Journal of Machine Learning Research 23 (2022) 1-61 Submitted 9/20; Revised 5/22; Published 8/22

A Wasserstein Distance Approach for Concentration of Empirical Risk Estimates*

Prashanth L.A. PRASHLA@CSE.IITM.AC.IN
Department of Computer Science and Engineering
Indian Institute of Technology Madras, Chennai 600036, India

Sanjay P. Bhat SANJAY.BHAT@TCS.COM
TCS Research
Hyderabad 500081, India

Editor: Ambuj Tewari

Abstract

This paper presents a unified approach based on Wasserstein distance to derive concentration bounds for empirical estimates for two broad classes of risk measures defined in the paper. The classes of risk measures introduced include as special cases well known risk measures from the finance literature such as conditional value at risk (CVaR), optimized certainty equivalent risk, spectral risk measures, utility-based shortfall risk, cumulative prospect theory (CPT) value, rank dependent expected utility and distorted risk measures. Two estimation schemes are considered, one for each class of risk measures. One estimation scheme involves applying the risk measure to the empirical distribution function formed from a collection of i.i.d. samples of the random variable (r.v.), while the second scheme involves applying the same procedure to a truncated sample. The bounds provided apply to three popular classes of distributions, namely sub-Gaussian, sub-exponential and heavy-tailed distributions. The bounds are derived by first relating the estimation error to the Wasserstein distance between the true and empirical distributions, and then using recent concentration bounds for the latter. Previous concentration bounds are available only for specific risk measures such as CVaR and CPT-value. The bounds derived in this paper are shown to either match or improve upon previous bounds in cases where they are available. The usefulness of the bounds is illustrated through an algorithm and the corresponding regret bound for a stochastic bandit problem involving a general risk measure from each of the two classes introduced in the paper.

Keywords: Concentration bounds, Wasserstein distance, conditional value-at-risk, optimized certainty equivalent risk, spectral risk measure, utility-based shortfall risk, cumulative prospect theory, rank-dependent expected utility, distorted risk measures, risk-sensitive bandits.

1. Introduction

Concentration of sample averages has received a lot of attention in statistics. Sample averages are usually used for estimating the expectation of a random variable (r.v.), and classic inequalities, such as those of Hoeffding and Bernstein, provide the necessary concentration bounds. However, the expected value has several shortcomings in the context of risk-sensitive optimization (Allais,

∗. A preliminary version of this manuscript was published in *NeurIPS 2019* - see (Bhat and Prashanth, 2019). In comparison to the conference version, this paper includes new theoretical results, all the convergence proofs, and a revised presentation.

1953; Ellsberg, 1961; Kahneman and Tversky, 1979; Rockafellar and Uryasev, 2000), and several measures have been proposed in the literature to capture the notion of risk in a practical application. Unlike the case of expected value, concentration bounds are either not available, or not optimal, for the estimation of several risk measures.

In this paper, we consider the estimation of the following popular risk measures: optimized certainty equivalent (OCE) risk (Ben-Tal and Teboulle, 1986) which includes Conditional Value-at-Risk (CVaR) (Rockafellar and Uryasev, 2000) as a special case, spectral risk measure (SRM) (Acerbi, 2002), utility-based shortfall risk (UBSR) (Föllmer and Schied, 2002), cumulative prospect theory (CPT) (Tversky and Kahneman, 1992), rank-dependent expected utility (RDEU) (Quiggin, 2012) and distorted risk measures (DRM) (Denneberg, 1990). CVaR is popular in financial applications, where it is necessary to minimize the worst-case losses, say in a portfolio optimization context. CVaR is a special instance of the class of OCE risk measures (Lee et al., 2020) as well as SRMs (Acerbi, 2002). CVaR is an appealing risk measure because it is coherent (Artzner et al., 1999), and spectral risk measures retain this property. UBSR belongs to the family of convex risk measures (Föllmer and Schied, 2002), which generalizes the class of coherent risk measures. CPT value is a risk measure that is useful for modeling human preferences. RDEU is a closely-related measure, since it shares with CPT the idea of employing a weight function to distort underlying probabilities.

We provide a novel categorization of risk measures based on their continuity properties in the space of distributions equipped with the Wasserstein metric. To elaborate, we define two broad types of risk measures, say (T1) and (T2), and show that, under suitable conditions, (i) CVaR, SRM and UBSR are (T1); and (ii) CPT, RDEU and DRM are (T2). Such a categorization provides a uniform framework for estimating the aforementioned risk measures as well as deriving concentration bounds for the estimates, which is the primary focus of this work.

The need for concentration bounds for estimation of risk measures is motivated by the fact that, while information about the underlying distribution is typically unavailable in practical applications, one can often obtain i.i.d. samples from the distribution. Our aim is to estimate the chosen risk measure using these samples, and derive concentration bounds for these estimates. We consider this problem of estimation in the broader context of (T1) and (T2) risk measures. In the former case, we examine the estimator obtained by applying the risk measure to the empirical distribution constructed from an i.i.d. sample. On the other hand, in the case of a (T2) risk measure, we consider an estimator that is obtained using the truncated empirical distribution. Our estimator for (T1) risk measure, when specialized to the case of CVaR, coincides with the one that is already available in the literature. On the other hand, the general (T1) estimator applied to SRM and UBSR leads to novel estimators. Next, in the case of CPT, the estimator in literature does not involve truncation, while our (T2) estimator does. For the case of RDEU, our estimator is novel, to the best of our knowledge.

We derive concentration bounds for estimators of (T1) and (T2) risk measures for three general classes of distributions, namely sub-Gaussian, sub-exponential and heavy-tailed distributions that satisfy a higher-moment bound. We achieve this in a novel manner by relating the estimation error to the Wasserstein distance between the empirical and true distributions, and then using concentration bounds for the latter. We specialize the bounds obtained for the case of a (T1) risk measure to provide concentration results for empirical versions of OCE (with CVaR as a special case), SRM and UBSR. We perform a similar exercise to obtain concentration results for CPT and RDEU, using the general result for (T2) risk measures.

We now summarize our results when the underlying distribution is sub-Gaussian, which is a popular class of distributions with possibly unbounded support.

(1) For the case of CVaR, we provide a tail bound of the order $O\left(\exp\left(-cn\epsilon^2\right)\right)$, where n is the number of samples, ϵ is the accuracy parameter, and c is a universal constant. Our bound matches the rate obtained for distributions with bounded support in Brown (2007), and features improved dependence on ϵ as compared to the one derived for sub-Gaussian distributions in Kolla et al. (2019). Further, unlike our results, the latter work imposes a minimum growth assumption on the underlying distribution.

(2) Tail bounds of the order $O\left(\exp\left(-cn\epsilon^2\right)\right)$ are shown to hold for any SRM having a bounded risk spectrum, and any UBSR with a Lipschitz utility function. Unlike CVaR, the estimators as well as the concentration results for SRM and UBSR are novel. A trapezoidal rule-based estimator has been proposed for SRM in Pandey et al. (2021), and the bounds there are under more stringent assumptions as compared to the one we derive for SRM.

(3) For the case of CPT-value, we obtain an order $O\left(\exp\left(-cn\epsilon^2\right)\right)$ bound for the case of distributions with bounded support, matching the rate in Cheng et al. (2018). For the case of sub-Gaussian distributions, we provide a bound that has an improved dependence on the number of samples n, as compared to the corresponding bound derived by Cheng et al. (2018). Similar bounds are shown to hold for any RDEU involving a utility function with derivative bounded above and bounded away from 0.

(4) The results outlined above for various risk measures in the sub-Gaussian case rely on a concentration bound on the Wasserstein distance between empirical and true distributions, which we derive. This bound, with explicit constants, may be of independent interest.

(5) As a minor contribution, our concentration bounds open avenues for bandit applications, and we illustrate this claim by considering a risk-sensitive bandit setting, with any (T1)/(T2) risk measure governing the objective. We use the concentration bounds for (T1)/(T2) risk measures to derive regret bounds for such a bandit problem in two cases: first, when the underlying arms' distribution is assumed to be sub-Gaussian and the risk measure is (T1), and second, when the arms' distribution has bounded support and the risk measure is (T2). Previous works (cf. Galichet (2015); Gopalan et al. (2017)) consider CVaR and CPT optimization in a bandit context, with arms' distributions having bounded support. In contrast, we consider unbounded, albeit sub-Gaussian distributions in the broader (T1) class of risk measures.

In addition, we also derive concentration bounds for the five risk measures mentioned above, for the case when the underlying distribution is either sub-exponential or heavy-tailed, but satisfying a higher moment bound. To the best of our knowledge, barring CVaR and SRM, tail bounds are not available for the aforementioned classes of distributions for the other risk measures.

Table 1 provides a risk-measure-wise summary of the bounds presented in this paper for various classes of distributions along with the relevant previous works in parentheses. After characterizing the two types of risk measures in Section 3, and establishing the type for several popular risk measures, we provide a map of our results in Section 4 for various risk measures under different assumptions on the underlying distribution through Tables 2–5. In particular, these tables list bounds for risk estimation that hold in expectation (under a bounded higher moment condition), and concentration bounds for sub-Gaussian and sub-exponential distributions, respectively.

Table 1: Summary of the bounds on $\mathbb{P}\left(|\rho_n - \rho(X)| > \epsilon\right)$, for various choices of risk measure $\rho(X)$. Here X is either a bounded, or sub-Gaussian, or sub-exponential r.v., and ρ_n is an estimate of $\rho(X)$ using n i.i.d. samples. The corollaries accompanied by a $*$ symbol feature improved bounds in comparison to those in the literature.

Risk measure	**Bounded support**	**Sub-Gaussian**	**Sub-exponential**
Conditional Value-at-Risk	(Wang and Gao, 2010)	(Prashanth et al., 2020) Corollaries 28 and 29	(Prashanth et al., 2020) Corollary 42
Spectral risk measure	(Pandey et al., 2021)	(Pandey et al., 2021) Corollaries 30 and 31 $*$	(Pandey et al., 2021) Corollary 43 $*$
Utility-based shortfall risk	Corollary 33	Corollaries 32 and 33	Corollary 44
Cumulative prospect theory	(Cheng et al., 2018) Corollary 36	(Cheng et al., 2018) Corollaries 39 and 40 $*$	Corollary 46
Rank-dependent expected utility	Bounds for CPT apply in light of Lemma 18		

Related work. Concentration of empirical CVaR has been the topic of many recent works, cf. Brown (2007); Wang and Gao (2010); Thomas and Learned-Miller (2019); Kolla et al. (2019); Kagrecha et al. (2019); Prashanth et al. (2020). The first three references address the case when the underlying distribution has bounded support, while the rest consider unbounded distributions which are either sub-Gaussian or sub-exponential or have a bounded higher-moment. SRM estimation has been considered in Pandey et al. (2021), where the authors provide tail bounds for the case of distributions that have either bounded support, or are sub-Gaussian/sub-exponential. In comparison to these works, the bounds that we derive are under less stringent assumptions. For instance, in Prashanth et al. (2020), the authors require a minimum growth condition on the underlying distribution, while in Pandey et al. (2021), the authors require that the underlying density be bounded above zero, or follow the Gaussian/exponential form. While estimation of UBSR has been considered in Dunkel and Weber (2010) and Hu and Zhang (2018), concentration bounds for UBSR has not been addressed in previous works, to the best of our knowledge. In the first of these two references, the authors present a stochastic approximation based method for UBSR estimation, while in the second reference, the authors propose a sample-average approximation for UBSR. In both works, there are no non-asymptotic bounds. Instead, an asymptotic normality result is provided. Finally, CPT estimation is the topic of Prashanth et al. (2016) and Cheng et al. (2018). Our estimate for (T2) measures resembles the one in the aforementioned references, except that we employ truncation, which aids in arriving at better concentration results for the sub-Gaussian case. Concentration bounds for empirical versions of RDEU are not available in the literature, to the best of our knowledge, and we fill this gap.

Since CVaR and spectral risk measures are weighted averages of the underlying distribution quantiles, a natural alternative to a Wasserstein-distance-based approach is to employ concentration results for quantiles such as in Kolla et al. (2019). While such an approach can provide bounds with better constants, the resulting bounds also involve distribution-dependent quantities (see Prashanth et al. (2020), for instance), and require different proofs for sub-Gaussian and sub-exponential r.v.s. In contrast, our approach provides a unified method of proof.

In Cassel et al. (2018), the authors consider a general risk measure along with an abstract norm on the space of distributions such that the risk measure satisfies a polynomial growth bound with respect to the norm, and the norm difference between the empirical and true distributions satisfies a given concentration inequality. There are thus parallels between how continuity properties of the risk measure and concentration inequalities for empirical distributions are combined in Cassel et al. (2018) and in this paper. Our definition of (T1) risk measures is also motivated by the stability criteria in the aforementioned reference. However, there are some key differences. First, instead of an abstract norm, we work with the Wasserstein metric, and impose a Hölder-continuity requirement on the risk measure in the space of distributions. Second, the concentration bounds we derive easily specialize to SRM and UBSR — two risk measures not considered there. Third, CPT and RDEU risk measures do not satisfy the stability criteria of Cassel et al. (2018). We address this gap by providing the (T2) class of risk measures, which CPT/RDEU belong to. Fourth, the requirement 2 in Definition 3 of Cassel et al. (2018), which specifies the stability criteria, relates to the concentration of empirical distribution, and is usually satisfied for sub-Gaussian distributions. We study concentration of empirical risk estimates for sub-Gaussian distributions as well as for the more general classes of sub-exponential and heavy-tailed distributions. Note that the aforementioned requirement in Cassel et al. (2018) does not hold for sub-exponential and higher-moment bounded distributions under standard metric such as the sup-norm and the Wasserstein distance.

The rest of the paper is organized as follows: In Section 2, we cover background material on Wasserstein distance, and state concentration bounds on Wasserstein distance between empirical and true distribution functions under different assumptions on the tail of the underlying distribution. In Section 3, we define two types of risk measures, and establish the type for five popular risk measures. In Section 4, we provide a map of the results in Sections 5–8. In Section 5, we provide bounds in expectation for risk estimation. In Section 6, we present concentration bounds for the two general types of risk measures for the case when the underlying distribution is sub-Gaussian. In Sections 7 and 8, we provide concentration bounds for the cases of sub-exponential and heavy-tailed distributions, respectively. In Section 9, we discuss bandit applications. Finally, in Section 12, we provide the concluding remarks.

2. Wasserstein Distance

In this section, we introduce the notion of Wasserstein distance, a popular metric for measuring the proximity between two distributions. The reader is referred to Chapter 6 of Villani (2008) for a detailed introduction.

Given two cumulative distribution functions (CDFs) F_1 and F_2 on $\mathbb{R}$, let $\Gamma(F_1, F_2)$ denote the set of all joint distributions on $\mathbb{R}^2$ having F_1 and F_2 as marginals.

Definition 1 *Given two CDFs F_1 and F_2 on $\mathbb{R}$ and $p \geq 1$, the p-Wasserstein distance between them is defined by*

$$W_p(F_1, F_2) \triangleq \inf_{F \in \Gamma(F_1, F_2)} \left[\int_{\mathbb{R}^2} |x - y|^p dF(x, y) \right]^{\frac{1}{p}}.$$

In this paper, we will mostly restrict ourselves to the case $p = 1$. For convenience, we shall refer to the 1-Wasserstein distance as 'the Wasserstein distance'.

Given $L > 0$ and $\kappa > 0$, a function $f : \mathbb{R} \to \mathbb{R}$ is L-Hölder of order κ if $|f(x) - f(y)| \leq L|x-y|^\kappa$ for all $x, y \in \mathbb{R}$. The function $f : \mathbb{R} \to \mathbb{R}$ is L-Lipschitz if it is L-Hölder of order 1. Finally, if F is a CDF on $\mathbb{R}$, we define the generalized inverse $F^{-1} : [0,1] \to \mathbb{R}$ of F by $F^{-1}(\beta) = \inf\{x \in \mathbb{R} : F(x) \geq \beta\}$. In the case where F is strictly increasing and continuous, F^{-1} equals the usual inverse of a bijective function.

The following lemma provides alternative characterizations of the Wasserstein distance between two CDFs.

Lemma 2 *Suppose X and Y are r.v.s having CDFs F_1 and F_2, respectively. Then,*

$$\sup |\mathbb{E}(f(X) - \mathbb{E}(f(Y))| = W_1(F_1, F_2) = \int_{-\infty}^{\infty} |F_1(s) - F_2(s)| \mathrm{d}s = \int_0^1 |F_1^{-1}(\beta) - F_2^{-1}(\beta)| \mathrm{d}\beta, \tag{1}$$

where the supremum is over all functions $f : \mathbb{R} \to \mathbb{R}$ that are 1-Lipschitz.

Proof See Section 10.1.1. ■

For deriving concentration bounds for empirical versions of risk measures, we shall use a bound on the Wasserstein distance between the empirical distribution function (EDF) and the underlying CDF. We first define the EDF of a r.v. X before stating the relevant Wasserstein distance bounds. Given i.i.d. samples $X_1, \ldots, X_n$ from the distribution F of a r.v. X, the EDF F_n is defined by

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\{X_i \leq x\}, \quad \text{for all } x \in \mathbb{R}. \tag{2}$$

In the above, $\mathbb{I}\{\cdot\}$ denotes the indicator function, i.e., for an event A, $\mathbb{I}\{A\} = 1$ if A happens, and $\mathbb{I}\{A\} = 0$ otherwise.

Concentration bounds on the Wasserstein distance between the EDF of an i.i.d. sample and the underlying CDF from which the sample is drawn, have been derived in Lei (2020); Weed and Bach (2019); Bolley et al. (2007); Boissard (2011); Fournier and Guillin (2015). In the following sections, we shall state Wasserstein distance bounds for two different classes of unbounded r.v.s.

2.1 Distributions satisfying an exponential moment bound

In this section, we state a Wasserstein distance bound assuming that the underlying distribution satisfies an exponential moment bound with an exponent greater than one. Sub-Gaussian distributions are a popular class of distributions that satisfy this assumption.

(C1) *There exist $\beta > 1$, $\gamma > 0$ and $\top > 0$ such that $\mathbb{E}\left(\exp\left(\gamma |X|^\beta\right)\right) < \top < \infty$.*

For the special case of distributions satisfying (C1), we extract the required Wasserstein concentration bound from Fournier and Guillin (2015). This result will be used to derive concentration bounds for empirical versions of various risk measures in Section 6.

Lemma 3 *Let X be a r.v. with CDF F, and suppose X satisfies (C1). Then, for every $\epsilon \geq 0$ and $n \geq 1$, we have*

$$\mathbb{P}\left(W_1(F_n, F) > \epsilon\right) \leq c_1 \left(\exp\left(-c_2 n \epsilon^2\right) \mathbb{I}\{\epsilon \leq 1\} + \exp\left(-c_3 n \epsilon^\beta\right) \mathbb{I}\{\epsilon > 1\}\right),$$

where c_1, c_2, and c_3 are constants that depend on the parameters β, γ and $\top$ specified in (C1).

Proof See Section 10.1.2. ∎

We next define a sub-Gaussian r.v., which is a popular sub-class of unbounded r.v.s satisfying assumption (C1).

Definition 4 *A r.v. X is sub-Gaussian if there exists $\sigma > 0$ such that*

$$P(X \geq \epsilon) \leq \exp\left(-\frac{\epsilon^2}{2\sigma^2}\right), \text{ and } P(X \leq -\epsilon) \leq \exp\left(-\frac{\epsilon^2}{2\sigma^2}\right), \tag{3}$$

for every $\epsilon > 0$.

A sub-Gaussian r.v. X satisfies

$$\mathbb{E}\left(\exp\left(\gamma X^2\right)\right) \leq 2,$$

where γ is a universal constant multiple of the sub-Gaussianity parameter σ. Thus, a sub-Gaussian random variable satisfies (C1) with $\beta = \top = 2$.

Remark 5 *Note that our definition of sub-Gaussianity above does not require the r.v. to have mean zero. If the r.v. X has mean zero, then X is sub-Gaussian if and only if there exists $\sigma > 0$ such that*

$$\mathbb{E}\left(\exp\left(\lambda X\right)\right) \leq \exp\left(\frac{\lambda^2\sigma^2}{2}\right) \text{ for every } \lambda \in \mathbb{R}.$$

The reader is referred to Section 2.5 of Vershynin (2018) for a detailed introduction to sub-Gaussian distributions.

Remark 6 *While sub-exponential distributions also satisfy an exponential moment bound, they do not satisfy (C1), as the exponent β in the sub-exponential case equals one. While Theorem 2 in Fournier and Guillin (2015) covers the case $\beta = 1$, the tail bound there is weak as it exhibits a polynomial decay for large deviations. In the next section, we handle the sub-exponential case separately by using a Wasserstein distance bound from the recent work of Lei (2020).*

The following result for sub-Gaussian distributions is an immediate corollary of Lemma 3.

Corollary 7 *Let X be a r.v. with CDF F. Suppose that X is sub-Gaussian with parameter σ. Then, for every $\epsilon \geq 0$ and $n \geq 1$, we have*

$$\mathbb{P}\left(W_1(F_n, F) > \epsilon\right) \leq c_1 \exp\left(-c_2 n \epsilon^2\right),$$

where c_1, and c_2 are constants that depend on the sub-Gaussianity parameter σ.

The bound in Corollary 7 does not make the constants explicit, since an expression that makes the dependence of the constants c_1, c_2 on the underlying parameters explicit is not available in Fournier and Guillin (2015). From the proof of Theorem 2 there, we were unable to extract the explicit dependence of the constants on the underlying parameters. The lack of explicit constants may be of concern in bandit applications. To address this shortcoming, we provide Wasserstein distance bounds with explicit constants for sub-Gaussian r.v.s in the result below. In the next section, we provide such bounds for sub-exponential r.v.s.

