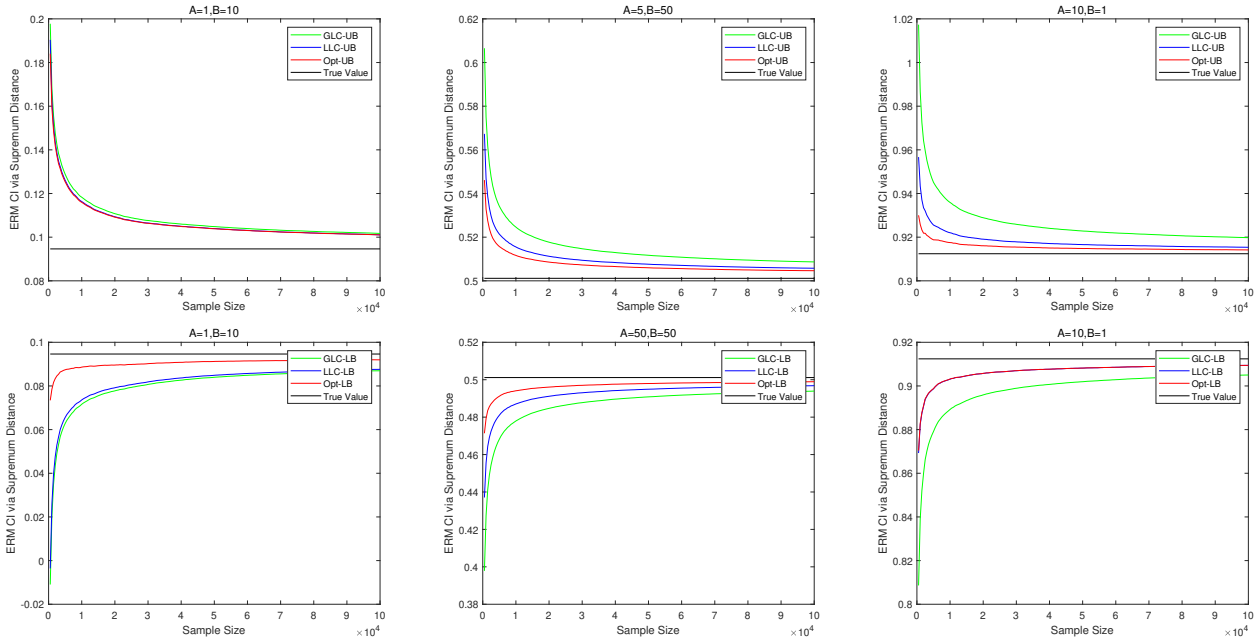


Figure 10: ERM CI with varying sample size

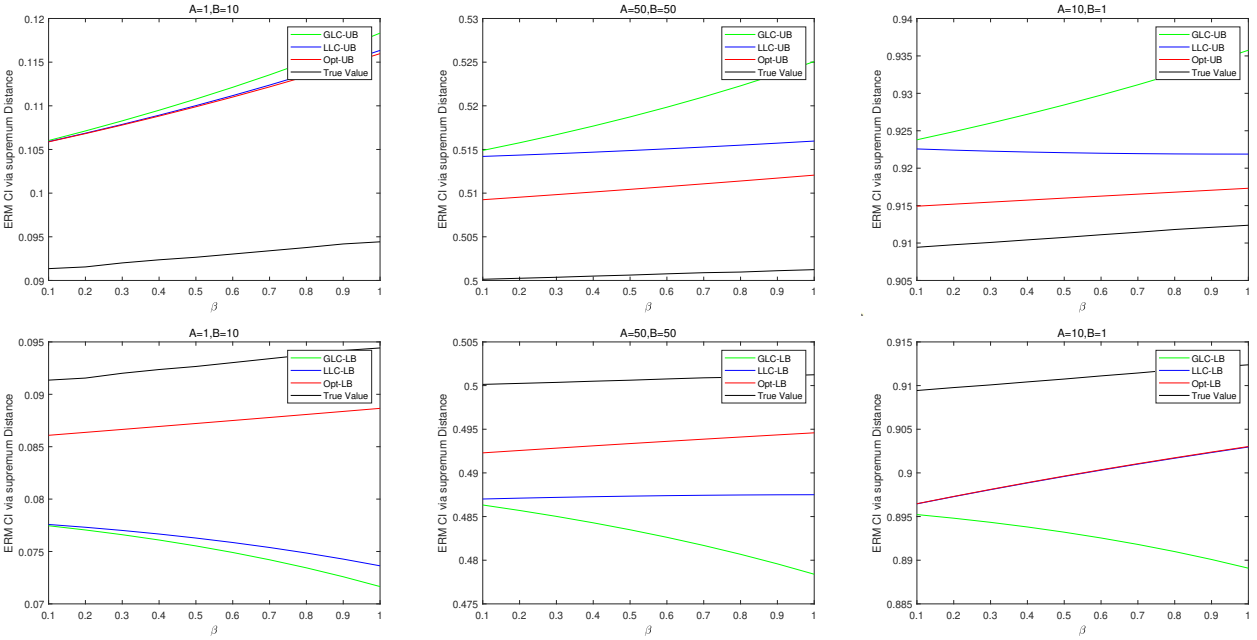