Lemma 8 ***(Wasserstein distance bound)*** *Let X be a sub-Gaussian r.v. with parameter σ. Let F denote the CDF of X. Then, for every $n \geq 1$ and ϵ such that $\frac{512\sigma}{\sqrt{n}} < \epsilon < \frac{512\sigma}{\sqrt{n}} + 16\sigma\sqrt{\mathsf{e}}$, we have*

$$\mathbb{P}\left(W_1(F_n, F) > \epsilon\right) \leq \exp\left(-\frac{n}{256\sigma^2\mathsf{e}}\left(\epsilon - \frac{512\sigma}{\sqrt{n}}\right)^2\right),$$

where e is Euler's number.

Proof See Section 10.1.3. ■

Notice that, unlike Lemma 3, the constants are made explicit in the bound above.

Remark 9 *The bound in the lemma above has a term of the form $\left(\epsilon - \frac{512\sigma}{\sqrt{n}}\right)^2$ inside the exponential function because we first derive a tail bound on the centered error $W_1(F_n, F) - \mathbb{E}[W_1(F_n, F)]$, and then use $\mathbb{E}[W_1(F_n, F)] \leq \frac{512\sigma}{\sqrt{n}}$. The tail bound uses arguments similar to those employed in establishing the well-known McDiarmid inequality (cf. Theorem 5.1 of Lei (2020)), and is of the form*

$$\mathbb{P}\left(W_1(F_n, F) - \mathbb{E}[W_1(F_n, F)] > \tilde{\epsilon}\right) \leq \exp\left(-\frac{n\tilde{\epsilon}^2}{128\sigma^2\mathsf{e}}\right), \textit{ for any } \tilde{\epsilon} \in (0, 16\sigma\sqrt{\mathsf{e}}).$$

The main bound in Lemma 8 follows by setting $\tilde{\epsilon} = \left(\epsilon - \frac{512\sigma}{\sqrt{n}}\right)$ in the above inequality, and using $\mathbb{E}[W_1(F_n, F)] \leq \frac{512\sigma}{\sqrt{n}}$. The reader is referred to Section 10.1.3 for the detailed proof.

2.2 Sub-exponential distributions

The second class of r.v.s that we treat in this paper are sub-exponential. As mentioned in Remark 6, sub-exponential r.v.s satisfy an exponential moment bound with exponent equal to one, and hence these r.v.s do not satisfy (C1). The reason for treating sub-exponential distributions will be made apparent after presenting the Wasserstein concentration bound below for the sub-exponential case. Before presenting this bound, we specify below the condition that characterizes a sub-exponential r.v..

(C2) *There exists a $c > 0$ such that*

$$\mathbb{P}\left(X > \epsilon\right) \leq \exp(-c\epsilon), \textit{ and } \mathbb{P}\left(X \leq -\epsilon\right) \leq \exp(-c\epsilon), \;\; \forall \epsilon \geq 0. \tag{4}$$

A sub-exponential r.v. X satisfies

$$\mathbb{E}\left(\exp\left(c'|X|\right)\right) \leq 2,$$

where c' is an universal constant multiple of the constant c from (C2). Further, if the r.v. X is mean zero, then an equivalent characterization of sub-exponential r.v.s is the following (see Vershynin (2018, Sec. 2.7)): there exist positive constants σ and b such that

$$\mathbb{E}(\exp\left(\lambda X\right)) \leq \exp\left(\frac{\lambda^2\sigma^2}{2}\right) \text{ for every } |\lambda| < \frac{1}{b}.$$

Since (C1) requires $\beta > 1$, sub-exponential r.v.s fall outside the class of r.v.s satisfying (C1). The reason for separately handling sub-exponential r.v.s is that the bound in Fournier and Guillin (2015) is not satisfactory for such r.v.s. In particular, the tail bound there exhibits a power law decay for large values of ϵ. To obtain an exponential decay in the tail bound, we rely on a bound from Lei (2020), which is presented in Lemma 10 below. In this bound, which is an analogue of Lemma 3 for the case of sub-exponential distributions, we make all the constants explicit by tracing through the proofs of Theorem 3.1 and Corollary 5.2 of Lei (2020).

Lemma 10 ***(Wasserstein distance bound: sub-exponential case)*** *Let X be a r.v. satisfying (C2) with parameter c. Let F denote the CDF of X. Then, for every $n \geq 1$ and ϵ satisfying $\epsilon > \frac{384}{c\sqrt{n}}$, we have*

$$\mathbb{P}\left(W_1(F_n, F) > \epsilon\right) \leq \exp\left(-\frac{n\left(\epsilon - \frac{384}{c\sqrt{n}}\right)^2}{\frac{32}{c^2} + \frac{4}{c}\left(\epsilon - \frac{384}{c\sqrt{n}}\right)}\right).$$

Proof See Section 10.1.4. ■

2.3 Heavy-tailed distributions

The third class of distributions that we consider in this paper are those that satisfy a higher-moment bound, as specified by the following condition:

(C3) *There exists $\beta > 2$ such that $\mathbb{E}\left(|X|^\beta\right) < \top < \infty$.*

Distributions satisfying (C3) fall under the broad class of heavy-tailed distributions (see Nair (2012)), which includes sub-Gaussian and sub-exponential distributions, as well as distributions with infinite variance.

Next, we provide an analogue of Lemma 3 for the case of distributions satisfying (C3).

Lemma 11 ***(Wasserstein distance bound)*** *Let X be a r.v. with CDF F. Suppose that X satisfies (C3). Then, for every $n \geq 1$, $\epsilon \geq 0$ and $\eta \in (0, \beta)$, we have*

$$\mathbb{P}\left(W_1(F_n, F) > \epsilon\right) \leq c_1 \left(\exp\left(-c_2 n \epsilon^2\right) \mathbb{I}\left\{\epsilon \leq 1\right\} + n\left(n\epsilon\right)^{-(\beta-\eta)} \mathbb{I}\left\{\epsilon > 1\right\}\right),$$

where c_1, c_2 are constants that depend on η and the parameters β and $\top$ specified in (C3).

Proof See Section 10.1.5. ■

The constants are not available explicitly in the tail bound presented above, as the corresponding result in Fournier and Guillin (2015) does not specify constants explicitly, and we could not obtain an explicit expression for the constants as a function of β, η and $\top$ from the proofs given therein.

3. Risk measures and estimators

Let $\mathcal{L}$ denote the space of CDFs on $\mathbb{R}$. A *risk measure* is simply a map $\rho : \mathcal{L} \to \mathbb{R}$. With a slight abuse of notation, we will find it convenient to also think of a risk measure as a real-valued function on the set of real-valued r.v.s. More precisely, given a risk measure ρ and a real-valued r.v. X with CDF F, we will write $\rho(X)$ to mean $\rho(F)$.

3.1 Empirical risk estimators

Given a risk measure ρ and CDF F, we consider two intuitive empirical estimates of $\rho(F)$ in this paper. The first estimate involves applying ρ to the EDF F_n formed by drawing n samples from F. To make this precise, we define the estimate ρ_n of $\rho(F)$ by

$$\rho_n = \rho(F_n), \tag{5}$$

where F_n is as defined in (2).

Our second estimate of $\rho(F)$ involves applying ρ to a truncated sample drawn from F. We introduce some notation to make this precise. Given $\tau > 0$, we denote $F|_\tau = F\mathbb{I}\left\{-\tau \leq x < \tau\right\} + \mathbb{I}\left\{x \geq \tau\right\}$. If F is the CDF of a r.v. X, then $F|_\tau$ is the CDF of the r.v. $X\mathbb{I}\left\{-\tau \leq X < \tau\right\}$.

As before, let F_n denote the EDF of r.v. X, formed using i.i.d. samples $X_1, \ldots, X_n$. Our second estimate of the risk measure $\rho(F)$ is then given by

$$\rho_{n,\tau} \triangleq \rho(F_n|_\tau). \tag{6}$$

We consider two types of risk measures in this paper. The first type satisfies a Hölder-continuity requirement in the metric space of distributions under the Wasserstein distance, which is made precise below. The second type, which is presented later, handles some classes of risk measures that may not satisfy a Hölder condition.

3.2 Type-1 (T1) risk measures

(T1) *Let $(\mathcal{L}, W_1)$ denote the metric space of distributions, with Wasserstein distance as the metric. The risk measure $\rho(\cdot)$ is Hölder -continuous on $(\mathcal{L}, W)$, i.e., there exists $\kappa \in (0, 1]$ and $L > 0$ such that, for any two distributions $F, G \in \mathcal{L}$, the following holds:*

$$|\rho(F) - \rho(G)| \leq L\left(W_1(F, G)\right)^\kappa. \tag{7}$$

Optimized certainty equivalent (OCE) risk (Ben-Tal and Teboulle, 1986, 2007), spectral risk measure (Acerbi, 2002), and utility-based shortfall risk (Föllmer and Schied, 2002) are three popular families of risk measures that are of type (T1). We introduce these risk measures in the following sections and specialize the estimate given in (5) to the cases of the aforementioned risk measures.

3.2.1 OCE RISK

We adapt the definition of an OCE risk given in Lee et al. (2020). Given a nondecreasing, convex disutility function $\phi : \mathbb{R} \to \mathbb{R}$ and a r.v. X such that $\phi(X)$ is integrable, the OCE risk of X determined by ϕ is

$$\mathrm{oce}^\phi(X) \triangleq \inf_\xi \left\{\xi + \mathbb{E}\left[\phi\left(X - \xi\right)\right]\right\}. \tag{8}$$

The following lemma shows that an OCE is a risk measure of type (T1).

Lemma 12 *Let X and Y be r.v.s with CDFs F_X and F_Y, respectively, and let ϕ be a disutility function as in* (8). *Assume that ϕ is L-Lipschitz for some $L > 0$. Then*

$$|oce^\phi(X) - oce^\phi(Y)| \leq LW_1(F_X, F_Y). \tag{9}$$

Proof See Section 10.2.1. ■

Next, we discuss estimation of an OCE risk from i.i.d. samples $X_1, \ldots, X_n$, which are drawn from the distribution of X. We estimate $\mathrm{oce}^{\phi}(X)$ from such a sample by

$$\mathrm{oce}_n^{\phi} = \inf_{\xi} \left\{ \xi + \frac{1}{n} \sum_{i=1}^{n} \phi(X_i - \xi) \right\}. \tag{10}$$

It is straightforward to see that $\mathrm{oce}_n^{\phi} = \mathrm{oce}^{\phi}(Z_n)$, where Z_n is a r.v. with distribution F_n. The estimator (10) is thus a special case of the general empirical risk estimate given in (5).

Next, we define CVaR, a risk measure that is popular in financial applications. The CVaR at level $\alpha \in (0,1)$ for an integrable r.v X is defined by

$$C_\alpha(X) \triangleq \inf_{\xi} \left\{ \xi + \frac{1}{(1-\alpha)} \mathbb{E}\left(X - \xi\right)^{+} \right\}, \text{ where } (y)^{+} = \max\{y, 0\}. \tag{11}$$

It is well known (see Rockafellar and Uryasev (2000)) that the infimum in the definition of CVaR above is achieved for $\xi = \mathrm{VaR}_\alpha(X)$, where $\mathrm{VaR}_\alpha(X) = F^{-1}(\alpha)$ is the value-at-risk of the r.v. X at confidence level α. Thus CVaR may also be written alternatively as given, for instance, in Kolla et al. (2019). In the special case where X has a continuous distribution, $C_\alpha(X)$ equals the expectation of X conditioned on the event that X exceeds $\mathrm{VaR}_\alpha(X)$.

A comparison of (11) and (8) shows that CVaR at level $\alpha \in (0,1)$ is an OCE risk with the disutility function $\phi(x) = (1-\alpha)^{-1}(x)^{+}$, which is nondecreasing, convex, and L-Lipschitz for $L = (1-\alpha)^{-1}$. Applying Lemma 12 lets us conclude that CVaR is a risk measure of type (T1) satisfying

$$|C_\alpha(X) - C_\alpha(Y)| \leq (1-\alpha)^{-1} W_1(F_X, F_Y) \tag{12}$$

for any two integrable random variables X and Y.

We may now apply (10) to estimate CVaR from i.i.d. samples $X_1, \ldots, X_n$, which are drawn from the distribution of X. The resulting estimate of CVaR appeared earlier in Brown (2007), and is given by

$$c_{n,\alpha} = \inf_{\xi} \left\{ \xi + \frac{1}{n(1-\alpha)} \sum_{i=1}^{n} (X_i - \xi)^{+} \right\}.$$

It is straightforward to see that $c_{n,\alpha} = C_\alpha(Z_n)$, where Z_n is a r.v. with distribution F_n.

3.2.2 SPECTRAL RISK MEASURES

Spectral risk measures are an alternative generalization of CVaR. Given a weighting function $\phi : [0,1] \to [0,\infty)$, the spectral risk measure M_ϕ associated with ϕ is defined by

$$M_\phi(X) = \int_0^1 \phi(\beta) F_X^{-1}(\beta) \mathrm{d}\beta, \tag{13}$$

where X is a r.v. with CDF F_X. If the weighting function, also known as the *risk spectrum*, is increasing and integrates to 1, then M_ϕ is a coherent risk measure like CVaR. In fact, CVaR is itself a special case of (13), with $C_\alpha(X) = M_\phi$ for the risk spectrum $\phi = (1-\alpha)^{-1}\mathbb{I}\left\{\beta \geq \alpha\right\}$ (see Acerbi

(2002) and Dowd and Blake (2006) for details). Assuming the r.v. X models losses in a financial application, it is apparent that CVaR treats all losses above a certain threshold equally, by assigning the same weight for $\beta \geq \alpha$. One could think of a spectral risk measure with a weight function chosen such that higher losses are given more weight. An example of such a weight function, proposed by Cotter and Dowd (2006), is $\phi(\beta) = \frac{\varkappa \mathrm{e}^{-\varkappa(1-\beta)}}{1-\mathrm{e}^{-\varkappa}}$, $\beta \in [0,1]$, where $\varkappa$ is a constant that controls risk-aversion. The dual-power risk measure Wang (1996) employs the weighting function given by $\phi(\beta) = \varkappa(1-\beta)^{\varkappa-1}$, with $\varkappa \geq 1$.

Our next lemma shows that a spectral risk measure having a bounded weighting function is a risk measure of type (T1). The proof is omitted as the lemma follows very easily from the definition (13), and the last characterization of the Wasserstein distance given in (1).

Lemma 13 *Let X and Y be r.v.s with CDFs F_X and F_Y, respectively, and let $\phi : [0,1] \to [0,\infty)$ be a weighting function as in* (13). *Assume that $\phi(u) \leq K$ for all $u \in [0,1]$. Then*

$$|M_\phi(X) - M_\phi(Y)| \leq K W_1(F_X, F_Y).$$

Remark 14 *The two examples of the weighting function given above satisfy the boundedness requirement in the lemma above. However, there are spectral risk measures with weighting function that are not bounded. For example, the weighting function of the proportional hazard transform risk measure (Wang, 1995) is unbounded, and given by $\phi(\beta) = \frac{1}{\varkappa}\beta^{\frac{1}{\varkappa}-1}$, with $\varkappa \geq 1$. Such a risk measure may not fall under the class of (T1) risk measures. We shall handle spectral risk measures with unbounded weight functions through the (T2) class of risk measures introduced in subsection 3.3 below.*

We now discuss the estimation of a spectral risk measure from an i.i.d. sample $X_1, \ldots, X_n$, which is drawn from the CDF F of a r.v. X. Using the EDF, a natural empirical estimate of the spectral risk measure $M_\phi(X)$ of X is

$$m_{n,\phi} = \int_0^1 \phi(\beta) F_n^{-1}(\beta) \mathrm{d}\beta. \tag{14}$$

As in the case of an OCE risk, the estimate defined above is also a special case of the general estimate given in (5).

3.2.3 UTILITY-BASED SHORTFALL RISK (UBSR)

VaR as a risk measure is not popular owing to the fact that it is not sub-additive. CVaR overcomes this limitation, and is a coherent risk measure. Convex risk measures (Föllmer and Schied, 2002) are a more general class of measures than coherent risk measures, because sub-additivity and homogeneity imply convexity. UBSR form a popular class of convex risk measures.

For a r.v. X, the utility-based shortfall risk $S_\alpha(X)$ is defined as

$$S_\alpha(X) = \inf \left\{ \xi \in \mathbb{R} \mid \mathbb{E}\left(l(X-\xi)\right) \leq \alpha \right\}, \tag{15}$$

where $l : \mathbb{R} \to \mathbb{R}$ is a utility function. $S_\alpha(X)$ can be seen to be the value as well as the minimizer of the following constrained minimization problem:

$$\min_{\xi \in \mathbb{R}} \xi \qquad \text{subject to} \qquad \mathbb{E}\left(l(X-\xi)\right) \leq \alpha. \tag{16}$$

UBSR is a risk measure that generalizes VaR, since one recovers VaR from (15) by employing an indicator function for l. While VaR is not a convex risk measure, a suitable choice for the utility function l in (15) can yield a convex risk measure, cf. Section 4.9 of Föllmer and Schied (2016). Furthermore, in contrast to CVaR, UBSR is invariant under randomization (Dunkel and Weber, 2010). Invariance under randomization formally means that given two r.v.s X_1, X_2 satisfying $S_\alpha(X_i) \leq 0, i = 1, 2$, the compound r.v. Z that chooses between X_1 and X_2 using an independent Bernoulli r.v. also satisfies $S_\alpha(Z) \leq 0$.

For the purpose of showing that UBSR is of type (T1), we require that l be Lipschitz. While our results below require no additional assumptions on l, we point out that a convexity assumption on l leads to a useful dual representation (Föllmer and Schied, 2002). Likewise, if the utility is increasing, then the estimation of UBSR from i.i.d. samples can be performed in a computationally efficient manner (see Hu and Zhang (2018)).

The following lemma shows that UBSR is a risk measure of type (T1).

Lemma 15 *Suppose the utility function l is non-decreasing, and there exist $K, k > 0$ such that l is K-Lipschitz and satisfies $l(x_2) \geq l(x_1) + k(x_2 - x_1)$ for every $x_1, x_2 \in \mathbb{R}$ satisfying $x_2 \geq x_1$. Let X and Y be r.v.s with CDFs F_X and F_Y, respectively. Then*

$$|S_\alpha(X) - S_\alpha(Y)| \leq \frac{K}{k} W_1(F_X, F_Y). \tag{17}$$

Proof See Section 10.2.2. ■

Next, we discuss the estimation of UBSR $S_\alpha(X)$. Given n i.i.d. samples $X_1, \ldots, X_n$, UBSR is estimated as the solution of the following constrained optimization problem (Hu and Zhang, 2018):

$$\min_{\xi \in \mathbb{R}} \xi \qquad \text{subject to} \qquad \frac{1}{n}\sum_{i=1}^{n} l(X_i - \xi) \leq \alpha. \tag{18}$$

The problem above can be seen as a sample-average approximation to (16). In other words, the solution of (18) is the UBSR value of a r.v. distributed according to the EDF F_n of X. Thus, as in the case of OCE risk and spectral risk measures, the estimation scheme provided above is a special case of the general estimate given in (5).

In Hu and Zhang (2018), the authors show that the problem (18) can be solved efficiently using a bisection method if the utility function l is increasing and an interval containing $S_\alpha(X)$ is known. The reader is referred to Hu and Zhang (2018) for further details.

3.2.4 TWO EXAMPLES

The (T1) measures seen so far, namely, OCE risk, spectral risk measures, and UBSR all satisfy (7) with $\kappa = 1$. To illustrate the point that the additional generality provided in (7) is not vacuous, we next present an example of a risk measure that satisfies the definition of a (T1) measure for some $\kappa < 1$, but not for $\kappa = 1$.

Example 1 *Consider the risk measure $\rho(F) = \int_0^1 (1 - F(x))^{\frac{1}{3}} \mathrm{d}x$. To show that ρ is (T1), let F and G be two distributions. We have*

$$\begin{aligned} |\rho(F) - \rho(G)| &\leq \int_0^1 |(1 - F(x))^{\frac{1}{3}} - (1 - G(x))^{\frac{1}{3}}| \mathrm{d}x \\ &\leq 2^{\frac{2}{3}} \int_0^1 |F(x) - G(x)|^{\frac{1}{3}} \mathrm{d}x \\ &\leq 2^{\frac{2}{3}} \left[\int_0^1 |F(x) - G(x)| \mathrm{d}x \right]^{\frac{1}{3}} \leq 2^{\frac{2}{3}} [W_1(F, G)]^{\frac{1}{3}}, \end{aligned}$$

where the last inequality above follows from Lemma 2, the first and third inequalities above follow from Jensen's inequality, and the second inequality follows from the fact that $|a^\alpha - b^\alpha| \leq 2^{1-\alpha}|a - b|^\alpha$ for every $a, b \in \mathbb{R}$ and $\alpha \in (0, 1]$ such that α equals the ratio of two odd positive integers (see Fact 2.2.77 in Bernstein (2018)). This shows that ρ satisfies (7) for all F and G with $\kappa = \frac{1}{3}$.

We claim that there exists no $L > 0$ such that ρ satisfies (7) with $\kappa = 1$ for all distributions. To see this, let $L > 0$, and pick $p \in (0, L^{-\frac{3}{2}})$. Define CDFs G and F by $G(x) \triangleq \mathbb{I}\{x \geq 0\}$ and $F(x) \triangleq (1 - p)\mathbb{I}\{x \geq 0\} + p\mathbb{I}\{x \geq 1\}$ for all $x \in \mathbb{R}$. It is a simple matter to calculate $\rho(F) = p^{\frac{1}{3}}$, $\rho(G) = 0$, and $W_1(F, G) = p$. Consequently, we have $\frac{|\rho(F) - \rho(G)|}{W_1(F,G)} = p^{-\frac{2}{3}} > L$. Since $L > 0$ was chosen arbitrarily, our claim follows.

To provide motivation for the class of risk measures that we introduce next, we provide an example of a risk measure that is not (T1).

Example 2 *Consider the risk measure $\rho(F) = \int_0^\infty \sqrt{1 - F(x)} \mathrm{d}x$. To show that ρ is not (T1), let $L > 0$ and $\kappa \in (0, 1]$ be given. Choose $p \in (0, L^{-2})$. Additionally, choose $a = 1$ or $a > [Lp^{\kappa - \frac{1}{2}}]^{\frac{1}{1-\kappa}}$ accordingly as $\kappa = 1$ or $\kappa < 1$, respectively. Define CDFs G and F by $G(x) \triangleq \mathbb{I}\{x \geq 0\}$ and $F(x) \triangleq (1 - p)\mathbb{I}\{x \geq 0\} + p\mathbb{I}\{x \geq a\}$ for all $x \in \mathbb{R}$. Note that G is the CDF of a r.v. that equals 0 a.s., while F is the CDF of a r.v. that takes the value 0 with probability $(1 - p)$ and the value a with probability p. It is a simple matter to evaluate $\rho(G) = 0$ and $\rho(F) = a\sqrt{p}$. Likewise, Lemma 2 can be used to compute $W_1(F, G) = ap$. One can now easily check that, in the case where $\kappa = 1$, $\frac{|\rho(F) - \rho(G)|}{[W_1(F,G)]^\kappa} = p^{-\frac{1}{2}} > L$, while in the case where $\kappa < 1$, $\frac{|\rho(F) - \rho(G)|}{[W_1(F,G)]^\kappa} = \frac{a^{1-\kappa}}{p^{\kappa - \frac{1}{2}}} > L$. We have thus shown that, for every $L > 0$ and $\kappa \in (0, 1]$, there exist CDFs F and G such that (7) fails to hold. It follows that the risk measure ρ is not (T1).*

3.3 Type-2 (T2) risk measures

In this subsection, we consider risk measures that satisfy a weaker version of the Hölder-continuity requirement specified in (T1) risk measures, and the definition below makes this continuity requirement precise.

(T2) *Let $(\mathcal{L}, W_1)$ denote the metric space of distributions, with Wasserstein distance as the metric. The risk measure $\rho(\cdot)$ satisfies* a truncated Hölder-continuity condition with a tail bound *on $(\mathcal{L}, W_1)$ if there exist positive constants α_1 ,α_2, α_3 L_1, L_2, L_3, K_1, K_2, and γ such that $\alpha_1, K_1, K_2 \leq 1$ and, for every choice of two distributions $F, G \in \mathcal{L}$ and every $\tau > 0$, the following holds:*

$$|\rho(F) - \rho(G|_\tau)| \leq L_1 \left(W_1(F, G)\right)^{\alpha_1} \tau^\gamma + L_2 \int_{K_1\tau}^\infty [1 - F(z)]^{\alpha_2} \mathrm{d}z$$

$$+ L_3 \int_{-\infty}^{-K_2\tau} [F(z)]^{\alpha_3} \mathrm{d}z. \tag{19}$$

To see the significance of the inequality (19), let ρ be a type (T2) risk measure satisfying (19), and consider two distributions F and G supported on the interval $[-K_2\tau, K_1\tau]$ for some $\tau > 0$. Then $G|_\tau = G$, and $1 - F(z) = 0 = F(y)$ for all $z > K_1\tau$ and $y < -K_2\tau$. The inequality (19) then yields

$$|\rho(F) - \rho(G)| \leq L_1 \left(W_1(F,G)\right)^{\alpha_1} \tau^\gamma.$$

The equation above indicates that, when restricted to distributions having bounded support, a type (T2) risk measure is Hölder continuous with respect to the Wasserstein distance with a Hölder constant that grows in an unbounded manner as the length of the support increases.

The reader will notice that the condition (19) is not symmetric in the distributions F and G, and it may seem that a more natural-looking condition might be to require that

$$\begin{aligned} |\rho(F) - \rho(G)| \leq L_1 \left(W_1(F,G)\right)^{\alpha_1} \tau^\gamma + L_2 \int_{\tau}^{\infty} |F(z) - G(z)|^{\alpha_2} \mathrm{d}z \\ + L_3 \int_{-\infty}^{-\tau} |F(z) - G(z)|^{\alpha_3} \mathrm{d}z \end{aligned} \tag{20}$$

hold for all $\tau \geq 1$. Note that the condition (20) implies (19), and hence (19) is the weaker of the two. Also, as we show below, (19) suffices for the bounds that we wish to derive and apply to known risk measures. Hence we choose to go with the weaker, albeit odd-looking, condition (19).

In the following subsection, we introduce cumulative prospect theory (CPT), which is a prominent risk measure in human-centered decision making systems, and an example of a (T2) risk measure. Subsequently, we describe rank-dependent expected utility (RDEU), which we show to be a special instance of the CPT-value that we define below. The risk measure that was seen to be not (T1) in example 2 is a special case of the CPT-value that we introduce next.

3.3.1 CUMULATIVE PROSPECT THEORY (CPT)

For any r.v. X, the CPT-value is defined as

$$C(X) = \int_0^\infty w^+ \left(\mathbb{P}\left(u^+(X) > z\right)\right) \mathrm{d}z - \int_0^\infty w^- \left(\mathbb{P}\left(u^-(X) > z\right)\right) \mathrm{d}z. \tag{21}$$

Let us deconstruct the above definition. First, the functions $u^+, u^- : \mathbb{R} \to \mathbb{R}_+$ are utility functions which are assumed to be continuous, with $u^+(x) = 0$ when $x \leq 0$ and increasing otherwise, and with $u^-(x) = 0$ when $x \geq 0$ and decreasing otherwise. The utility functions capture the human inclination to play safe with gains and take risks with losses – see Fig 1. Second, $w^+, w^- : [0,1] \to [0,1]$ are weight functions, which are assumed to be continuous, non-decreasing and satisfy $w^+(0) = w^-(0) = 0$ and $w^+(1) = w^-(1) = 1$. The weight functions w^+, w^- capture the human inclination to view probabilities in a non-linear fashion. Tversky and Kahneman (1992); Barberis (2013) (see Fig 2 from Tversky and Kahneman (1992)) recommend the following choices for w^+ and w^-, based on inference from experiments involving human subjects:

$$w^+(p) = \frac{p^{0.61}}{(p^{0.61} + (1-p)^{0.61})^{\frac{1}{0.61}}}, \text{ and } w^-(p) = \frac{p^{0.69}}{(p^{0.69} + (1-p)^{0.69})^{\frac{1}{0.69}}}.$$

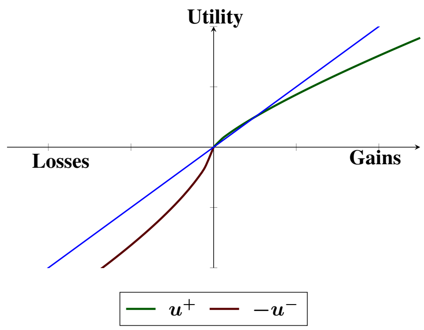

Figure 1: Utility function

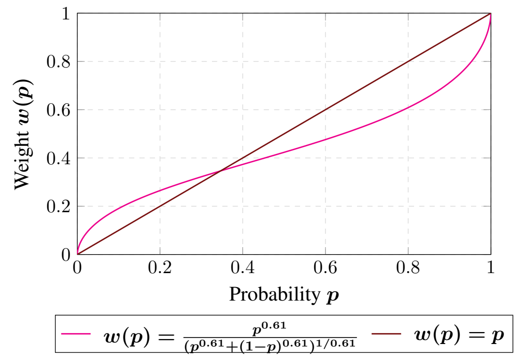

Figure 2: Weight function

The lemma below shows that CPT-value is a risk measure of type (T2) under certain assumptions.

Lemma 16 *Suppose that the utility functions $u^+, u^- : \mathbb{R} \to \mathbb{R}_+$ are differentiable, and their derivatives are bounded above and below by $K^+ > 0$ and $k^+ > 0$, and K^- and $k^- > 0$, respectively, in absolute value. Further assume that the weight functions w^+ and w^- are Hölder continuous with exponent $\alpha \in (0,1]$ and Hölder constant $L > 0$, and let $F, G \in \mathcal{L}$ be CDFs. Then, for every $\tau > 0$, we have*

$$|C(F) - C(G|_\tau)| \le L(K^+ + K^-)[W_1(F,G)]^\alpha \tau^{(1-\alpha)} + LK^+ \int_{\frac{k^+}{K^+}\tau}^{\infty} [1 - F(z)]^\alpha \mathrm{d}z$$
$$+ LK^- \int_{-\infty}^{-\frac{k^-}{K^-}\tau} [F(z)]^\alpha \mathrm{d}z.$$

Proof See Section 10.2.3. ■

Remark 17 *It is not difficult to see that the risk measure considered in Example 2 can be written as a right-tailed version of CPT value by letting u^- be identically zero, u^+ to be the identity map, and w^+ to be the square root function. Although these choices do not satisfy the assumptions of Lemma 16, one can follow the steps in the proof of Lemma 16 to show that the risk measure in the example is (T2). Example 2 thus provides a risk measure that is (T2) but not (T1).*

We now describe the CPT-value estimation scheme, which is a variant of the one proposed in Prashanth et al. (2016). In particular, unlike the aforementioned reference, we employ a truncated EDF to obtain the CPT-value estimate. Let $X_i, i = 1, \ldots, n$, denote n independent samples from the distribution of X. For any given τ_n, and real-valued functions u^+ and u^-, let $(F_n|_{\tau_n})^+$ and $(F_n|_{\tau_n})^-$ denote the truncated EDFs formed from the samples $\{u^+(X_i), i = 1, \ldots, n\}$ and $\{u^-(X_i), i = 1, \ldots, n\}$, respectively. Using the truncated EDFs, the CPT-value is estimated as follows:

$$C_n = \int_0^{\tau_n} w^+(1 - (F_n|_{\tau_n})^+(x))\mathrm{d}x - \int_0^{\tau_n} w^-(1 - (F_n|_{\tau_n})^-(x))\mathrm{d}x. \tag{22}$$

Notice that we have substituted the complementary (truncated) EDFs $\left(1-(F_n|_{\tau_n})^+(x)\right)$ and $\left(1-(F_n|_{\tau_n})^-(x)\right)$ for $\mathbb{P}\left(u^+(X)>x\right)$ and $\mathbb{P}\left(u^-(X)>x\right)$, respectively, in (21), and then performed an integration of the weight function composed with the complementary EDF. It is apparent that the CPT-value estimator in (22) equals the CPT value of the truncated EDF $F_n|_{\tau_n}$, and thus is a special case of the estimator (6) for a general risk measure.

Let $\tilde{X}_i = X_i\mathbb{I}\left\{X_i \le \tau_n\right\}$, for $i=1,\ldots,n$. Using arguments similar to that in Section III of Prashanth et al. (2016), the first and second integral, say C_n^+ and C_n^-, in (22) can be easily computed using the order statistics $\{\tilde{X}_{(1)},\ldots,\tilde{X}_{(n)}\}$ of the truncated samples $\{\tilde{X}_i : i=1,\ldots,n\}$ as follows:

$$C_n^+ = \sum_{i=1}^{n} u^+(\tilde{X}_{[i]})\left[w^+\left(\frac{n+1-i}{n}\right)-w^+\left(\frac{n-i}{n}\right)\right],$$
$$C_n^- = \sum_{i=1}^{n} u^-(\tilde{X}_{[i]})\left[w^-\left(\frac{i}{n}\right)-w^-\left(\frac{i-1}{n}\right)\right].$$

3.3.2 RANK-DEPENDENT EXPECTED UTILITY (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 a continuous, increasing function satisfying $u(0)=0$. Then, following Quiggin (2012), the RDEU-value $V(F)$ is defined by

$$V(F)=\int_{-\infty}^{\infty} u(x)\mathrm{d}(w\circ F)(x).$$

The result below shows that RDEU-value is a special case of CPT-value as defined in (21) To elaborate, CPT allows one the freedom of choosing two different weight functions w^+ and w^-. By suitably defining these, we show that RDEU is a special case of CPT.

Lemma 18 *Let $w:[0,1]\to[0,1]$ and $u:\mathbb{R}\to\mathbb{R}$ be a weight function and a utility function as above. Assume that u is unbounded above and below. Define $u^+(x)=u(x\mathbb{I}\{x\ge 0\})$, $u^-(x)=-u(x\mathbb{I}\{x<0\})$, for $x\in\mathbb{R}$. Let $w^-(p)=w(p)$, and $w^+(p)=1-w(1-p)$, for $p\in[0,1]$. Then, for any r.v. X with CDF F, we have*

$$V(F)=\int_0^{\infty} w^+\left(\mathbb{P}\left(u^+(X)>z\right)\right)\mathrm{d}z-\int_0^{\infty} w^-\left(\mathbb{P}\left(u^-(X)>z\right)\right)\mathrm{d}z.$$

Proof See Section 10.2.4. ■

Thus, the RDEU-value V is a special case of CPT-value, with $u^\pm$ and $w^\pm$ chosen as in the statement of the lemma above. Note that for the choice of w^+ and w^- given in Lemma 16, we have

$$w^+(p)+w^-(1-p)=1.$$

CPT-value, as defined in (21), is more general, as the weight functions w^+ and w^- are not required to satisfy the aforementioned equality. Owing to the fact that RDEU is a special case of CPT, the estimation scheme presented in (22) applies to RDEU as well.

3.3.3 DISTORTED RISK MEASURE (DRM)

For a r.v. X with CDF F, the DRM $D(X)$ is defined as

$$D(X) = \int_{-\infty}^{0} \left[w\left(1 - F(z)\right) - 1\right] \mathrm{d}z + \int_{0}^{\infty} w\left(1 - F(z)\right) dz,$$

where $w : [0, 1] \to [0, 1]$ is a weight function that satisfies $w(0) = 0$ and $w(1) = 1$. If w is concave, then DRM is a coherent risk measure. Further, DRMs are equivalent to spectral risk measures if the weight function is increasing and differentiable. In this case, the risk spectrum for the equivalent spectral risk measure is given by $\phi(\beta) = w'(1 - \beta)$, for $\beta \in [0, 1]$.

It is easy to see that DRMs are of type (T2), as they can be treated as a special case of CPT-value. Thus, the bounds we derive for CPT-value estimation can be easily specialized to handle the case of DRMs. While DRMs are equivalent to spectral risk measures, the bounds for spectral risk measures derived earlier require that the risk spectrum be bounded. As noted in Remark 14, there are spectral risk measures (or DRMs) that do not satisfy this boundedness requirement, and for such DRMs, one could take the (T2) route to arrive at estimation bounds. An example of such a risk measure is the proportional hazard transform, which has the following risk spectrum: $\phi(\beta) = \frac{1}{\varkappa}\beta^{\frac{1}{\varkappa}-1}$, with $\varkappa \geq 1$.

4. Map of the results

In the next three sections, we provide bounds in expectation as well as concentration bounds for estimates of the risk measures described in the previous section under the assumptions ((C1))-((C3)) (as well as (C4) appearing in the next section) on the underlying distribution. Since the number of combinations of risk measures and distributional assumptions is rather large, we provide here a tabular summary of all the bounds that will be presented in the succeeding sections.

Table 2: Summary of the bounds in expectation of the form $\mathbb{E}\left(|\rho_n - \rho(X)|\right)$ for various choices of risk measure $\rho(X)$. Here X is a r.v. satisfying $\mathbb{E}\left(|X|^{\beta}\right) < \top < \infty$ for some $\beta > 1$, and ρ_n is an estimate of $\rho(X)$ using n i.i.d. samples.

Risk measure	Bound in expectation	Reference
OCE (includes CVaR)	$O\left(\frac{1}{n^{\min\{\frac{1}{2}, 1-\frac{1}{\beta}\}}}\right)$	Corollary 20
Spectral risk measure	$O\left(\frac{1}{n^{\min\{\frac{1}{2}, 1-\frac{1}{\beta}\}}}\right)$	Corollary 21
Utility-based shortfall risk	$O\left(\frac{1}{n^{\min\{\frac{1}{2}, 1-\frac{1}{\beta}\}}}\right)$	Corollary 22
Cumulative prospect theory	$O\left(\frac{1}{\left(n^{\min\{\frac{1}{2}, 1-\frac{1}{\beta}\}}\right)^{\frac{\beta\alpha-1}{\beta-1}}}\right)$	Corollary 24

Table 3: Summary of the concentration bounds of the form $\mathbb{P}\left(|\rho_n - \rho(X)| > \epsilon\right)$, for various choices of risk measure $\rho(X)$. Here X is a sub-Gaussian r.v. with parameter σ. Here the constants c_1, c_2 are functions of the sub-Gaussianity parameter σ, but an explicit expression is not available. For bounds with explicit constants, see Table 4

Risk measure	**Concentration bound**	**Reference**
OCE	$2c_1 \exp\left(-\frac{c_2 n \epsilon^2}{L^2}\right)$	Corollary 28 with $\beta = 2$
Spectral risk measure	$2c_1 \exp\left(-\frac{c_2 n \epsilon^2}{K^2}\right)$	Corollary 30 with $\beta = 2$
Utility-based shortfall risk	$2c_1 \exp\left(-\frac{c_2 k^2 n \epsilon^2}{K^2}\right)$	Corollary 32 with $\beta = 2$
Cumulative prospect theory	$c_1 \exp\left(-c_2 n \left(\frac{\epsilon - c_3(n)}{L(K^+ + K^-)\tau_n^{1-\alpha}}\right)^{\frac{2}{\alpha}}\right)$, $c_3(n) = \frac{\text{const}}{\sqrt{\log n}\, n^{\alpha(1-\alpha)}}$ $\tau_n = \text{const}\sqrt{\log n}$.	Corollary 39 with $\beta = 2$

Table 4: Summary of the concentration bounds of the form $\mathbb{P}\left(|\rho_n - \rho(X)| > \epsilon\right)$, for various choices of risk measure $\rho(X)$. Here X is a sub-Gaussian r.v. with parameter σ. Unlike Table 3, the bounds here feature explicit constants.

Risk measure	**Concentration bound**	**Reference**
OCE	$\exp\left(-\frac{n}{256\sigma^2 \mathsf{e}}\left(\frac{\epsilon}{L} - \frac{512\sigma}{\sqrt{n}}\right)^2\right)$	Corollary 29
Spectral risk measure	$\exp\left(-\frac{n}{256\sigma^2 \mathsf{e}}\left(\frac{\epsilon}{K} - \frac{512\sigma}{\sqrt{n}}\right)^2\right)$	Corollary 31
Utility-based shortfall risk	$\exp\left(-\frac{n}{256\sigma^2 \mathsf{e}}\left(\frac{k\epsilon}{K} - \frac{512\sigma}{\sqrt{n}}\right)^2\right)$	Corollary 33
Cumulative prospect theory	$\exp\left(-\frac{n}{256\sigma^2 \mathsf{e}}\right. \times$ $\left.\left(\left(\frac{\epsilon - c_3(n)}{L(K^+ + K^-)\left[\max\left\{\frac{K^+}{k^+}, \frac{K^-}{k^-}\right\}(\log n)^{\frac{1}{2}}\right]^{1-\alpha}}\right)^{\frac{1}{\alpha}} - \frac{512\sigma}{\sqrt{n}}\right)^2\right)$	Corollary 40

Table 5: Summary of the concentration bounds of the form $\mathbb{P}\left(|\rho_n - \rho(X)| > \epsilon\right)$, for various choices of risk measure $\rho(X)$. Here X is a sub-exponential r.v. with parameter c.

Risk measure	**Concentration bound**	**Reference**
OCE	$\exp\left(-\frac{n}{\frac{32}{c^2}+\frac{4}{c}\left(\frac{\epsilon}{L}-\frac{384}{c\sqrt{n}}\right)}\left(\frac{\epsilon}{L}-\frac{384}{c\sqrt{n}}\right)^2\right)$	Corollary 42
Spectral risk measure	$\exp\left(-\frac{n}{\frac{32}{c^2}+\frac{4}{c}\left(\frac{\epsilon}{K}-\frac{384}{c\sqrt{n}}\right)}\left(\frac{\epsilon}{K}-\frac{384}{c\sqrt{n}}\right)^2\right)$	Corollary 43
Utility-based shortfall risk	$\exp\left(-\frac{n}{\frac{32}{c^2}+\frac{4}{c}\left(\frac{k\epsilon}{K}-\frac{384}{c\sqrt{n}}\right)}\left(\frac{k\epsilon}{K}-\frac{384}{c\sqrt{n}}\right)^2\right)$	Corollary 44
Cumulative prospect theory	$\exp\Bigg(-\frac{n}{\frac{32}{c^2}+\frac{4}{c}\left(\left(\frac{\epsilon-\frac{(K^++K^-)L}{c\alpha n^\alpha}}{(K^++K^-)L\tau_n^{1-\alpha}}\right)^{\frac{1}{\alpha}}-\frac{384}{c\sqrt{n}}\right)}$ $\times\left(\left(\left(\frac{\epsilon-\frac{(K^++K^-)L}{c\alpha n^\alpha}}{(K^++K^-)L\tau_n^{1-\alpha}}\right)^{\frac{1}{\alpha}}-\frac{384}{c\sqrt{n}}\right)^2\right)\Bigg),$ $\tau_n = \text{const}\sqrt{\log n}$	Corollary 46

5. Bounds in expectation for risk estimation

In this section, we provide non-asymptotic bounds, which hold in expectation, for estimation of risk measures. For these bounds, we assume that the underlying distribution has a finite higher-moment bound. More precisely, we assume that the following condition holds.

(C4) *There exists* $\beta > 1$ *such that* $\mathbb{E}\left(|X|^\beta\right) < \top < \infty$.

5.1 Bounds for risk measures of type (T1)

Theorem 19 *Suppose* X *is a r.v. satisfying (C4),* $\rho : \mathcal{L} \to \mathbb{R}$ *is a risk measure of type (T1) with parameters* $L > 0$ *and* $\kappa > 0$ *as in* (7)*, and* ρ_n *is given by* (5)*. Then, for every* $n \geq 1$*, we have*

$$\mathbb{E}\left(|\rho_n - \rho(X)|\right) \leq L\left(\frac{2^{\beta+3}\top}{n^{\min\{\frac{1}{2},1-\frac{1}{\beta}\}}}\right)^{\kappa}.$$

Proof See Section 10.3.1. ∎

The following corollaries provide bounds in expectation for empirical OCE, SRM and UBSR. The proofs of these corollaries follow in a straightforward fashion using the result in Theorem 19, and we omit the proof details.