Figure 11: ERM CI with varying β

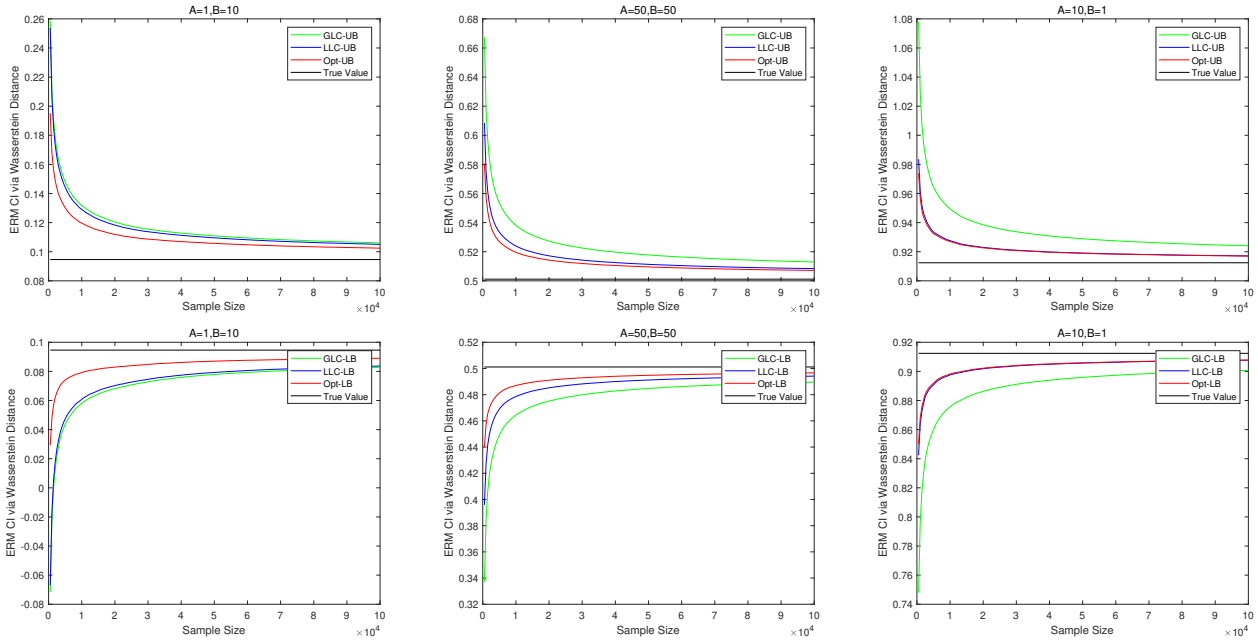


Figure 12: ERM CI with varying sample size