Corollary 20 (Bound in expectation: OCE) *Suppose X is a r.v. satisfying (C4), $\phi : \mathbb{R} \to \mathbb{R}$ is a L-Lipschitz disutility function, $oce^{\phi}(X)$ is the OCE risk of X as defined in* (8)*, and oce_n^{ϕ} is its empirical estimate as in* (10)*. Then, for every $n \geq 1$, we have*

$$\mathbb{E}\left(\left|oce_n^{\phi} - oce^{\phi}(X)\right|\right) \leq L\left(\frac{2^{\beta+3}\top}{n^{\min\{\frac{1}{2},1-\frac{1}{\beta}\}}}\right).$$

In light of the fact that CVaR is an OCE risk satisfying (12), it is clear that the bound above holds for empirical CVaR with $L = (1-\alpha)^{-1}$.

Corollary 21 (Bound in expectation: SRM) *Suppose X is a r.v. satisfying (C4), $\phi : [0,1] \to [0,\infty)$ is a weighting function uniformly bounded above by $K > 0$, $M_{\phi}(X)$ is the OCE risk of X as defined in* (13)*, and $m_{n,\phi}$ is its empirical estimate as in* (14)*. Then, for every $n \geq 1$, we have*

$$\mathbb{E}\left(\left|m_{n,\phi} - M_{\phi}(X)\right|\right) \leq K\left(\frac{2^{\beta+3}\top}{n^{\min\{\frac{1}{2},1-\frac{1}{\beta}\}}}\right).$$

Corollary 22 (Bound in expectation: UBSR) *Suppose X is a r.v. satisfying (C4), $l : \mathbb{R} \to \mathbb{R}$ is a utility function satisfying the assumptions in Lemma 15, and $S_{\alpha}(X)$ is the UBSR of X as defined in* (15)*. Let $\xi_{n,\alpha}$ denote the solution to the constrained problem in* (18)*. Then, for every $n \geq 1$, we have*

$$\mathbb{E}\left(\left|\xi_{n,\alpha} - S_{\alpha}(X)\right|\right) \leq \frac{K}{k}\left(\frac{2^{\beta+3}\top}{n^{\min\{\frac{1}{2},1-\frac{1}{\beta}\}}}\right),$$

where $K, k > 0$ are as in Lemma 15.

5.2 Bounds for risk measures of type (T2)

In this subsection, we provide bounds on the expected estimation error for a risk measure of type (T2).

Theorem 23 *Suppose X is a r.v. satisfying (C4), and $\rho : \mathcal{L} \to \mathbb{R}$ is a risk measure of type (T2), with parameters $\alpha_1, \alpha_2, \alpha_3, L_1, L_2, L_3, K_1, K_2$ as defined in* (19)*. Further, assume that $\min\{\beta\alpha_2, \beta\alpha_3\} > 1$. Fix $\tau > 0$. Then, for every $n \geq 1$, we have*

$$\mathbb{E}\left(\left|\rho_{n,\tau} - \rho(X)\right|\right) \leq L_1\tau^{\gamma}\left(\frac{2^{\beta+3}\top}{n^{\min\{\frac{1}{2},1-\frac{1}{\beta}\}}}\right)^{\alpha_1} + \frac{L_2\top^{\alpha_2}}{(\beta\alpha_2-1)(K_1\tau)^{\beta\alpha_2-1}} + \frac{L_3\top^{\alpha_3}}{(\beta\alpha_3-1)(K_2\tau)^{\beta\alpha_3-1}}, \tag{23}$$

where $\rho_{n,\tau} = \rho(F_n|_{\tau})$.

Proof See Section 10.3.2. ∎

We now specialize the result in Theorem 23 for the case of CPT-value. We consider the CPT value estimator based on truncation, defined in (22).

Corollary 24 (Bound in expectation: CPT value) *Assume that the conditions in Lemma 16 hold. Suppose that X satisfies (C4) for some $\beta > 1$ such that $\beta\alpha > 1$, where $\alpha \in (0, 1]$ is as in Lemma 16. For each $n \geq 1$, set $\tau_n = \left(n^{\min\{\frac{1}{2}, 1-\frac{1}{\beta}\}}\right)^{\frac{1}{(\beta-1)}}$, and form the CPT-value estimate C_n using* (22). *Then we have*

$$\mathbb{E}\left(|C_n - C(X)|\right) \leq \frac{L\top^{\alpha}}{\left(n^{\min\{\frac{1}{2}, 1-\frac{1}{\beta}\}}\right)^{\frac{\beta\alpha-1}{\beta-1}}} \times \left(2^{(\beta+3)\alpha}(K^+ + K^-) + \frac{K^+}{(\beta\alpha - 1)\left(\frac{k^+}{K^+}\right)^{\beta\alpha-1}} + \frac{K^-}{(\beta\alpha - 1)\left(\frac{k^-}{K^-}\right)^{\beta\alpha-1}}\right).$$

Proof See Section 10.3.3. . ■

As discussed in subsection 3.3, RDEU and DRM can be viewed as special cases of CPT value, and are hence covered by Corollary 24.

6. Concentration bounds for distributions satisfying (C1)

In this section, we provide concentration bounds for the risk estimates introduced in section 3 using the Wasserstein distance bound in Lemma 3. The concentration bounds assume that the underlying distribution satisfies condition (C1), that is, an exponential moment bound with an exponent greater than one. Sub-Gaussian distributions are a popular class of distributions that satisfy this assumption. As mentioned before, sub-exponential distributions are treated separately in the next section.

In the following sections, we relate the estimation errors for both risk measure types to the Wasserstein distance between the EDF and the underlying CDF. The resulting concentration bounds apply to unbounded random variables as long as they belong to one of the three classes mentioned above. An alternative approach involves an application of the Dvoretzky-Kiefer-Wolfowitz (DKW) theorem (Wasserman, 2015, Chapter 2). Such an approach has been used to obtain concentration bounds for CVaR/CPT (cf. Thomas and Learned-Miller (2019); Cheng et al. (2018), but the sup norm in the DKW inequality will make the resulting bounds applicable only to bounded r.v.s. In contrast, employing the Wasserstein metric allows us to derive concentration bounds for a broader class of r.v.s.

6.1 Bounds for Risk Measures of Type (T1)

Given a risk measure ρ, a r.v. X with CDF F, and $n > 0$, we form an empirical estimate ρ_n of $\rho(X)$ using (5). Recall that the latter estimate is obtained by applying ρ to the EDF formed by drawing n samples from F.

Theorem 25 *Suppose X is a r.v. satisfying (C1), and $\rho : \mathcal{L} \to \mathbb{R}$ is a risk measure of type (T1) with parameters $L > 0$ and $\kappa > 0$ as in 7. Then, for every $\epsilon > 0$ and $n \geq 1$, we have*

$$\mathbb{P}\left(|\rho_n - \rho(X)| > \epsilon\right) \leq c_1 \left(\exp\left(-c_2 n \left(\frac{\epsilon}{L}\right)^{\frac{2}{\kappa}}\right) \mathbb{I}\left\{\epsilon \leq L\right\} + \exp\left(-c_3 n \left(\frac{\epsilon}{L}\right)^{\frac{\beta}{\kappa}}\right) \mathbb{I}\left\{\epsilon > L\right\}\right), \tag{24}$$

where the constants c_1, c_2, and c_3 are as in Lemma 3.

Proof See Section 10.4.1. ■

The result above can be easily specialized for the case of sub-Gaussian distributions by using $\beta = 2$ in the bound above.

Remark 26 *It is easy to see from the first equality in* (1) *that $X \mapsto \mathbb{E}(X)$ is a (T1) risk measure with parameters $L = \kappa = 1$. Theorem 25 applies with $\beta = 2$, and yields the well-known Hoeffding's inequality for the concentration of sample mean in the case of sub-Gaussian r.v.s.*

Next, we present a concentration bound for the sub-Gaussian case, but with explicit constants.

Theorem 27 *Let X be a sub-Gaussian r.v. with parameter $\sigma > 0$. Suppose $\rho : \mathcal{L} \to \mathbb{R}$ is a risk measure of type (T1), with parameters L and κ. Then, for every $n \geq 1$ and ϵ such that ϵ such that $\frac{512\sigma}{\sqrt{n}} < \left(\frac{\epsilon}{L}\right)^{\frac{1}{\kappa}} < \frac{512\sigma}{\sqrt{n}} + 16\sigma\sqrt{\mathsf{e}}$, we have*

$$\mathbb{P}\left(|\rho_n - \rho(X)| > \epsilon\right) \leq \exp\left(-\frac{n}{256\sigma^2\mathsf{e}}\left(\left(\frac{\epsilon}{L}\right)^{\frac{1}{\kappa}} - \frac{512\sigma}{\sqrt{n}}\right)^2\right).$$

Proof See Section 10.4.2. ■

6.1.1 BOUNDS FOR EMPIRICAL OCE

We now provide a concentration bound for the empirical OCE estimate (10), by relating the estimation error $\left|\mathrm{oce}_n^{\phi} - \mathrm{oce}^{\phi}(X)\right|$ to the Wasserstein distance between the true and empirical distribution functions, and subsequently invoking Lemma 3 that bounds the Wasserstein distance between these two distributions.

Corollary 28 (OCE concentration) *Suppose X satisfies (C1) for some $\beta > 1$. Let $\phi : \mathbb{R} \to \mathbb{R}$ be a L-Lipschitz disutility function. Let $oce^{\phi}(X)$ be the OCE risk of X as defined in* (8)*, and let oce_n^{ϕ} be its empirical estimate as in* (10)*. Then, for every $\epsilon > 0$ and $n \geq 1$, we have*

$$\begin{aligned}\mathbb{P}\left(\left|oce_n^{\phi} - oce^{\phi}(X)\right| > \epsilon\right) \leq c_1 &\left[\exp\left[-c_2 n(\epsilon/L)^2\right]\mathbb{I}\left\{\epsilon \leq L\right\}\right.\\ &\left.+ \exp\left[-c_3 n(\epsilon/L)^{\beta}\right]\mathbb{I}\left\{\epsilon > L\right\}\right],\end{aligned}$$

where the constants c_1, c_2 and c_3 are as in Lemma 3.

Proof See Section 10.4.3. ■

One can specialize the result above to handle the case of sub-Gaussian random r.v.s. by using $\beta = 2$ in the bound above. However, the constants c_1, c_2 are unknown functions of the parameter σ. The following result provides an alternative OCE concentration bound with explicit constants.

Corollary 29 (OCE concentration with explicit constants) *Suppose X is a sub-Gaussian r.v. with parameter σ, and let ϕ be as in Corollary 28. Then, for every $n \geq 1$ and ϵ such that $\frac{512\sigma}{\sqrt{n}} < \frac{\epsilon}{L} < \frac{512\sigma}{\sqrt{n}} + 16\sigma\sqrt{\mathsf{e}}$, we have*

$$\mathbb{P}\left(\left|oce_n^{\phi} - oce^{\phi}(X)\right| > \epsilon\right) \leq \exp\left(-\frac{n}{256\sigma^2\mathsf{e}}\left(\frac{\epsilon}{L} - \frac{512\sigma}{\sqrt{n}}\right)^2\right).$$

Proof See Section 10.4.4. ■

As discussed in subsection 3.2.1, CVaR is a special case of an OCE risk, and satisfies the assumptions of corollaries 28 and 29 with $L = (1-\alpha)^{-1}$. Hence concentration bounds for empirical CVaR follow from the preceding two results.

The case considered in Corollary 29, as applied to empirical CVaR, was also treated by Kolla et al. (2019); Prashanth et al. (2020), while the special case of bounded r.v.s has been treated in Brown (2007); Wang and Gao (2010). In terms of dependence on n and ϵ, the tail bound in Corollary 29 is better than the one-sided concentration bound in Kolla et al. (2019). In fact, the dependence on n and ϵ given in Corollary 29 matches that in the case of bounded distributions (Brown, 2007; Wang and Gao, 2010). More recently, Prashanth et al. (2020) have derived a two-sided concentration result for CVaR estimation. Their bound requires the knowledge of the density in a neighborhood of the true VaR, while the constants in our bounds depend only on the parameter σ of the underlying sub-Gaussian distribution. On the other hand, though our bounds do not depend on density-related information, they probably involve conservative constants. Finally, unlike Prashanth et al. (2020), our bounds allow a multi-armed bandit application in the spirit of the classic UCB algorithm (Auer et al., 2002), as we require only the knowledge of the sub-Gaussianity parameter in arriving at a confidence term. The reader is referred to Section 9 for more details.

6.1.2 BOUNDS FOR EMPIRICAL SPECTRAL RISK MEASURE

In this section, we restrict ourselves to a spectral risk measure M_ϕ whose associated risk spectrum ϕ is bounded. Specifically, we assume that $|\phi(\beta)| \leq K$ for all $\beta \in [0,1]$ for some $K > 0$. It immediately follows from Lemma 13 that, if X and Y are r.v.s with CDFs F_1 and F_2, then

$$|M_\phi(X) - M_\phi(Y)| \leq K W_1(F_1, F_2). \tag{25}$$

On noting from (14) that the empirical estimate $m_{n,\phi}$ of $M_\phi(X)$ is simply the spectral risk measure M_ϕ applied to a r.v. whose CDF is F_n, we conclude from (25) that

$$|M_\phi(X) - m_{n,\phi}| \leq K W_1(F, F_n). \tag{26}$$

Equation (26) relates the estimation error $|M_\phi(X) - m_{n,\phi}|$ to the Wasserstein distance between the true and empirical CDFs of X. As in the case of CVaR, invoking Lemma 3 provides concentration bounds for the empirical spectral risk measure estimate (14).

Corollary 30 (SRM concentration) *Suppose X satisfies (C1) for some $\beta > 1$. Let $K > 0$ and let $\phi : [0,1] \to [0,K]$ be a risk spectrum. Then, for every $\epsilon > 0$ and $n \geq 1$, we have*

$$\mathbb{P}\left(|m_{n,\phi} - M_\phi(X)| > \epsilon\right) \leq c_1\left[\exp\left[-c_2 n\left[\frac{\epsilon}{K}\right]^2\right]\mathbb{I}\left\{\epsilon \leq K\right\} + \exp\left[-c_3 n\left[\frac{\epsilon}{K}\right]^\beta\right]\mathbb{I}\left\{\epsilon > K\right\}\right],$$

where the constants c_1, c_2 and c_3 are as in Lemma 3.

Proof See Section 10.4.5. ■

As before, setting $\beta = 2$ in the bound above handles the special case of sub-Gaussian random r.v.s.

Next, we present an SRM concentration bound with explicit constants.

Corollary 31 (SRM concentration with explicit constants) *Let X be a sub-Gaussian r.v. with parameter σ, and let ϕ be a risk spectrum as in Corollary 30. Then, for every $n \geq 1$ and ϵ such that $\frac{512\sigma}{\sqrt{n}} < \frac{\epsilon}{K} < \frac{512\sigma}{\sqrt{n}} + 16\sigma\sqrt{\mathsf{e}}$, we have*

$$\mathbb{P}\left(|m_{n,\phi} - M_\phi(X)| > \epsilon\right) \leq \exp\left(-\frac{n}{256\sigma^2\mathsf{e}}\left(\frac{\epsilon}{K} - \frac{512\sigma}{\sqrt{n}}\right)^2\right).$$

Proof See Section 10.4.6. ■

The bound above is an improvement over the SRM concentration bound derived in Pandey et al. (2021) for two reasons. First, our bound is for the un-truncated estimator, while their bound involves a truncated estimate. Second, our bound applies for all $n \geq 1$, while their bound applies only for a sufficiently large number of samples.

6.1.3 BOUNDS FOR EMPIRICAL UTILITY-BASED SHORTFALL RISK

As in the case of OCE and spectral risk measure, using the fact that UBSR is (T1) leads to the following concentration bound.

Proposition 32 *Suppose X satisfies (C1) for some $\beta > 1$. Let the utility function l in the definition* (15) *of $S_\alpha(X)$ satisfy the assumptions in Lemma 15, and let $\xi_{n,\alpha}$ denote the solution to the constrained problem in* (18)*. Then, for every $\epsilon > 0$ and $n \geq 1$, we have*

$$\begin{aligned}\mathbb{P}\left(|\xi_{n,\alpha} - S_\alpha(X)| > \epsilon\right) \leq c_1 \Bigg[&\exp\left[-c_2 n\left[\frac{k\epsilon}{K}\right]^2\right]\mathbb{I}\left\{\epsilon \leq \frac{K}{k}\right\} \\ &+ \exp\left[-c_3 n\left[\frac{k\epsilon}{K}\right]^\beta\right]\mathbb{I}\left\{\epsilon > \frac{K}{k}\right\}\Bigg],\end{aligned}$$

where the constants c_1, c_2 and c_3 are as in Lemma 3, while $K, k > 0$ as are as in Lemma 15.

Proof See Section 10.4.7. ■

The specialization to sub-Gaussian r.v.s is immediate. Next, we present an UBSR concentration bound with explicit constants.

Corollary 33 *Let X be a sub-Gaussian r.v. with parameter σ, and let l be a utility function as in Proposition 32. Then, for every $n \geq 1$ and ϵ such that $\frac{512\sigma}{\sqrt{n}} < \frac{\epsilon k}{K} < \frac{512\sigma}{\sqrt{n}} + 16\sigma\sqrt{\mathsf{e}}$, we have*

$$\mathbb{P}\left(|\xi_{n,\alpha} - S_\alpha(X)| > \epsilon\right) \leq \exp\left(-\frac{n}{256\sigma^2\mathsf{e}}\left(\frac{k\epsilon}{K} - \frac{512\sigma}{\sqrt{n}}\right)^2\right).$$

Proof See Section 10.4.8. ■

6.2 Bounds for Risk Measures of Type (T2)

We consider two cases in this subsection. The first case is when the distribution F has bounded support, while the second case is that of sub-Gaussian distributions. Although the first case is subsumed under the case of sub-Gaussian distributions, we still consider it separately as the concentration bound that we obtain is stronger than the one in the sub-Gaussian case.

6.2.1 DISTRIBUTIONS WITH BOUNDED SUPPORT

Our first result concerns the risk estimator (5) for a Type (T2) risk measure applied to a distribution with bounded support.

Theorem 34 *Suppose X takes values in $[-B_2, B_1]$ a.s. for some $B_1, B_2 \geq 0$ such that at least one of B_1, B_2 is positive. Let $\rho : \mathcal{L} \to \mathbb{R}$ be a risk measure of type (T2). For each n, let $\rho_n = \rho(F_n)$, where F is the CDF of X. Then, for every $\epsilon > 0$ and $n \geq 1$, we have*

$$\mathbb{P}\left(|\rho_n - \rho(X)| > \epsilon\right) \leq c_1 \exp\left(-c_2 n \left[\frac{\epsilon}{L_1 \tau^\gamma}\right]^{\frac{2}{\alpha_1}}\right),$$

where $\tau = \max\left\{\frac{B_1}{K_1}, \frac{B_2}{K_2}\right\}$, $K_1, K_2, L_1, \gamma, \alpha_1$ are as in the definition (19) *of a Type (T2) risk measure, and c_1, c_2 are constants that depend on B_1, B_2.*

Proof See Section 10.4.9. ■

As mentioned earlier, the constants c_1, c_2 are not explicitly known, and this lack of knowledge hinders bandit applications. To handle such an application, we next present a concentration bound with explicit constants.

Theorem 35 *Assume that the conditions of Theorem 34 hold. Then, for every $n \geq 1$ and ϵ such that $\frac{256(B_1+B_2)}{\sqrt{n}} < \left(\frac{\epsilon}{L_1\tau^\gamma}\right)^{\frac{1}{\alpha_1}} < \frac{256(B_1+B_2)}{\sqrt{n}} + 8(B_1 + B_2)\sqrt{\mathsf{e}}$, we have*

$$\mathbb{P}\left(|\rho_n - \rho(X)| > \epsilon\right) \leq \exp\left(-\frac{n}{64\mathsf{e}(B_1 + B_2)^2}\left(\left(\frac{\epsilon}{L_1\tau^\gamma}\right)^{\frac{1}{\alpha_1}} - \frac{256(B_1 + B_2)}{\sqrt{n}}\right)^2\right),$$

where L_1, γ and τ are as specified in Theorem 34.

Proof See Section 10.4.10. ■

We now turn our attention to deriving a concentration bound for the CPT estimator in (22) for the case of a bounded r.v. To put things in context, in Cheng et al. (2018), the authors derive a concentration bound for the same estimator assuming that the underlying distribution has bounded support, and for this purpose, they employ the DKW theorem. Interestingly, we are able to provide a matching bound for the case of distributions with bounded support, using a proof technique that relates the the estimation error $|C_n - C(X)|$ to the Wasserstein distance between the empirical and true CDF, and this is the content of the proposition below.

Corollary 36 ***(CPT concentration for bounded r.v.s)*** *Suppose X is a r.v. that assumes values in $[-B_2, B_1]$ a.s., where $B_1, B_2 \geq 0$, and at least one of B_1, B_2 is positive. Further, assume that the conditions of Lemma 16 on the CPT value defined by* (21) *hold. Then, for every $\epsilon > 0$ and $n \geq 1$, we have*

$$\mathbb{P}\left(|C_n - C(X)| > \epsilon\right) \leq c_1 \exp\left(-c_2 n \left[\frac{\epsilon}{L\tau^{1-\alpha}}\right]^{\frac{2}{\alpha}}\right),$$

where $\tau = \max\left\{B_1 \frac{K^+}{k^+}, B_2 \frac{K^-}{k^-}\right\}$, the constants L, K^+, k^+, K^-, k^- and α are as in Lemma 16, and c_1, and c_2 are as in Theorem 34.

Proof See Section 10.4.11. ∎

It is apparent from the bound above that, given $\delta \in (0,1)$, if the number of samples n is of the order $O\left(\frac{1}{\epsilon^{2/\alpha}} \log\left(\frac{1}{\delta}\right)\right)$, then $|C_n - C(X)| < \epsilon$ with probability $1 - \delta$.

Next, we provide a CPT concentration bound with explicit constants.

Corollary 37 ***(CPT concentration for bounded r.v.s with explicit constants)*** *Under the conditions of Corollary 36, for every $n \geq 1$ and ϵ such that $\frac{256(B_1+B_2)}{\sqrt{n}} < \left(\frac{\epsilon}{L\tau^{1-\alpha}}\right)^{\frac{1}{\alpha}} < \frac{256(B_1+B_2)}{\sqrt{n}} + 8(B_1 + B_2)\sqrt{\mathsf{e}}$, we have*

$$\mathbb{P}\left(|C_n - C(X)| > \epsilon\right) \leq \exp\left(-\frac{n}{64\mathsf{e}(B_1+B_2)^2}\left(\left(\frac{\epsilon}{L\tau^{1-\alpha}}\right)^{\frac{1}{\alpha}} - \frac{256(B_1+B_2)}{\sqrt{n}}\right)^2\right),$$

where τ, α and L are as in Corollary 36.

Proof See Section 10.4.12. ∎

6.2.2 DISTRIBUTIONS SATISFYING (C1)

Next, we provide a concentration bound for type (T2) risk measures in the case of r.v.s that satisfy (C1).

Theorem 38 *Let X be a r.v. with a distribution F that satisfies (C1) for some $\beta > 1$, and suppose $\rho : \mathcal{L} \to \mathbb{R}$ is a risk measure of type (T2) with parameters $\alpha_1, \alpha_2, \alpha_3, L_1, L_2, L_3, K_1, K_2$ as defined in* (19)*. Fix $\tau > 0$ and let $\rho_{n,\tau} = \rho(F_n|_\tau)$. Fix $\epsilon > 0$ such that*

$$\begin{aligned}\epsilon' = \epsilon &- \frac{L_2}{(K_1\tau)^{\beta-1}\,\alpha_2\gamma(\beta-1)} \exp\left(-\alpha_2\gamma\,(K_1\tau)^{\beta}\right)\\ &- \frac{L_3}{(K_2\tau)^{\beta-1}\,\alpha_3\gamma(\beta-1)} \exp\left(-\alpha_3\gamma\,(K_2\tau)^{\beta}\right)\end{aligned} \tag{27}$$

is positive. Then, for every $n \geq 1$, we have

$$\mathbb{P}\left(|\rho_{n,\tau} - \rho(X)| > \epsilon\right) \leq c_1 \exp\left(-c_2 n \left(\frac{\epsilon'}{L_1\tau^{\gamma}}\right)^{\frac{2}{\alpha_1}}\right), \tag{28}$$

where c_1 and c_2 are constants that depend on the parameters β, γ and $\top$ specified in (C1).

Proof See Section 10.4.13. ■

Next, we apply Theorem 38 to obtain a CPT concentration result for a r.v. satisfying (C1). For this result, we consider the CPT value estimator based on truncation defined in (22). The cut-off value for truncation is chosen as a function of the sample size to get an exponential decay in the tail bound.

Corollary 39 ***(CPT concentration)*** *Assume that the conditions in Lemma 16 hold, and suppose that X satisfies (C1) for some $\beta > 1$. For each $n \geq 1$, set*

$$\tau_n = \left((\log n)^{\frac{1}{\beta}} + 1 \right) \max \left\{ \frac{K^+}{k^+}, \frac{K^-}{k^-} \right\},$$

$c_3(n) = \frac{L(K^+ + K^-)}{(\log n)^{\frac{\beta-1}{\beta}} \alpha(1-\alpha)(\beta-1) n^{\alpha(1-\alpha)}}$, *and form the CPT-value estimate C_n using* (22)*, where L, K^+, k^+, K^-, k^- and α are as in Lemma 16. Then, for every $n \geq 1$ and $\epsilon > c_3(n)$, we have*

$$\mathbb{P}\left(|C_n - C(X)| > \epsilon \right) \leq c_1 \exp\left(-c_2 n \left(\frac{\epsilon - c_3(n)}{L(K^+ + K^-)\tau_n^{1-\alpha}} \right)^{\frac{2}{\alpha}} \right),$$

where c_1, c_2 are constants that depend on the parameters β, γ and $\top$ specified in (C1).

Proof See Section 10.4.14. ■

In Proposition 3 of Cheng et al. (2018), the authors provide a $\left[2n\mathsf{e}^{-n^{\frac{\alpha}{2+\alpha}}} + 2\mathsf{e}^{-n^{\frac{\alpha}{2+\alpha}} \left(\frac{\epsilon}{2H} \right)^{\frac{2}{\alpha}}} \right]$ bound for CPT-estimation in the case where the underlying distribution is sub-Gaussian. It is apparent that the bound we obtain with $\beta = 2$ in the theorem above is significantly improved in comparison to the bound of Cheng et al. (2018).

In Bhat and Prashanth (2019), a tail bound for CPT-value estimation is presented for the sub-Gaussian case. In comparison, the bound we have in Corollary 39 applies to the more general class of distributions satisfying (C1). More importantly, our tail bound is applicable for all $n \geq 1$, while the bound in Bhat and Prashanth (2019) applies only when the number of samples n is sufficiently large.

Our next result is a variation on Corollary 39, where we specify the constants, but under the slighly more restrictive assumption of sub-Gaussianity.

Proposition 40 ***(CPT concentration bound with explicit constants for sub-Gaussian r.v.s)*** *Assume that the conditions in Lemma 16 hold, and suppose that X is sub-Gaussian with parameter σ. For each $n \geq 1$, choose τ_n and $c_3(n)$ as given in Corollary 39 by setting $\beta = 2$, and form the CPT-value estimate C_n using* (22)*, where L, K^+, k^+, K^-, k^- and α are as in Lemma 16. Then, for every $n \geq 1$ and ϵ such that $\frac{512\sigma}{\sqrt{n}} < \left(\frac{\epsilon - c_3(n)}{c_4(n)} \right)^{\frac{1}{\alpha}} < \frac{512\sigma}{\sqrt{n}} + 16\sigma\sqrt{\mathsf{e}}$, we have*

$$\mathbb{P}\left(|C_n - C(X)| > \epsilon \right) \leq \exp\left(-\frac{n}{256\sigma^2\mathsf{e}} \left(\left(\frac{\epsilon - c_3(n)}{c_4(n)} \right)^{\frac{1}{\alpha}} - \frac{512\sigma}{\sqrt{n}} \right)^2 \right),$$

where $c_4(n) = L(K^+ + K^-)\tau_n^{1-\alpha}$.

Proof See Section 10.4.15. ■

7. Concentration bounds for distributions satisfying (C2)

In this section, we present concentration bounds for estimators of the risk measures considered so far under the assumption that the underlying distribution is sub-exponential. First, we consider risk measures of type (T1).

7.1 Bounds for Risk Measures of Type (T1)

Using Lemma 10, we provide an analogue of Theorem 25 for the case of sub-exponential r.v.s below.

Theorem 41 *Let X be a r.v. satisfying (C2) with parameter c. Suppose $\rho : \mathcal{L} \to \mathbb{R}$ is a risk measure of type (T1), with parameters L and κ. Then, for every $n \geq 1$ and ϵ satisfying $\left(\frac{\epsilon}{L}\right)^{\frac{1}{\kappa}} > \frac{384}{c\sqrt{n}}$, we have*

$$\mathbb{P}\left(|\rho_n - \rho(X)| > \epsilon\right) \leq \exp\left(-\frac{n}{\frac{32}{c^2} + \frac{4}{c}\left(\left(\frac{\epsilon}{L}\right)^{\frac{1}{\kappa}} - \frac{384}{c\sqrt{n}}\right)}\left(\left(\frac{\epsilon}{L}\right)^{\frac{1}{\kappa}} - \frac{384}{c\sqrt{n}}\right)^2\right).$$

Proof See Section 10.5.1. ■

The following corollaries regarding concentration of empirical OCE, SRM and UBSR follow in a straightforward fashion using the result above. We omit the proof details.

Corollary 42 (OCE concentration) *Let X be a r.v. satisfying (C2) with parameter c, and let $\phi : \mathbb{R} \to \mathbb{R}$ be a L-Lipschitz disutility function. Let $oce^{\phi}(X)$ be the OCE risk of X as defined in* (8)*, and let oce_n^{ϕ} be its empirical estimate as in* (10)*. Then, for every $n \geq 1$ and every ϵ satisfying $\frac{\epsilon}{L} > \frac{384}{c\sqrt{n}}$, we have*

$$\mathbb{P}\left(\left|oce_n^{\phi} - oce^{\phi}(X)\right| > \epsilon\right) \leq \exp\left(-\frac{n}{\frac{32}{c^2} + \frac{4}{c}\left(\frac{\epsilon}{L} - \frac{384}{c\sqrt{n}}\right)}\left(\frac{\epsilon}{L} - \frac{384}{c\sqrt{n}}\right)^2\right).$$

Recall that CVaR is an OCE risk satisfying (12). It is therefore clear that the bound above holds for empirical CVaR with $L = (1-\alpha)^{-1}$.

Corollary 43 (SRM concentration) *Let X be a r.v. satisfying (C2) with parameter c. Let $K > 0$ and let $\phi : [0,1] \to [0,K]$ be a risk spectrum. Then, for all $n \geq 1$ and every ϵ satisfying $\frac{\epsilon}{K} > \frac{384}{c\sqrt{n}}$, we have*

$$\mathbb{P}\left(|m_{n,\phi} - M_{\phi}(X)| > \epsilon\right) \leq \exp\left(-\frac{n}{\frac{32}{c^2} + \frac{4}{c}\left(\frac{\epsilon}{K} - \frac{384}{c\sqrt{n}}\right)}\left(\frac{\epsilon}{K} - \frac{384}{c\sqrt{n}}\right)^2\right).$$

Corollary 44 (UBSR concentration) *Let X be a r.v. satisfying (C2) with parameter c. Let the utility function l in the definition* (15) *of $S_\alpha(X)$ satisfy the assumptions of Lemma 15. For each $n \geq 1$, let $\xi_{n,\alpha}$ denote the solution to the constrained problem in* (18)*. Then, for all $n \geq 1$ and every ϵ satisfying $\frac{k\epsilon}{K} > \frac{384}{c\sqrt{n}}$, we have*

$$\mathbb{P}\left(|\xi_{n,\alpha} - S_\alpha(X)| > \epsilon\right) \leq \exp\left(-\frac{n}{\frac{32}{c^2} + \frac{4}{c}\left(\frac{k\epsilon}{K} - \frac{384}{c\sqrt{n}}\right)}\left(\frac{k\epsilon}{K} - \frac{384}{c\sqrt{n}}\right)^2\right),$$

where the constants $K, k > 0$ are as in Lemma 15.

7.2 Bounds for Risk Measures of Type (T2)

We now provide an analogue of Theorem 38 for the case of sub-exponential r.v.s below.

Theorem 45 *Let X be a r.v. having CDF F and satisfying (C2) with parameter c. Suppose $\rho : \mathcal{L} \to \mathbb{R}$ is a risk measure of type (T2) with parameters $\alpha_1, \alpha_2, \alpha_3, L_1, L_2, L_3, K_1, K_2$ as defined in* (19)*. Fix $\tau > 0$, and let $\rho_n = \rho(F_n|_\tau)$. Fix $\epsilon > 0$ such that $\epsilon' \triangleq \epsilon - \frac{L_2}{c\alpha_2}\exp\left(-\alpha_2 c K_1 \tau\right) - \frac{L_3}{c\alpha_3}\exp\left(-\alpha_3 c K_2 \tau\right) > 0$ and $\left(\frac{\epsilon'}{L_1\tau^\gamma}\right)^{\frac{1}{\alpha_1}} > \frac{384}{c\sqrt{n}}$. Then, for every $n \geq 1$, we have*

$$\mathbb{P}\left(|\rho_n - \rho(X)| > \epsilon\right) \leq \exp\left(-\frac{n}{\frac{32}{c^2} + \frac{4}{c}\left(\left(\frac{\epsilon'}{L_1\tau^\gamma}\right)^{\frac{1}{\alpha_1}} - \frac{384}{c\sqrt{n}}\right)}\left(\left(\frac{\epsilon'}{L_1\tau^\gamma}\right)^{\frac{1}{\alpha_1}} - \frac{384}{c\sqrt{n}}\right)^2\right). \tag{29}$$

Proof See Section 10.5.2. ∎

Finally, we provide a concentration bound for CPT estimation, when the underlying distribution is sub-exponential.

Corollary 46 ***(CPT concentration for sub-exponential r.v.s)*** *Assume that the conditions of Lemma 16 hold. Let X be a r.v. satisfying (C2) with parameter c. For each $n \geq 1$, set*

$$\tau_n = \left(\frac{\log n}{c} + 1\right)\max\left\{\frac{K^+}{k^+}, \frac{K^-}{k^-}\right\},$$

and form the CPT-value estimate C_n using (22)*. Then,*

$$\mathbb{P}\left(|C_n - C(X)| > \epsilon\right) \leq$$

$$\exp\left(-\frac{n}{\frac{32}{c^2} + \frac{4}{c}\left(\left(\frac{\epsilon - \frac{(K^+ + K^-)L}{c\alpha n^\alpha}}{(K^+ + K^-)L\tau_n^{1-\alpha}}\right)^{\frac{1}{\alpha}} - \frac{384}{c\sqrt{n}}\right)}\left(\left(\frac{\epsilon - \frac{(K^+ + K^-)L}{c\alpha n^\alpha}}{(K^+ + K^-)L\tau_n^{1-\alpha}}\right)^{\frac{1}{\alpha}} - \frac{384}{c\sqrt{n}}\right)^2\right)$$

holds for every $\epsilon > \frac{(K^+ + K^-)L}{c\alpha n^\alpha}$ satisfying

$$\left(\frac{\epsilon - \frac{(K^+ + K^-)L}{c\alpha n^\alpha}}{(K^+ + K^-)L\tau_n^{1-\alpha}}\right)^{\frac{1}{\alpha}} > \frac{384}{c\sqrt{n}}.$$

Proof See Section 10.5.3. ■

8. Concentration bounds for distributions satisfying (C3)

In this section, we consider heavy-tailed distributions that satisfy (C3), i.e., a higher moment bound $\mathbb{E}\left(|X|^\beta\right) < \top < \infty$ for some $\beta > 2$.

We consider risk measures of type (T1), and derive concentration bounds for the empirical estimate of such a risk measure. Using Lemma 11, we provide below an analogue of Theorem 25 for the case of r.v.s satisfying (C3).

Theorem 47 *Suppose X is a r.v. that satisfies (C3) with parameter β, and $\rho : \mathcal{L} \to \mathbb{R}$ is a risk measure of type (T1) with parameters L and κ. Then, for every $\epsilon > 0$, $n \geq 1$ and $\eta \in (0, \beta)$, we have*

$$\mathbb{P}\left(|\rho_n - \rho(X)| > \epsilon\right) \leq c_1 \left(\exp\left(-c_2 n \left(\frac{\epsilon}{L}\right)^{\frac{2}{\kappa}} \right) \mathbb{I}\left\{\epsilon \leq L\right\} + n \left(n \left(\frac{\epsilon}{L}\right)^{\frac{1}{\kappa}} \right)^{-(\beta-\eta)} \mathbb{I}\left\{\epsilon > L\right\} \right),$$

where the constants c_1, c_2 are as in Lemma 11.

Proof See Section 10.6.1. ■

The following corollaries on concentration of empirical CVaR, SRM and UBSR follow in a straightforward fashion from the result above, and we omit the proof details.

Corollary 48 (OCE concentration) *Suppose X is a r.v. that satisfies (C3) with parameter β. Let $\phi : \mathbb{R} \to \mathbb{R}$ be a L-Lipschitz disutility function. Let $oce^\phi(X)$ be the OCE risk of X as defined in* (8)*, and let oce_n^ϕ be its empirical estimate as in* (10)*. Then, for every $n \geq 1$, $\epsilon > 0$ and $\eta \in (0, \beta)$, we have*

$$\begin{aligned}\mathbb{P}\left(\left|oce_n^\phi - oce^\phi(X)\right| > \epsilon\right) \leq c_1 \Big[&\exp\left[-c_2 n (\epsilon/L)^2\right] \mathbb{I}\left\{\epsilon \leq L\right\} \\ &+ n\left(n\epsilon/L\right)^{-(\beta-\eta)} \mathbb{I}\left\{\epsilon > L\right\} \Big],\end{aligned}$$

where the constants c_1 and c_2 are as in Lemma 11.

In light of the fact that CVaR is an OCE risk satisfying (12), it is clear that the bound above holds for empirical CVaR with $L = (1-\alpha)^{-1}$.

Corollary 49 (SRM concentration) *Suppose X is a r.v. that satisfies (C3) with parameter β. Let $K > 0$ and let $\phi : [0, 1] \to [0, K]$ be a risk spectrum. Then, for every $n \geq 1$, $\epsilon > 0$ and $\eta \in (0, \beta)$, we have*

$$\mathbb{P}\left(|m_{n,\phi} - M_\phi(X)| > \epsilon\right) \leq c_1 \left[\exp\left(-c_2 n \left[\frac{\epsilon}{K}\right]^2 \right) \mathbb{I}\left\{\epsilon \leq K\right\} + n \left(n \left[\frac{\epsilon}{K}\right] \right)^{-(\beta-\eta)} \mathbb{I}\left\{\epsilon > K\right\} \right],$$

where the constants c_1 and c_2 are as in Lemma 11.

Corollary 50 (UBSR concentration) *Suppose X is a r.v. that satisfies (C3) with parameter β. Let the utility function l in the definition* (15) *of $S_\alpha(X)$ satisfy the assumptions in Lemma 15. For every $n \geq 1$, let $\xi_{n,\alpha}$ denote the solution to the constrained problem in* (18)*. Then, for every $n \geq 1$, $\epsilon > 0$ and $\eta \in (0, \beta)$, we have*

$$\mathbb{P}\left(|\xi_{n,\alpha} - S_\alpha(X)| > \epsilon\right) \leq c_1 \left[\exp\left(-c_2 n \left[\frac{k\epsilon}{K}\right]^2\right) \mathbb{I}\left\{\epsilon \leq \frac{K}{k}\right\} + n\left(n\left[\frac{k\epsilon}{K}\right]\right)^{-(\beta-\eta)} \mathbb{I}\left\{\epsilon > \frac{K}{k}\right\}\right],$$

where the constants c_1 and c_2 are as in Lemma 11 and $K, k > 0$ are as in Lemma 15.

For small deviations, i.e., $\epsilon \leq 1$, the bounds presented above are satisfactory, as the tail decay matches that of a Gaussian r.v. with constant variance. On the other hand, for large ϵ, the second term exhibits polynomial decay. The latter polynomial term is not an artifact of our analysis. Instead, it relates to the rate obtained in Lemma 11. It would be an interesting research direction to investigate if the rate improves in the large ϵ case by employing the truncated estimator (6).

We have not presented bounds for empirical risk estimates of type (T2) risk measures, when the underlying distribution satisfies (C3). It would be an interesting direction of future research to fill this gap.

9. Application: Risk-sensitive bandits

The concentration bounds for (T1) and (T2) risk measures in previous sections open avenues for bandit applications. We illustrate this claim by using the regret minimization framework in a stochastic K-armed bandit problem, with an objective based on an abstract risk measure. Our algorithm can be used as a template for risk-sensitive bandits, where the notion of risk could be any of the five risk measures discussed earlier namely, OCE (including CVaR as a special case), spectral risk measure, UBSR, CPT (including DRM as a special case) and RDEU.

9.1 Risk-sensitive bandit problem

We are given K arms with unknown distributions $P_i, i = 1, \ldots, K$. The interaction of the bandit algorithm with the environment proceeds, over n rounds, as follows: (i) At round t, select an arm $I_t \in \{1, \ldots, K\}$; (ii) Observe a sample cost from the distribution P_{I_t} corresponding to the arm I_t.

Let ρ be a risk measure, and let $\rho(i) = \rho(P_i)$ denote the risk associated with arm i, for $i = 1, \ldots, K$. Let $\rho_* = \min_{i=1,\ldots,K} \rho(i)$ denote the lowest risk among the K distributions, and $\Delta_i = (\rho(i) - \rho_*)$ denote the gap in risk values of arm i and that of the best arm.

The classic objective in a bandit problem is to find the arm with the lowest expected value. We consider an alternative formulation, where the goal is to find the arm with the lowest risk. Using the notion of regret, this objective is formalized as follows:

$$R_n = \sum_{i=1}^{K} \rho(i) T_i(n) - n\rho_* = \sum_{i=1}^{K} T_i(n)\Delta_i,$$

where $T_i(n) = \sum_{t=1}^{n} \mathbb{I}\{I_t = i\}$ is the number of pulls of arm i up to time instant n. The regret definition above is in the spirit of those for CVaR and CPT-sensitive bandits in Galichet (2015) and Gopalan et al. (2017), respectively.