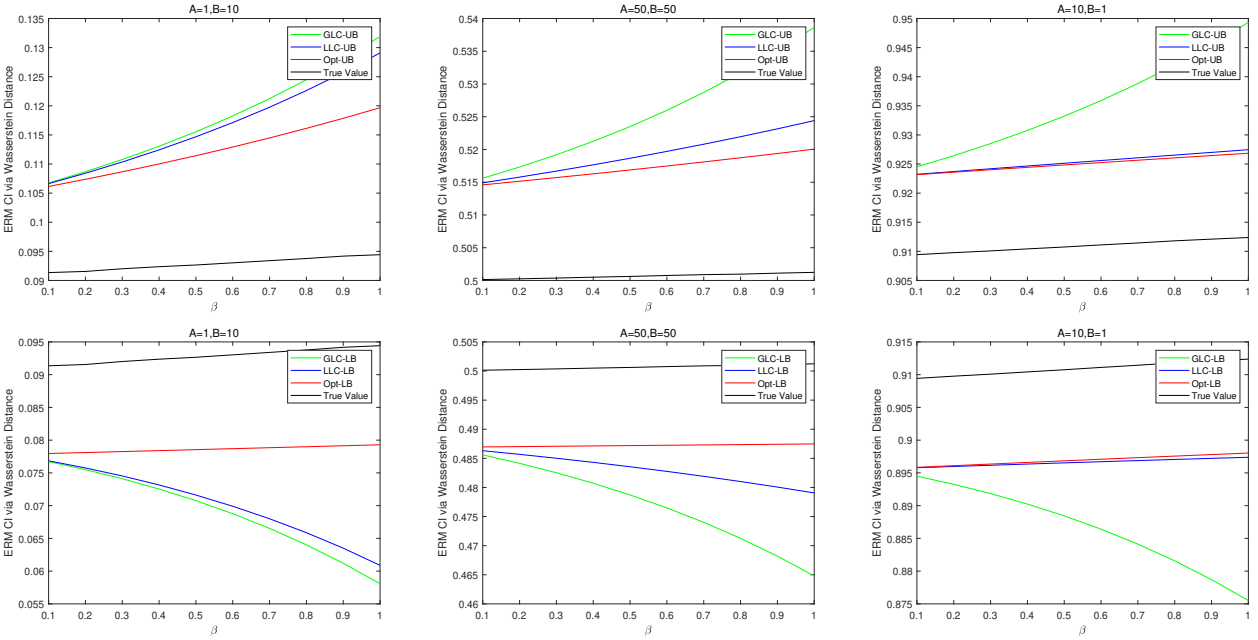


Figure 13: ERM CI with varying β

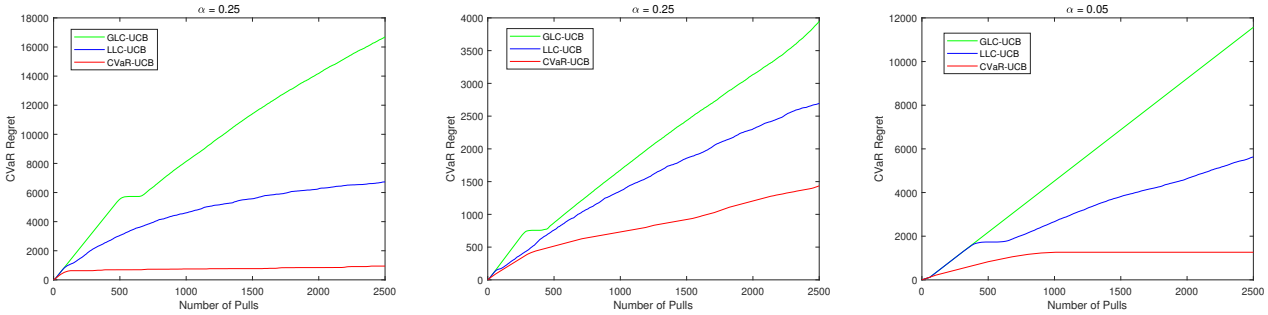


Figure 14: CVaR bandit