9.2 (T1) risk measures with sub-Gaussian arms

We present a straightforward adaptation Risk-LCB of the well-known UCB algorithm (Auer et al., 2002) to handle an objective based on the abstract risk measure ρ. The algorithm caters to (T1) risk measures and arms' distributions that are sub-Gaussian with common parameter σ. The relevant concentration bound for the former case is in Theorem 27, which we recall below.

$$\mathbb{P}\left(|\rho_m - \rho(X)| > \epsilon\right) \leq \exp\left(-\frac{m}{256\sigma^2 \mathsf{e}}\left(\left(\frac{\epsilon}{L}\right)^{\frac{1}{\kappa}} - \frac{512\sigma}{\sqrt{m}}\right)^2\right), \tag{30}$$

where ρ_m, which is formed using (5), is an m-sample estimate of the risk measure $\rho(X)$, and L and κ come from the defining inequality (7) of a (T1) risk measure. The tail bound above holds for ϵ satisfying the constraint given by

$$\frac{512\sigma}{\sqrt{m}} < \left(\frac{\epsilon}{L}\right)^{\frac{1}{\kappa}} < \frac{512\sigma}{\sqrt{m}} + 16\sigma\sqrt{\mathsf{e}}. \tag{31}$$

Simple algebraic manipulations show that, for every $\delta \in (\exp(-m), 1)$, ϵ defined by

$$\epsilon = L\left[\left(\frac{256\sigma^2 \mathsf{e} \log(\frac{1}{\delta})}{m}\right)^{\frac{1}{2}} + \frac{512\sigma}{\sqrt{m}}\right]^{\kappa}$$

satisfies the constraint in (31). This allows us to rewrite the tail bound for a (T1) risk measure from Theorem 27 in the high confidence form as

$$\mathbb{P}\left(|\rho_m - \rho(X)| \leq L\left[\left(\frac{256\sigma^2 \mathsf{e} \log(\frac{1}{\delta})}{m}\right)^{\frac{1}{2}} + \frac{512\sigma}{\sqrt{m}}\right]^{\kappa}\right) \geq 1 - \delta, \text{ for every } \delta \in (\mathsf{e}^{-m}, 1) \tag{32}$$

For the classic bandit setup with a expected value objective, the finite sample analysis provided by Auer et al. (2002) chose to set $\delta = \frac{1}{t^4}$ for the tth round in a tail bound based on Hoeffding's inequality. This choice of δ was shown to be good enough to guarantee a sub-linear regret in (Auer et al., 2002). However, unlike (31), Hoeffding's inequality, which forms the basis for the UCB definition in a risk-neutral bandit setting, does not have a constraint on ϵ. In our setting, for an arm i that is pulled $T_i(t-1)$ times up to round t, we require $\delta \in (\mathsf{e}^{-T_i(t-1)}, 1)$ for using the tail bound (30). Setting $\delta = \frac{8}{t^4}$ and using the fact that $T_i(t-1) \leq t$, it is easy to see that this choice of δ satisfies the necessary constraint coming from (32). Under this choice of δ for the risk-sensitive bandit problem that we consider, in any round t of Risk-LCB, we have the following high-confidence guarantee for any arm $k \in \{1, \ldots, K\}$:

$$\mathbb{P}\left(\rho(i) \in [\rho_{i,T_i(t-1)} - w_{i,T_i(t-1)}, \rho_{i,T_i(t-1)} + w_{i,T_i(t-1)}]\right) \geq 1 - \frac{8}{t^4}, \text{ where}$$
$$w_{i,T_i(t-1)} = L\sigma\left[\frac{32\sqrt{\mathsf{e}\log(t)} + 512}{\sqrt{T_i(t-1)}}\right]^{\kappa}. \tag{33}$$

where, $\rho_{i,T_i(t-1)}$ is the estimate of the risk measure for arm i computed using (5) from $T_i(t-1)$ samples, and $w_{i,T_i(t-1)}$ is the confidence width. In arriving at the form for the confidence width, we have ignored a subtractive factor involving $\log 8$ that arises from taking the logarithm of $\delta = \frac{8}{t^4}$. Ignoring such a factor does not affect the high-confidence guarantee since we have increased the size of the confidence interval through this simplification.

Risk-LCB algorithm

Initialization: Play each arm once.
For $t = K+1, \ldots, n$**, repeat**

1. For each arm $i = 1, \ldots, K$, define
$$\mathrm{LCB}_t(i) = \rho_{i,T_i(t-1)} - w_{i,T_i(t-1)},$$
where $\rho_{i,T_i(t-1)}$ is the estimate of the risk measure for arm i computed using (5) from $T_i(t-1)$ samples, and $w_{i,T_i(t-1)}$ is defined in (33).
2. Play arm $I_t = \underset{i=1,\ldots,K}{\arg\min} \, \mathrm{LCB}_t(i)$.
3. Observe sample X_t from the distribution P_{I_t} corresponding to the arm I_t.

The result below bounds the regret of Risk-LCB algorithm, and the proof is a straightforward adaptation of that used to establish the regret bound of the regular UCB algorithm in Auer et al. (2002).

Theorem 51 *Consider a K-armed stochastic bandit problem with a risk measure ρ that is (T1) with parameters L and κ. Assume that the arms' distributions are sub-Gaussian with a common parameter σ. Then the expected regret of Risk-LCB at the end of $n \geq 1$ rounds satisfies*

$$\mathbb{E}(R_n) \leq \sum_{\{i:\Delta_i>0\}} \left(\frac{\sigma^2 (32\sqrt{\mathsf{e}\log n} + 512)^2 \, (2L)^{\frac{2}{\kappa}}}{\Delta_i^{\frac{2}{\kappa}-1}} + \left(1 + \frac{8\pi^2}{3}\right) \Delta_i \right).$$

Further, R_n also satisfies the following bound that does not scale inversely with the gaps:

$$\mathbb{E}(R_n) \leq \left(K\sigma^2 (32\sqrt{\mathsf{e}\log n} + 512)^2 (2L)^{\frac{2}{\kappa}} + \left(1 + \frac{8\pi^2}{3}\right) \sum_{\{i:\Delta_i>0\}} \Delta_i^{2/\kappa} \right)^{\frac{\kappa}{2}} n^{\frac{2-\kappa}{2}}.$$

Proof See Section 11.1. ■

For the case of OCE, SRM and UBSR, we have $\kappa = 1$. Thus, the regret bound obtained above matches, up to log factors, that of a classic K-armed bandit problem w.r.t. the dependence on the underlying gaps, and the horizon n.

A UCB-type algorithm for optimization of risk measures has been proposed earlier in Cassel et al. (2018). In comparison to the bound in the result above, the regret bound in Cassel et al. (2018) exhibits a sub-optimal dependence on the underlying gaps. In particular, the bound scales inversely with $\min(1, \Delta)$, where Δ is the smallest gap. In contrast, the bound we derive for a UCB-type algorithm scales inversely with the smallest gap. Our bound is thus better for problems with gaps bounded below by one.

CVaR optimization has been considered in a bandit setting in the literature. For instance, in Galichet (2015), the authors assume that the underlying arms' distributions have bounded support, and propose a UCB-type algorithm. We relax this assumption, and consider the case of sub-Gaussian distributions for the K arms. The tail bounds in Kolla et al. (2019) and Prashanth et al. (2020) do not allow a bandit application, because forming the confidence term (required for UCB-type algorithms) using their bound would require knowledge of the density in a neighborhood of the true VaR. In contrast, the constants in our bounds depend only on the sub-Gaussian parameter σ, and several classic MAB algorithms (including UCB) assume this information. More recently, in Baudry et al. (2021), the authors propose and analyze a Thompson sampling-based algorithm for CVaR, under the assumption that the arms' distributions have bounded support.

9.3 CPT risk measure with sub-Gaussian arms

We now consider the case of CPT, which is a prominent (T2) risk measure, in conjunction with sub-Gaussian arms. CPT-based bandits have been considered in Gopalan et al. (2017) for the case of arms' distributions with bounded support, and an UCB-based algorithm has been proposed therein. For handling the case of CPT-value in a bandit context, we propose a straightforward variant of the Risk-LCB algorithm presented earlier. Our algorithm can be seen as a generalization of the scheme in Gopalan et al. (2017), as it can handle arms' distributions that are sub-Gaussian.

The overall algorithm follows the template in Risk-LCB with the only modifications being to the confidence widths used in defining the LCBs for each arm. For defining the confidence widths, we start with the concentration bound in Proposition 40 after recalling that the proposition applies to sub-Gaussian r.v.s with $\beta = 2$. To write this bound in the high confidence form, note that for every $m \geq 1$ and $\delta \in (\exp(-m), 1)$, ϵ defined by

$$\epsilon = \left[L(K^+ + K^-)\tau_m^{1-\alpha}\right] \left[\left(\frac{256\sigma^2 \mathsf{e} \log(\frac{1}{\delta})}{m}\right)^{\frac{1}{2}} + \frac{512\sigma}{\sqrt{m}}\right]^{\alpha} + \frac{(K^+ + K^-)L}{\alpha(1-\alpha)n^{\alpha(1-\alpha)}\sqrt{\log m}}$$

satisfies the constraint $\frac{512\sigma}{\sqrt{m}} < \left(\frac{\epsilon - c_3(m)}{c_4(m)}\right)^{\frac{1}{\alpha}} < \frac{512\sigma}{\sqrt{m}} + 16\sigma\sqrt{\mathsf{e}}$ appearing in Proposition 40, where $\tau_m, L, K^+, K^-, \alpha, c_3(m)$ and $c_4(m)$ are as defined in that proposition and σ is the sub-Gaussianity parameter. This observation along with the fact that $\frac{1}{\sqrt{\log m}} < 2$ for all $m \geq 2$ allows us to rewrite the bound in Proposition 40 in the high confidence form as follows: for every $\delta \in (\mathsf{e}^{-m}, 1)$

$$\mathbb{P}\left(|C_m - C(X)| \leq \left[L(K^+ + K^-)\tau_m^{1-\alpha}\right] \times \left[\left(\frac{256\sigma^2 \mathsf{e} \log(\frac{1}{\delta})}{m}\right)^{\frac{1}{2}} + \frac{512\sigma}{\sqrt{m}}\right]^{\alpha} + \frac{2(K^+ + K^-)L}{\alpha(1-\alpha)n^{\alpha(1-\alpha)}}\right) \geq 1 - \delta,$$

where C_n is the m-sample estimate of the CPT-value $C(X)$ formed using (22).

Let $X_1, \ldots, X_K$ denote the r.v.s corresponding to arms $1, \ldots, K$, respectively. Assume that $X_i, i = 1, \ldots, K$ are sub-Gaussian with parameter σ. Then, as in Section 9.2, the choice $\delta = \frac{8}{t^4}$ is contained in $(\exp(-t), 1)$, and for this choice, we have the following high-confidence guarantee for any arm $k \in \{1, \ldots, K\}$ at any round t of CPT-LCB:

$$\mathbb{P}\left(C(X_i) \in [C_{i,T_i(t-1)} - w_{i,T_i(t-1)}, C_{i,T_i(t-1)} + w_{i,T_i(t-1)}]\right) \geq 1 - \frac{8}{t^4}, \text{ where}$$

$$w_{i,T_i(t-1)} = L(K^+ + K^-)\left[\max\left\{\frac{K^+}{k^+}, \frac{K^-}{k^-}\right\}\left(\sqrt{\log T_i(t-1)} + 1\right)\right]^{1-\alpha} \times \left[\frac{\sigma(32\sqrt{\mathsf{e}\log t} + 512)}{\sqrt{T_i(t-1)}}\right]^{\alpha} + \frac{2(K^+ + K^-)L}{\alpha(1-\alpha)T_i(t-1)^{\alpha(1-\alpha)}}, \tag{34}$$

where $C_{i,T_i(t-1)}$ is the estimate of the CPT-value for arm i computed using (22) from $T_i(t-1)$ samples. This high-confidence guarantee can be used to prove the following theorem.

Theorem 52 *Consider a K-armed stochastic bandit problem with CPT as the risk measure. Assume that the arms' distributions are sub-Gaussian with common parameter σ. Then the expected regret of Risk-LCB at the end of $n \geq 1$ rounds with the LCB given by* (34) *satisfies*

$$\mathbb{E}(R_n) \leq \sum_{\{i:\Delta_i>0\}} \left(\frac{\left[\tilde{c}_4[(32\sqrt{\mathsf{e}\log n} + 512)\sigma]^{\alpha} + \tilde{c}_3\right]^{\frac{1}{\alpha\min\left\{\frac{1}{2},1-\alpha\right\}}}}{\Delta_i^{\frac{1}{\alpha\min\left\{\frac{1}{2},1-\alpha\right\}}-1}} + \left(1 + \frac{8\pi^2}{3}\right)\Delta_i\right), \tag{35}$$

and

$$\mathbb{E}(R_n) \leq \left(K\left[\tilde{c}_4[(32\sqrt{\mathsf{e}\log n} + 512)\sigma]^{\alpha} + \tilde{c}_3\right]^{\frac{1}{\alpha\min\left\{\frac{1}{2},1-\alpha\right\}}} + \left(1 + \frac{8\pi^2}{3}\right)\sum_{\{i:\Delta_i>0\}} \Delta_i^{\frac{1}{\alpha\min\left\{\frac{1}{2},1-\alpha\right\}}}\right)^{\alpha\min\left\{\frac{1}{2},1-\alpha\right\}} n^{1-\alpha\min\left\{\frac{1}{2},1-\alpha\right\}}, \tag{36}$$

where $\tilde{c}_3 = \frac{2L(K^+ + K^-)}{\alpha(1-\alpha)}$ *and* $\tilde{c}_4 = L(K^+ + K^-)\left(\max\left\{\frac{K^+}{k^+}, \frac{K^-}{k^-}\right\}(\sqrt{\log n} + 1)\right)^{1-\alpha}$.

Proof See Section 11.2. ■

Ignoring log factors, the regret bound above is $\tilde{O}\left(\sum_i \frac{1}{\Delta_i^{\frac{1}{\alpha\min\left\{\frac{1}{2},1-\alpha\right\}}-1}}\right)$. The parameter α is the Hölder exponent of the weight function, and is less than 1 for the weight function shown in Figure 2. Thus, the regret bound for CPT is weaker than the one for classical UCB that finds the arm with the best mean.

In Gopalan et al. (2017), the authors provide a regret upper bound for a UCB-type algorithm with CPT-value as the risk measure under the assumption that the underlying arms' distributions have bounded support. In particular, the authors in Gopalan et al. (2017) establish a regret upper bound of the order $\tilde{O}\left(\sum_i \frac{1}{\Delta_i^{\frac{2}{\alpha}-1}}\right)$, and also show that this upper bound cannot be improved as far as the dependence on gaps and the horizon are concerned through a minimax regret lower bound. We relax this assumption to consider the more general class of sub-Gaussian arms' distributions. For $\alpha < \frac{1}{2}$, our regret bound in Theorem 52 matches the bound in Gopalan et al. (2017), while for $\alpha > \frac{1}{2}$, our bound is weaker. It would be interesting future research to check if one can obtain an improved regret bound with sub-Gaussian arms for the latter case, or if the lower bound in Gopalan et al. (2017) is sub-optimal in this case.

10. Proofs

10.1 Proofs of the claims in Section 2

10.1.1 Proof of Lemma 2

Proof The first equality in (1) is given by the Kantorovich-Rubinstein theorem (Givens and Shortt, 1984; Edwards, 2011). The second equality is given in Vallander (1974).

To prove the third inequality in (1), we note that the integral on the left hand side of the third inequality is unchanged if we replace F_1 and F_2 by the point-wise maximum and minimum, respectively, of F_1 and F_2. Hence, without loss of generality, we may assume that $F_1(s) \geq F_2(s)$ for all $s \in \mathbb{R}$. The integral in question then reduces to

$$\int_{-\infty}^{\infty} |F_1(s) - F_2(s)| \mathrm{d}s = \int_{-\infty}^{\infty} (F_1(s) - F_2(s)) \mathrm{d}s = \int_{-\infty}^{\infty} \int_{F_2(s)}^{F_1(s)} \mathrm{d}\beta \mathrm{d}s. \tag{37}$$

It can easily be shown from the definition of the generalized inverse that

$$\begin{aligned} \{(\beta, s) \in \mathbb{R}^2 : F_2(s) < \beta < F_1(s)\} \subseteq & \quad \{(\beta, s) \in \mathbb{R}^2 : F_1^{-1}(\beta) \leq s \leq F_2^{-1}(\beta)\} \\ \subseteq & \quad \{(\beta, s) \in \mathbb{R}^2 : F_2(s) \leq \beta \leq F_1(s)\}. \end{aligned}$$

This justifies interchanging the order of integration (see Theorem 14.14 of Apostol (1974)) in (37), which yields

$$\int_{-\infty}^{\infty} |F_1(s) - F_2(s)| \mathrm{d}s = \int_0^1 \int_{F_1^{-1}(\beta)}^{F_2^{-1}(\beta)} \mathrm{d}s \mathrm{d}\beta = \int_0^1 [F_2^{-1}(\beta) - F_1^{-1}(\beta)] \mathrm{d}\beta.$$

The third inequality in (1) now follows by noting that, under our assumption that $F_1(s) \geq F_2(s)$ for all $s \in \mathbb{R}$, we have $F_2^{-1}(\beta) \geq F_1^{-1}(\beta)$ for all $\beta \in [0, 1]$. ■

10.1.2 Proof of Lemma 3

Proof The lemma follows directly by applying case (1) of Theorem 2 in (Fournier and Guillin, 2015) to the r.v. X. ■

10.1.3 Proof of Lemma 8

We first state and prove a variation of McDiarmid's inequality, which will be used subsequently to prove Lemma 8. Let $\mathcal{X} = (X_1, \ldots, X_n)$ denote a vector of n i.i.d. samples from a common distribution, say F. For each $i = 1, \ldots, n$, let $\mathcal{X}'_{(i)} = (X_1, \ldots, X_{i-1}, X'_i, X_{i+1}, \ldots, X_n)$, where X'_i denotes an independent copy of X_i. Let f be a real-valued function on $\mathbb{R}^n$ satisfying $\mathbb{E}[|f(\mathcal{X})|] < \infty$. For each $i = 1, \ldots, n$, define

$$D_i = f(\mathcal{X}) - f(\mathcal{X}'_{(i)}).$$

Finally, let $\mathcal{F}_0$ denote the trivial σ-field and, for each $k = 1, \ldots, n$, let $\mathcal{F}_k$ denote the σ-field generated by the samples $\{X_i, i \leq k\}$

Lemma 53 *Let $n \geq 1$, f and D_i, $i = 1, \ldots, n$, be as above. Suppose there exists $\tilde{\sigma} > 0$ such that, for each $i = 1, \ldots, n$, D_i satisfies*

$$\mathbb{E}\left(|D_i|^k | \mathcal{F}_{i-1}\right) \leq 4^k \tilde{\sigma}^k k^{k/2}, \forall k \geq 1.$$

Then, for every $\tilde{\epsilon} \in (0, 16n\tilde{\sigma}\sqrt{\mathsf{e}})$, we have

$$\mathbb{P}\left(f(\mathcal{X}) - \mathbb{E}[f(\mathcal{X})] > \tilde{\epsilon}\right) \leq \exp\left(-\frac{\tilde{\epsilon}^2}{256 n \tilde{\sigma}^2 \mathsf{e}}\right),$$

where e is Euler's number.

Proof To begin, choose $\tilde{\epsilon} \in (0, 16n\tilde{\sigma}\sqrt{\mathsf{e}})$ and $i \in \{1, \ldots, n\}$. Since X_i and X'_i have the same conditional distribution given $\mathcal{F}_{i-1}$, it is easy to see that $\mathbb{E}(D_i | \mathcal{F}_{i-1}) = 0$. Let $c = \tilde{\epsilon}/(128 n \tilde{\sigma}^2 \mathsf{e})$, and note that $32 c^2 \tilde{\sigma}^2 \mathsf{e} < 1/2$. Using $\mathsf{e}^x \leq x + \mathsf{e}^{x^2}$ and $\mathbb{E}(D_i | \mathcal{F}_{i-1}) = 0$, we obtain

$$\begin{aligned}
\mathbb{E}[\exp(cD_i) | \mathcal{F}_{i-1}] &\leq \mathbb{E}\left(cD_i + \exp(c^2 D_i^2) | \mathcal{F}_{i-1}\right) \\
&= \mathbb{E}\left(\exp(c^2 D_i^2) | \mathcal{F}_{i-1}\right) = 1 + \sum_{k \geq 1} \frac{c^{2k} \mathbb{E}(D_i^{2k} | \mathcal{F}_{i-1})}{k!} \\
&\leq 1 + \sum_{k \geq 1} \frac{c^{2k} 4^{2k} \tilde{\sigma}^{2k} (2k)^k}{k!} \\
&\leq 1 + \sum_{k \geq 1} \frac{(4c\tilde{\sigma})^{2k} (2k)^k}{(k/\mathsf{e})^k} \quad (\text{ Using Stirling's approximation } k! \geq (k/\mathsf{e})^k) \\
&= \sum_{k \geq 0} (32 c^2 \tilde{\sigma}^2 \mathsf{e})^k = \frac{1}{1 - 32 c^2 \tilde{\sigma}^2 \mathsf{e}} \quad (\text{ since } 32 c^2 \tilde{\sigma}^2 \mathsf{e} < 1) \\
&\leq \exp\left(64 c^2 \tilde{\sigma}^2 \mathsf{e}\right) \quad (\text{ since } 32 c^2 \tilde{\sigma}^2 \mathsf{e} < 1/2),
\end{aligned} \tag{38}$$

where we have used the inequality $1/(1-x) \leq \mathsf{e}^{2x}$ for $0 < x < 1/2$.

Next, define $\Delta_i = \mathbb{E}[f(\mathcal{X}) \,|\mathcal{F}_k] - \mathbb{E}[f(\mathcal{X}) \,|\mathcal{F}_{k-1}]$ for each $i = 1, \ldots, n$. Fix $i \in \{1, \ldots, n\}$. Since X_i and X'_i are identically distributed and all the samples are independent, we have

$$\mathbb{E}(f(\mathcal{X}) | \mathcal{F}_{i-1}) = \mathbb{E}(f(\mathcal{X}'_{(i)}) | \mathcal{F}_{i-1}) = \mathbb{E}(f(\mathcal{X}'_{(i)}) | \mathcal{F}_i).$$

As a result, we can write $\Delta_i = \mathbb{E}(D_i | \mathcal{F}_i)$. Applying Jensen's inequality for conditional expectations now yields

$$\mathbb{E}(\exp(c\Delta_i) | \mathcal{F}_{i-1}) = \mathbb{E}[\exp(c\mathbb{E}(D_i | \mathcal{F}_i)) | \mathcal{F}_{i-1}] \leq \mathbb{E}[\mathbb{E}(\exp(cD_i) | \mathcal{F}_i) | \mathcal{F}_{i-1}] = \mathbb{E}(\exp(cD_i) | \mathcal{F}_{i-1}).$$

Combining the above inequality with (38) gives

$$\mathbb{E}(\exp(c\Delta_i)|\mathcal{F}_{i-1}) \le \exp\left(64c^2\tilde{\sigma}^2\mathsf{e}\right) \tag{39}$$

On noting that $f(\mathcal{X}) - \mathbb{E}(f(\mathcal{X})) = \sum_{i=1}^{n} \Delta_i$ and using (39), we have

$$\begin{aligned}
\mathbb{P}\left(f(\mathcal{X}) - \mathbb{E}(f(\mathcal{X})) > \tilde{\epsilon}\right) &= \mathbb{P}\left(\sum_{k=1}^{n} \Delta_k > \tilde{\epsilon}\right) \\
&\le \exp(-c\tilde{\epsilon})\mathbb{E}\left[\exp\left(c\sum_{k=1}^{n}\Delta_k\right)\right] \\
&\le \exp(-c\tilde{\epsilon})\mathbb{E}\left[\exp\left(c\sum_{k=1}^{n-1}\Delta_k\right)\mathbb{E}\left[\exp(c\Delta_n)|\mathcal{F}_{n-1}\right]\right] \\
&\le \exp(-c\tilde{\epsilon})\mathbb{E}\left[\exp\left(c\sum_{k=1}^{n-1}\Delta_k\right)\exp\left(64c^2\tilde{\sigma}^2\mathsf{e}\right)\right] \\
&\vdots \\
&\le \exp(-c\tilde{\epsilon})\exp\left(64nc^2\tilde{\sigma}^2\mathsf{e}\right) \\
&= \exp\left(-\frac{\tilde{\epsilon}^2}{256n\tilde{\sigma}^2\mathsf{e}}\right),
\end{aligned}$$

where the suppressed steps involve successively conditioning over $\mathcal{F}_{n-2}, \mathcal{F}_{n-3}, \ldots, \mathcal{F}_0$ and using (38) at each step. The final equality comes from substituting the value of c. ■

Before proving Lemma 8, we provide a standard result on sub-Gaussian r.v.s that establishes bounds on its moments. For the sake of completeness, we prove this result so that the constants in the bound can be inferred easily.

Lemma 54 *Suppose a r.v. X is sub-Gaussian with parameter σ. Then X satisfies*

$$\left(\mathbb{E}\left|X\right|^k\right)^{\frac{1}{k}} \le 2\sigma\sqrt{k},\ \forall k \ge 1.$$

Proof Notice that

$$\begin{aligned}
\mathbb{E}\left(|X|^k\right) &= \int_0^\infty \mathbb{P}\left(|X|^k \ge u\right) du \\
&= \int_0^\infty \mathbb{P}\left(|X| \ge \epsilon\right) k\epsilon^{k-1} d\epsilon \\
&\le \int_0^\infty 2\exp\left(-\frac{\epsilon^2}{2\sigma^2}\right) k\epsilon^{k-1} d\epsilon \\
&\le 2^{\frac{k}{2}}\sigma^k k \int_0^\infty \exp(-s)\, s^{\frac{k}{2}-1} ds \\
&= 2^{\frac{k}{2}}\sigma^k k\Gamma\left(\frac{k}{2}\right) \le 2^{\frac{k}{2}}\sigma^k k\left(\frac{k}{2}\right)^{\frac{k}{2}} \quad \text{(Since } \Gamma(x) \le x^x\text{)}
\end{aligned}$$

$$\leq \sigma^k k\,(k)^{\frac{k}{2}}.$$

Hence,

$$\left(\mathbb{E}\,|X|^k\right)^{\frac{1}{k}} \leq \sigma\sqrt{k}\,(k)^{\frac{1}{k}} \leq 2\sigma\sqrt{k}.$$

Hence proved. ■

Proof [Lemma 8] Choose n and ϵ as in the lemma, let $f(\mathcal{X}) = W_1(F_n, F)$, and fix $i \in \{1, \ldots, n\}$. On using the triangle inequality for the Wasserstein distance, we obtain

$$|f(\mathcal{X}) - f(\mathcal{X}'_{(i)})| \leq W_1(F_n, F'_n) \leq \frac{1}{n}|X_i - X'_i|,$$

where F'_n denotes the EDF obtained from the sample $\mathcal{X}'_{(i)}$, and the last inequality follows from the definition of an EDF. Setting $D_i = f(\mathcal{X}) - f(\mathcal{X}'_{(i)})$, using the independence of the samples and the inequalities $|x-y|^k \leq (|x|+|y|)^k \leq 2^{k-1}(|x|^k + |y|^k)$ for $k \geq 1$ (see Fact 2.2.59 in Bernstein (2018)), and then applying Lemma 54 to the sub-Gaussian r.v.s X_i, X'_i, we obtain

$$\mathbb{E}(|D_i|^k|\mathcal{F}_{i-1}) = \frac{1}{n^k}\mathbb{E}(|X_i - X'_i|^k) \leq \frac{2^k}{n^k}\mathbb{E}(|X_i|^k) \leq 4^k\left(\frac{\sigma}{n}\right)^k k^{k/2} \tag{40}$$

for every $k \geq 1$, where we have also used $\mathbb{E}(|X_i|^k) = \mathbb{E}(|X'_i|^k)$.

The inequalities above show that the assumptions of Lemma 53 hold with $\tilde{\sigma} = \sigma/n$. Letting $\tilde{\epsilon} = \epsilon - \frac{512\sigma}{\sqrt{n}}$, we see that $\tilde{\epsilon} \in (0, 16n\tilde{\sigma}\sqrt{\mathsf{e}})$. Applying Lemma 53 now yields

$$\mathbb{P}\left(f(\mathcal{X}) - \mathbb{E}(f(\mathcal{X})) > \tilde{\epsilon}\right) \leq \exp\left(-\frac{\tilde{\epsilon}^2}{256n(\sigma/n)^2\mathsf{e}}\right) = \exp\left(-\frac{n\tilde{\epsilon}^2}{256\sigma^2\mathsf{e}}\right). \tag{41}$$

To infer the final claim in the lemma statement, we need to bound $\mathbb{E}(f(\mathcal{X}))$. For this purpose, we first specify the bound from Theorem 3.1 of Lei (2020) and later specialize to our setting. For a r.v. X satisfying $\mathbb{E}\left(|X|^q\right) < \top^q < \infty$ and some $q > p \geq 1$,

$$\mathbb{E}(W_p(F_n, F)) \leq c_{p,q}\top n^{-\min\left\{\frac{1}{\max\{2p,1\}}, \frac{1}{p} - \frac{1}{q}\right\}}(\log n)^{\frac{\zeta}{p}}, \tag{42}$$

where

$$\zeta = \begin{cases} 2 & \text{if } 1 = q = 2p \\ 1 & \text{if ``} 1 \neq 2p \text{ and } q = \min(\frac{p}{1-p}, 2p)\text{'' or ``}q > 1 = 2p\text{''}, \\ 0 & \text{else.} \end{cases}$$

and $c_{p,q}$ is a constant that depends on p and q. Tracing through the proof of the aforementioned theorem, we found that $c_{p,q} = (1+3^p)2^{2p-2}2^{q+1}$.

In our setting, $p = 1$. Choosing $q = 4$, we obtain $c_{p,q} = 128$, $\top = 4\sigma$ using Lemma 54. Further, $\zeta = 0$ for this choice of p, q. Thus, applying the bound in (42) leads to

$$\mathbb{E}(f(\mathcal{X})) \leq \frac{512\sigma}{\sqrt{n}}. \tag{43}$$

The main claim now follows by combining (41) and (43), and substituting for $\tilde{\epsilon}$. ■

10.1.4 PROOF OF LEMMA 10

For establishing the bound in Lemma 10, we need to invoke a Wasserstein concentration result for r.v.s satisfying the 'Bernstein's condition'. We specify this condition and show that a sub-exponential r.v. satisfies the same.

A r.v. X satisfies the Bernstein's condition if there exist $\sigma, b > 0$ such that

$$\mathbb{E}\left(|X|^k\right) \leq \frac{1}{2}\sigma^2(k!)b^{k-2} \text{ for all } k \geq 2. \tag{44}$$

Note that (44) does not require the r.v. to necessarily have mean zero. The result below shows that a sub-exponential r.v. satisfies (44)

Lemma 55 *Suppose a r.v. X satisfies (C2) with parameter $c > 0$. Then X satisfies the Bernstein's condition* (44) *with parameters $\sigma = \frac{2}{c}$ and $b = \frac{1}{c}$.*

Proof Suppose X is sub-exponential. Then, we have

$$\begin{aligned}
\mathbb{E}\left(|X|^k\right) &= \int_0^\infty \mathbb{P}\left(|X|^k \geq u\right) du \\
&= \int_0^\infty \mathbb{P}\left(|X| \geq \epsilon\right) k\epsilon^{k-1} d\epsilon \\
&\leq \int_0^\infty 2\exp(-c\epsilon) k\epsilon^{k-1} d\epsilon \\
&\leq \int_0^\infty \frac{2k}{c^k} \exp(-s) s^{k-1} ds \\
&= \frac{2k}{c^k}\Gamma(k) = \frac{(2/c)^2}{2} k! \left(\frac{1}{c}\right)^{k-2}.
\end{aligned}$$

Thus, X satisfies the Bernstein's condition (44). ■

Next, we state a variant of Lemma 53 for the sub-exponential case, and subsequently prove Lemma 10 by applying this result. The lemma appears as Theorem 5.1 in Lei (2020), but we provide a proof here for the sake of completeness.

Lemma 56 *Let $n \geq 1$, f and D_i, $i = 1, \ldots, n$, be as defined above Lemma 53. Suppose there exist $\tilde{\sigma}, \tilde{b} > 0$ such that, for each $i = 1, \ldots, n$, D_i satisfies*

$$\mathbb{E}\left(|D_i|^k \middle| \mathcal{F}_{i-1}\right) \leq \frac{1}{2}\tilde{\sigma}^2(k!)\tilde{b}^{k-2} \text{ for all } k \geq 2, \tag{45}$$

where the σ-fields $\mathcal{F}_0, \ldots, \mathcal{F}_{n-1}$ are as defined in subsection 10.1.3. Then, for every $\tilde{\epsilon} > 0$, we have

$$\mathbb{P}\left(f(\mathcal{X}) - \mathbb{E}[f(\mathcal{X})] > \tilde{\epsilon}\right) \leq \exp\left(-\frac{\tilde{\epsilon}^2}{2n\tilde{\sigma}^2 + 2\tilde{\epsilon}\tilde{b}}\right).$$

Proof To begin, choose $\tilde{\epsilon} > 0$ and $i \in \{1, \ldots, n\}$. Since X_i and X_i' have the same conditional distribution given $\mathcal{F}_{i-1}$, it is easy to see that $\mathbb{E}(D_i|\mathcal{F}_{i-1}) = 0$. Let $c = \frac{\tilde{\epsilon}}{n\tilde{\sigma}^2 + \tilde{b}\tilde{\epsilon}}$, and note that $\tilde{b}c < 1$. By expanding the exponential and using $\mathbb{E}(D_i|\mathcal{F}_{i-1}) = 0$, we may write

$$\mathbb{E}[\exp(cD_i)|\mathcal{F}_{i-1}] = \mathbb{E}\left(1 + cD_i + \sum_{k=2}^{\infty} \frac{c^k D_i^k}{k!} \middle| \mathcal{F}_{i-1}\right)$$

$$
\begin{aligned}
&\leq 1 + \sum_{k\geq 2} \frac{c^k \mathbb{E}(|D_i|^k|\mathcal{F}_{i-1})}{k!} \leq 1 + \frac{1}{2}\sum_{k\geq 2} c^k \tilde{\sigma}^2 \tilde{b}^{k-2} \quad \text{(Using (45))}\\
&= 1 + \frac{1}{2} c^2 \tilde{\sigma}^2 \sum_{k\geq 0} (c\tilde{b})^k\\
&\leq \exp\left(\frac{1}{2} c^2 \tilde{\sigma}^2 \sum_{k\geq 0} (c\tilde{b})^k\right) \quad \text{(Using the inequality } 1 + ax \leq \mathsf{e}^{ax} \text{ for } a, x > 0)\\
&= \exp\left(\frac{1}{2}\frac{c^2\tilde{\sigma}^2}{(1-c\tilde{b})}\right) \quad \text{(since } c\tilde{b} < 1). \qquad (46)
\end{aligned}
$$

Next, define $\Delta_i = \mathbb{E}[f(\mathcal{X})\,|\mathcal{F}_i] - \mathbb{E}[f(\mathcal{X})\,|\mathcal{F}_{i-1}]$ for each $i = 1, \ldots, n$. Fix $i \in \{1, \ldots, n\}$. Since X_i and X_i' are identically distributed and all the samples are independent, we have

$$\mathbb{E}(f(\mathcal{X})|\mathcal{F}_{i-1}) = \mathbb{E}(f(\mathcal{X}'_{(i)})|\mathcal{F}_{i-1}) = \mathbb{E}(f(\mathcal{X}'_{(i)})|\mathcal{F}_i).$$

As a result, we can write $\Delta_i = \mathbb{E}(D_i|\mathcal{F}_i)$. Applying Jensen's inequality for conditional expectations now yields

$$\mathbb{E}(\exp(c\Delta_i)|\mathcal{F}_{i-1}) = \mathbb{E}[\exp(c\mathbb{E}(D_i|\mathcal{F}_i))|\mathcal{F}_{i-1}] \leq \mathbb{E}[\mathbb{E}(\exp(cD_i)|\mathcal{F}_i)|\mathcal{F}_{i-1}] = \mathbb{E}(\exp(cD_i)|\mathcal{F}_{i-1}).$$

Combining the above inequality with (46) gives

$$\mathbb{E}(\exp(c\Delta_i)|\mathcal{F}_{i-1}) \leq \exp\left(\frac{1}{2}\frac{c^2\tilde{\sigma}^2}{(1-c\tilde{b})}\right) \qquad (47)$$

On noting that $f(\mathcal{X}) - \mathbb{E}(f(\mathcal{X})) = \sum_{i=1}^{n} \Delta_i$ and using (47), we have

$$
\begin{aligned}
\mathbb{P}\left(f(\mathcal{X}) - \mathbb{E}(f(\mathcal{X})) > \tilde{\epsilon}\right) &= \mathbb{P}\left(\sum_{i=1}^{n} \Delta_i > \tilde{\epsilon}\right)\\
&\leq \exp(-c\tilde{\epsilon})\mathbb{E}\left[\exp\left(c\sum_{i=1}^{n}\Delta_i\right)\right]\\
&\leq \exp(-c\tilde{\epsilon})\mathbb{E}\left[\exp\left(c\sum_{i=1}^{n-1}\Delta_i\right)\mathbb{E}\left[\exp(c\Delta_n)|\mathcal{F}_{n-1}\right]\right]\\
&\leq \exp(-c\tilde{\epsilon})\mathbb{E}\left[\exp\left(c\sum_{i=1}^{n-1}\Delta_i\right)\exp\left(\frac{1}{2}\frac{c^2\tilde{\sigma}^2}{(1-c\tilde{b})}\right)\right]\\
&\;\vdots\\
&\leq \exp(-c\tilde{\epsilon})\exp\left(\frac{1}{2}\frac{nc^2\tilde{\sigma}^2}{(1-c\tilde{b})}\right) = \exp\left(-\frac{\tilde{\epsilon}^2}{2n\tilde{\sigma}^2 + 2\tilde{b}\tilde{\epsilon}}\right),
\end{aligned}
$$

where the suppressed steps involve successively conditioning over $\mathcal{F}_{n-2}, \mathcal{F}_{n-3}, \ldots, \mathcal{F}_0$ and using (47) at each step. The final equality comes from substituting the value of c. ■

Proof [Lemma 10] Define n and ϵ as in the lemma, and let $\tilde{\epsilon} = \epsilon - \frac{384}{c\sqrt{n}}$. Recall the notation $\mathcal{X}$ and $\mathcal{X}'_{(i)}$ introduced in subsection 10.1.3. Define f such that $f(\mathcal{X}) = W_1(F_n, F)$ is the Wasserstein distance between the EDF F_n formed from the samples $\mathcal{X}$ and the CDF F of X. For each $i = 1, \ldots, n$, let $D_i = f(\mathcal{X}) - f(\mathcal{X}'_{(i)})$.

By Lemma 55, X satisfies Bernstein's condition with $\sigma = 2/c$ and $b = 1/c$. Consider $i \in \{1, \ldots, n\}$. Using arguments similar to those employed in deriving (40) in the proof of Lemma 8 along with the Bernstein's condition for X_i and X'_i, we can show that

$$\mathbb{E}\left(|D_i|^k \middle| \mathcal{F}_{i-1}\right) \leq \frac{1}{2}\left(\frac{2\sigma}{n}\right)^2 (k!) \left(\frac{2b}{n}\right)^{k-2}$$

for every $i = 1, \ldots, n$ and every $k \geq 2$. Invoking Lemma 56 with $f(\mathcal{X}) = W_1(F_n, F)$, $\tilde{\sigma} = \frac{2\sigma}{n}$ and $\tilde{b} = \frac{2b}{n}$ now yields

$$\mathbb{P}\left(W_1(F_n, F) - \mathbb{E}(W_1(F_n, F)) > \tilde{\epsilon}\right) \leq \exp\left(-\frac{n\tilde{\epsilon}^2}{8\sigma^2 + 4b\tilde{\epsilon}}\right).$$

Substituting for σ and b gives

$$\mathbb{P}\left(W_1(F_n, F) - \mathbb{E}(W_1(F_n, F)) > \tilde{\epsilon}\right) \leq \exp\left(-\frac{n\tilde{\epsilon}^2}{\frac{32}{c^2} + \frac{4}{c}\tilde{\epsilon}}\right). \tag{48}$$

Next, applying the bound in (42) with $p = 1, q = 4$ leads to $\zeta = 0$. From Lemma 55, we have $\mathbb{E}\left(|X|^4\right) \leq \frac{(2/c)^2}{2} 4! \left(\frac{1}{c}\right)^{4-2} \leq \left(\frac{3}{c}\right)^4$, which implies $\top = \frac{3}{c}$. Using these values in (42), we obtain

$$\mathbb{E}(W_1(F_n, F)) \leq \frac{384}{c\sqrt{n}}.$$

The lemma now follows by using the last inequality in (48) and then substituting for $\tilde{\epsilon}$. ■

10.1.5 PROOF OF LEMMA 11

Proof The lemma follows directly by applying case (3) of Theorem 2 in Fournier and Guillin (2015) to the r.v. X. ■

10.2 Proofs of the claims in Section 3

10.2.1 PROOF OF LEMMA 12

Proof Choose $\xi \in \mathbb{R}$ arbitrarily and let $g_\xi(x) = L^{-1}[\xi + \phi(x - \xi)]$. Then,

$$\int_{\mathbb{R}} g_\xi(x) dF_X(x) = L^{-1}[\xi + \mathbb{E}\{\phi(X - \xi)\}] \triangleq D_X(\xi), \text{ and}$$
$$\int_{\mathbb{R}} g_\xi(x) dF_Y(x) = L^{-1}[\xi + \mathbb{E}\{\phi(Y - \xi)\}] \triangleq D_Y(\xi).$$

Observing that g_ξ is 1-Lipschitz in x for every $\xi \in \mathbb{R}$ and using (1), we obtain

$$|D_X(\xi) - D_Y(\xi)| \leq W_1(F_X, F_Y) \text{ for every } \xi \in \mathbb{R}.$$

Choose $m > 0$ arbitrarily, and let $\xi_1, \xi_2 \in \mathbb{R}$ be such that

$$D_X(\xi_1) \leq \inf_\xi D_X(\xi) + \frac{1}{m}, \text{ and } D_Y(\xi_2) \leq \inf_\xi D_Y(\xi) + \frac{1}{m}.$$

Then, we obtain

$$\begin{aligned}
&- W_1(F_X, F_Y) - \frac{1}{m} \leq D_X(\xi_1) - D_Y(\xi_1) - \frac{1}{m} \leq \inf_\xi D_X(\xi) - \inf_\xi D_Y(\xi) \\
&\leq D_X(\xi_2) - D_Y(\xi_2) + \frac{1}{m} \leq W_1(F_X, F_Y) + \frac{1}{m}.
\end{aligned}$$

Since the chain of inequalities above hold for every $m > 0$, we conclude that

$$\left| \inf_\xi D_X(\xi) - \inf_\xi D_Y(\xi) \right| \leq W_1(F_X, F_Y). \tag{49}$$

By definition, $\inf_\xi D_X(\xi) = L^{-1}\text{oce}^\phi(X)$ and $\inf_\xi D_Y(\xi) = L^{-1}\text{oce}^\phi(Y)$. Hence (9) follows immediately from (49). ■

10.2.2 PROOF OF LEMMA 15

Proof For convenience, define $f_X : \mathbb{R} \to \mathbb{R}$ by $f_X(\xi) = \mathbb{E}[l(X - \xi))]$, and note that $S_\alpha(X) = \inf\{\xi \in \mathbb{R} : f_X(\xi) \leq \alpha\}$. Also, since l is nondecreasing, it follows that f_X is nonincreasing. Define f_Y in an identical fashion.

Given $\xi \in \mathbb{R}$, note that $|f_X(\xi) - f_Y(\xi)| = K|\mathbb{E}[K^{-1}l(X - \xi)] - \mathbb{E}[K^{-1}l(Y - \xi)]|$. Since the function $x \mapsto K^{-1}l(x - \xi)$ is 1-Lipschitz, it follows from Lemma 2 and the last equality that

$$|f_X(\xi) - f_Y(\xi)| \leq KW_1(F_X, F_Y) \tag{50}$$

for every $\xi \in \mathbb{R}$.

Next, let $\epsilon > 0$, and choose $\xi_1, \xi_2 \in \mathbb{R}$ such that

$$S_\alpha(X) - \epsilon \leq \xi_1 < S_\alpha(X) < \xi_2 < S_\alpha(X) + \epsilon. \tag{51}$$

It follows from the definition of $S_\alpha(X)$ and the nonincreasing nature of f_X that $f_X(\xi_1) > \alpha$ while $f_X(\xi_2) \leq \alpha$. Next, define $\xi_1' = \xi_1 - \frac{K}{k}W_1(F_X, F_Y)$ and $\xi_2' = \xi_2 + \frac{K}{k}W_1(F_X, F_Y)$. Since $\xi_1' \leq \xi_1$, the third assumption on l implies that $l(X - \xi_1') - l(X - \xi_1) \geq k(\xi_1 - \xi_1') = KW_1(F_X, F_Y)$ almost surely. Taking expectations gives

$$f_X(\xi_1') - f_X(\xi_1) \geq KW_1(F_X, F_Y). \tag{52}$$

Similarly, it can be shown that

$$f_X(\xi_2) - f_X(\xi_2') \geq KW_1(F_X, F_Y). \tag{53}$$

Applying (50) and (52) along with our choice of ξ_1, we get

$$f_Y(\xi_1') \geq f_X(\xi_1') - KW_1(F_X, F_Y) \geq f_X(\xi_1) > \alpha. \tag{54}$$

Similarly, applying (50) and (53) along with our choice of ξ_2, we get

$$f_Y(\xi_2') \leq f_X(\xi_2') + KW_1(F_X, F_Y) \leq f_X(\xi_2) \leq \alpha. \tag{55}$$

The inequalities (54) and (55) together imply that $\xi_1' \leq S_\alpha(Y) \leq \xi_2'$. Substituting for ξ_1' and ξ_2' and using the inequalities (51) gives

$$S_\alpha(X) - \frac{K}{k} W_1(F_X, F_Y) - \epsilon \leq S_\alpha(Y) \leq S_\alpha(X) + \frac{K}{k} W_1(F_X, F_Y) + \epsilon.$$

that $g_{\mu,\xi}(\cdot)$ is a 1-Lipschitz function for each ξ and μ. Hence, by Lemma 1, (17) holds. This completes the proof. ■

10.2.3 PROOF OF LEMMA 16

Proof Choose $\tau > 0$. Let X and Y be r.v.s having CDFs F and G, respectively. Recall that the CDF of $\hat{Y} = Y\mathbb{I}\{-\tau \leq Y < \tau\}$ is $G|_\tau$. Then

$$C(F) - C(G|_\tau) = \Delta^+ - \Delta^-, \tag{56}$$

where

$$\Delta^+ = \int_0^\infty w^+\left(\mathbb{P}\left(u^+(X) > z\right)\right)\mathrm{d}z - \int_0^\infty w^+\left(\mathbb{P}\left(u^+(\hat{Y}) > z\right)\right)\mathrm{d}z, \text{ and}$$
$$\Delta^- = \int_0^\infty w^-\left(\mathbb{P}\left(u^-(X) > z\right)\right)\mathrm{d}z - \int_0^\infty w^-\left(\mathbb{P}\left(u^-(\hat{Y}) > z\right)\right)\mathrm{d}z.$$

On noting that $\mathbb{P}\left(u^+(\hat{Y}) > z\right) = 0$ for $z \geq u^+(\tau)$ and $\mathbb{P}\left(u^+(\hat{Y}) > z\right) = \mathbb{P}\left(u^+(Y) > z\right)$ for $0 \leq z < u^+(\tau)$, we get

$$\Delta^+ = \int_0^{u^+(\tau)} w^+\left(\mathbb{P}\left(u^+(X) > z\right)\right)\mathrm{d}z - \int_0^{u^+(\tau)} w^+\left(\mathbb{P}\left(u^+(Y) > z\right)\right)\mathrm{d}z$$
$$+ \int_{u^+(\tau)}^\infty w^+\left(\mathbb{P}\left(u^+(X) > z\right)\right)\mathrm{d}z.$$

Using the assumption of Hölder continuity on the weight function w^+ along with $w^+(0) = 0$ gives

$$\left|\Delta^+\right| \leq L\int_0^{u^+(\tau)} |F^+(z) - G^+(z)|^\alpha \mathrm{d}z + L\int_{u^+(\tau)}^\infty [1 - F^+(z)]^\alpha \mathrm{d}z, \tag{57}$$

where $F^+(\cdot)$ and $G^+(\cdot)$ are the CDFs of the r.v.s $u^+(X)$ and $u^+(Y)$, respectively.

Applying Jensen's inequality to the concave function $x \mapsto x^\alpha$ after normalizing the Lebesgue measure on the interval $[0, u^+(\tau)]$, we obtain

$$\frac{1}{u^+(\tau)}\int_0^{u^+(\tau)} |F^+(z) - G^+(z)|^\alpha \mathrm{d}z \leq \left[\frac{1}{u^+(\tau)}\int_0^{u^+(\tau)} |F^+(z) - G^+(z)|\mathrm{d}z\right]^\alpha$$

$$
\begin{aligned}
&\leq \left[\frac{1}{u^+(\tau)}\int_0^\tau |F(v)-G(v)|(u^+)'(v)\mathrm{d}v\right]^\alpha \\
&\leq \left[\frac{K^+}{u^+(\tau)}\int_{-\infty}^\infty |F(v)-G(v)|\mathrm{d}v\right]^\alpha.
\end{aligned}
$$

In the second inequality above, which follows by using the substitution $z = u^+(v)$, $(u^+)'(v)$ denotes the derivative of $u^+(\cdot)$ at v. Applying the second equality in Lemma 2 to the CDFs F and G gives

$$\int_0^{u^+(\tau)} |F^+(z)-G^+(z)|^\alpha \mathrm{d}z \leq [K^+W_1(F,G)]^\alpha [u^+(\tau)]^{1-\alpha}.$$

Our assumption that the derivative of u^+ on $[0,\infty)$ takes values in $[k^+,K^+]$ implies that $k^+\tau \leq u^+(\tau) \leq K^+\tau$. Hence, we get

$$\int_0^{u^+(\tau)} |F^+(z)-G^+(z)|^\alpha \mathrm{d}z \leq K^+[W_1(F,G)]^\alpha \tau^{1-\alpha}. \tag{58}$$

We also see that $F^+(z) = \mathbb{P}(u^+(X)\leq z) \geq \mathbb{P}(K^+X \leq z) = F(z/K^+)$. Using this along with $u^+(\tau)\geq k^+\tau$ in the second integral in (57) gives

$$\int_{u^+(\tau)}^\infty [1-F^+(z)]^\alpha \mathrm{d}z \leq \int_{k^+\tau}^\infty [1-F(z/K^+)]^\alpha \mathrm{d}z. \tag{59}$$

Performing a simple change of variables in (59), and using the outcome along with (58) in (57) gives

$$\left|\Delta^+\right| \leq LK^+[W_1(F,G)]^\alpha\tau^{1-\alpha} + LK^+\int_{\frac{k^+}{K^+}\tau}^\infty [1-F(z)]^\alpha \mathrm{d}z. \tag{60}$$

It follows from almost identical arguments that

$$\left|\Delta^-\right| \leq LK^-[W_1(F,G)]^\alpha\tau^{1-\alpha} + LK^-\int_{-\infty}^{-\frac{k^-}{K^-}\tau} [F(z)]^\alpha \mathrm{d}z. \tag{61}$$

Using (60) and (61) in (56) completes the proof. ■

10.2.4 PROOF OF LEMMA 18

Proof

$$
\begin{aligned}
V(F) &= \int_{-\infty}^\infty u(x)\mathrm{d}(w\circ F)(x) \\
&= \int_0^\infty \left(\int_0^{u(x)} \mathrm{d}z\right)\mathrm{d}(w\circ F)(x) + \int_{-\infty}^0 \left(\int_0^{u(x)} \mathrm{d}z\right)\mathrm{d}(w\circ F)(x) \\
&= \int_0^\infty \left(\int_0^{u(x)} \mathrm{d}z\right)\mathrm{d}(w\circ F)(x) - \int_{-\infty}^0 \left(\int_0^{u^-(x)} \mathrm{d}z\right)\mathrm{d}(w\circ F)(x)
\end{aligned}
$$

$$
\begin{aligned}
&= \int_0^{\infty} \left(\int_{u^{-1}(z)}^{\infty} \mathrm{d}(w \circ F)(x) \right) \mathrm{d}z - \int_0^{\infty} \left(\int_{-\infty}^{(u^-)^{-1}(z)} \mathrm{d}(w \circ F)(x) \right) \mathrm{d}z \\
&= \underbrace{\int_0^{\infty} \left(1 - w\left(F(u^{-1}(z))\right)\right) \mathrm{d}z}_{(A)} - \underbrace{\int_0^{\infty} w\left(F((u^-)^{-1}(z))\right) \mathrm{d}z}_{(B)},
\end{aligned}
$$

The first term on the RHS above may be simplified as

$$
\begin{aligned}
(A) &= \int_0^{\infty} \left(1 - w\left(\mathbb{P}\left(X \le u^{-1}(z)\right)\right)\right) \mathrm{d}z = \int_0^{\infty} \left(1 - w\left(\mathbb{P}\left(u(X) \le z\right)\right)\right) \mathrm{d}z \\
&= \int_0^{\infty} w^+\left(\mathbb{P}\left(u(X) > z\right)\right) \mathrm{d}z = \int_0^{\infty} w^+\left(\mathbb{P}\left(u^+(X) > z\right)\right) \mathrm{d}z,
\end{aligned}
$$

where the third equality follows by the definition of w^+, and the last equality by observing that $\{x : u(x) > z\} = \{x : u^+(x) > z\}$ for every $z > 0$.

Along similar lines, we obtain

$$
\begin{aligned}
(B) &= \int_0^{\infty} w\left(F((u^-)^{-1}(z))\right) \mathrm{d}z = \int_0^{\infty} w\left(\mathbb{P}\left(X \le (u^-)^{-1}(z)\right)\right) \mathrm{d}z \\
&= \int_0^{\infty} w^-\left(\mathbb{P}\left(u^-(X) \ge z\right)\right) \mathrm{d}z = \int_0^{\infty} w^-\left(\mathbb{P}\left(u^-(X) > z\right)\right) \mathrm{d}z,
\end{aligned}
$$

where we used the fact that u^- is decreasing. The claim follows. ■

10.3 Proofs of the claims in Section 5

10.3.1 Proof of Theorem 19

Proof Let F denote the distribution of X. Then, we have

$$
|\rho_n - \rho(X)| = |\rho(F_n) - \rho(F)| \le L(W_1(F_n, F))^{\kappa}. \tag{63}
$$

Using Theorem 3.1 of Lei (2020), we have

$$
\begin{aligned}
\mathbb{E}[W_1(F_n, F)] &\le \frac{2^{\beta+3}\top}{n^{\min(\frac{1}{2}, 1-\frac{1}{\beta})}}, \text{ implying} \\
\mathbb{E}\left[W_1(F_n, F)^{\kappa}\right] \le \left[\mathbb{E}(W_1(F_n, F))\right]^{\kappa} &\le \left(\frac{2^{\beta+3}\top}{n^{\min(\frac{1}{2}, 1-\frac{1}{\beta})}}\right)^{\kappa},
\end{aligned}
$$

where we used Jensen's inequality to infer $\mathbb{E}(W_1(F_n, F)^{\kappa}) \le [\mathbb{E}(W_1(F_n, F))]^{\kappa}$, since $\kappa \in (0, 1]$. The main claim follows by substituting the bound obtained above in (63). ■

10.3.2 PROOF OF THEOREM 23

Proof Since ρ is (T2) measure, we have

$$|\rho(X) - \rho_{n,\tau}| = |\rho(X) - \rho(F_n|_\tau)| \leq \underbrace{L_1 \left(W_1(F, F_n)\right)^{\alpha_1} \tau^\gamma}_{I_1} + \underbrace{L_2 \int_{K_1\tau}^{\infty} [1 - F(z)]^{\alpha_2} \mathrm{d}z}_{I_2}$$
$$+ \underbrace{L_3 \int_{-\infty}^{-K_2\tau} [F(z)]^{\alpha_3} \mathrm{d}z}_{I_3}. \quad (64)$$

The second term on the RHS above can be bounded as follows:

$$I_2 = L_2 \int_{K_1\tau}^{\infty} [1 - F(z)]^{\alpha_2} \mathrm{d}z \leq L_2 \top^{\alpha_2} \int_{K_1\tau}^{\infty} \frac{1}{z^{\beta\alpha_2}} \mathrm{d}z = \frac{L_2 \top^{\alpha_2}}{(\beta\alpha_2 - 1)(K_1\tau)^{\beta\alpha_2 - 1}}, \quad (65)$$

where we used the fact that $1 - F(z) = \mathbb{P}(X > z) \leq \mathbb{P}\left(|X|^\beta > z^\beta\right) \leq \frac{\top}{z^\beta}$, since the r.v. X satisfies (C4).

Along similar lines, the third term on the RHS of (64) can be bounded as follows:

$$I_3 = L_3 \int_{-\infty}^{-K_2\tau} [F(-z)]^{\alpha_3} \mathrm{d}z \leq \frac{L_3 \top^{\alpha_3}}{(\beta\alpha_3 - 1)(K_2\tau)^{\beta\alpha_3 - 1}}. \quad (66)$$

As in the proof of Theorem 19, we have

$$\mathbb{E}\left(W_1(F_n, F)^{\alpha_1}\right) \leq \left[\mathbb{E}(W_1(F_n, F))\right]^{\alpha_1} \leq \left(\frac{2^{\beta+3}\top}{n^{\min(\frac{1}{2}, 1 - \frac{1}{\beta})}}\right)^{\alpha_1}. \quad (67)$$

The claim follows by taking expectations in (64) followed by substitution of the bounds given by (65), (66) and (67). ∎

10.3.3 PROOF OF COROLLARY 24

Proof From Lemma 16, we know that CPT is (T2), with the following parameters:

$$L_1 = (K^+ + K^-)L, L_2 = LK^+, L_3 = LK^-, \gamma = 1 - \alpha, \alpha_1 = \alpha_2 = \alpha_3 = \alpha, K_1 = \frac{k^+}{K^+}, \text{ and}$$

$K_2 = \frac{k^-}{K^-}$. Here, α, L are the exponent and constant of the Hölder-continuous weight function in the definition of CPT.

The claim follows by substituting in (23) the values above, and the choice of τ specified in the corollary statement. ∎

10.4 Proofs of the claims in Section 6

10.4.1 Proof of Theorem 25

Proof Let F denote the CDF of X, and F_n denote the EDF formed from n independent samples of X. Consider the event $A = \{W_1(F, F_n) > \left(\frac{\epsilon}{L}\right)^{\frac{1}{\kappa}}\}$. Note that, by (7), the event $\{|\rho_n - \rho(X)| > \epsilon\}$ is contained in A. Equation (24) now follows by applying Lemma 3 to the event A. ■

10.4.2 Proof of Theorem 27

Proof The proof is similar to that of Theorem 25, except that one invokes Lemma 8 instead of Lemma 3. ■

10.4.3 Proof of Corollary 28

Proof From Lemma 12, we have that OCE is a (T1) risk measure with parameters $L = \frac{1}{1-\alpha}$, and $\kappa = 1$. The proof now follows by an application of Theorem 25. ■

10.4.4 Proof of Corollary 29

Proof The proof is similar to that of Corollary 28 except that one invokes Theorem 27 instead of Theorem 25. ■

10.4.5 Proof of Corollary 30

Proof The result follows by applying Theorem 25, after observing from Lemma 13 that a spectral risk measure is of type (T1) with parameters $L = K$ and $\kappa = 1$. ■

10.4.6 Proof of Corollary 31

Proof The proof is similar to that of Corollary 30 except that one invokes Theorem 27 instead of Theorem 25. ■

10.4.7 Proof of Corollary 32

Proof The result follows by applying Theorem 25, after observing that UBSR is a (T1) risk measure with parameters $L = K/k$ and $\kappa = 1$ (see Lemma 15). ■

10.4.8 PROOF OF COROLLARY 33

Proof The proof is similar to that of Corollary 32 except that one invokes Theorem 27 instead of Theorem 25. ■

10.4.9 PROOF OF THEOREM 34

Proof Recall that the since the r.v. X is bounded in $[-B_2, B_1]$, we have $F(z) = 0$, for $z < -B_2$ and $F(z) = 1$ for $z > B_1$. Thus, on choosing $\tau = \max\left(\frac{B_1}{K_1}, \frac{B_2}{K_2}\right)$, the second and third integrals on the RHS of (19) vanish, and the risk measure ρ satisfies

$$|\rho_n - \rho(X)| \leq L_1 \left(W_1(F, F_n)\right)^{\alpha_1} \tau^{\gamma}, \tag{68}$$

where α_1, γ and L_1 are parameters of the (T2) risk measure ρ as in (19).

Fix $\epsilon > 0$ and consider the event $A = \{W_1(F, F_n) > [\epsilon/\{L_1\tau^\gamma\}]^{1/\alpha_1}\}$, where τ is specified above. Recall that a bounded r.v. is sub-Gaussian with parameter σ determined by its bounds, and hence satisfies (C1) with $\beta = 2$. Hence applying Lemma 3 with $\beta = 2$, we obtain

$$\mathbb{P}\left(A\right) \leq c_1 \exp\left(-c_2 n \left(\frac{\epsilon}{L_1 \tau^\gamma}\right)^{\frac{2}{\alpha_1}}\right),$$

where c_1, and c_2 are constants that depend on B_1 and B_2. By (68), the event A contains the event $\{|\rho_n - \rho(X)| > \epsilon\}$, and the claim follows. ■

10.4.10 PROOF OF THEOREM 35

Proof Follows in a similar manner as Theorem 34, except that we invoke Lemma 8 in place of Lemma 3. For this invocation, we have used the fact that the sub-Gaussianity paramter $\sigma = \frac{(B_1+B_2)}{2}$ for a r.v. bounded within $[-B_2, B_1]$. ■

10.4.11 PROOF OF COROLLARY 36

Proof The result follows in a straightforward fashion by applying Theorem 34 to CPT-value, which is a (T2) risk measure (see Lemma 16). ■

10.4.12 PROOF OF COROLLARY 37

Proof The result follows in a straightforward fashion by applying Theorem 35. ■

10.4.13 PROOF OF THEOREM 38

Proof Since ρ is (T2) measure, we have

$$|\rho(X) - \rho_{n,\tau}| = |\rho(X) - \rho(F_n|_\tau)| \leq \underbrace{L_1\left(W_1(F,F_n)\right)^{\alpha_1}\tau^\gamma}_{I_1} + \underbrace{L_2\int_{K_1\tau}^\infty [1-F(z)]^{\alpha_2}\mathrm{d}z}_{I_2} + \underbrace{L_3\int_{-\infty}^{-K_2\tau}[F(z)]^{\alpha_3}\mathrm{d}z}_{I_3}. \tag{69}$$

The second term on the RHS above can be bounded as follows:

$$\begin{aligned} L_2\int_{K_1\tau}^\infty [1-F(z)]^{\alpha_2}\mathrm{d}z &\leq L_2\top^{\alpha_2}\int_{K_1\tau}^\infty \exp\left(-\alpha_2\gamma z^\beta\right)\mathrm{d}z \\ &\leq L_2\top^{\alpha_2}\int_{K_1\tau}^\infty \left(\frac{z}{K_1\tau}\right)^{\beta-1}\exp\left(-\alpha_2\gamma z^\beta\right)\mathrm{d}z \\ &= \frac{L_2}{(K_1\tau)^{\beta-1}\alpha_2\gamma(\beta-1)}\exp\left(-\alpha_2\gamma\left(K_1\tau\right)^\beta\right), \end{aligned} \tag{70}$$

where we used the fact that, for $z > 0$, we have

$$1-F(z) = \mathbb{P}\left(X > z\right) = \mathbb{P}\left(|X| > z\right) = \mathbb{P}\left(\exp(\gamma|X|^\beta) > \exp(\gamma z^\beta)\right) \leq \top\exp\left(-\gamma z^\beta\right),$$

since the r.v. X satisfies (C1).

Along similar lines, the third term on the RHS of (69) can be bounded as follows:

$$\begin{aligned} L_3\int_{-\infty}^{-K_2\tau}[F(z)]^{\alpha_3}\mathrm{d}z &\leq L_3\top^{\alpha_3}\int_{-\infty}^{-K_2\tau}\exp\left(-\alpha_3\gamma(-z)^\beta\right)\mathrm{d}z \\ &= L_3\top^{\alpha_3}\int_{K_2\tau}^\infty \exp\left(-\alpha_3\gamma z^\beta\right)\mathrm{d}z \\ &\leq L_3\top^{\alpha_3}\int_{K_2\tau}^\infty \left(\frac{z}{K_2\tau}\right)^{\beta-1}\exp\left(-\alpha_3\gamma z^\beta\right)\mathrm{d}z \\ &= \frac{L_3}{(K_2\tau)^{\beta-1}\alpha_3\gamma(\beta-1)}\exp\left(-\alpha_3\gamma\left(K_2\tau\right)^\beta\right), \end{aligned} \tag{71}$$

where the first inequality holds because, for $z < 0$, we have

$$F(z) = \mathbb{P}\left(X \leq z\right) = \mathbb{P}\left(|X| \geq -z\right) = \mathbb{P}\left(\exp(\gamma|X|^\beta) \geq \exp(\gamma(-z)^\beta)\right) \leq \top\exp\left(-\gamma(-z)^\beta\right).$$

The last inequality above uses Markov's inequality and (C1).

Using (70) and (71), we have

$$\mathbb{P}\left(|\rho(X) - \rho_{n,\tau}| > \epsilon\right) = \mathbb{P}\left(I_1 + I_2 + I_3 > \epsilon\right) = \mathbb{P}\left(I_1 > \epsilon - I_1 - I_2\right) \leq \mathbb{P}\left(I_1 > \epsilon'\right), \tag{72}$$

where ϵ' is as defined in the theorem statement.

Applying Lemma 3, we obtain

$$\mathbb{P}\left(I_1 > \epsilon'\right) = \mathbb{P}\left(W_1(F, F_n) > \left(\frac{\epsilon'}{L_1 \tau^\gamma}\right)^{\frac{1}{\alpha_1}}\right) \leq c_1 \exp\left(-\frac{c_2 n (\epsilon')^{2/\alpha_1}}{(L_1 \tau^\gamma)^{2/\alpha_1}}\right), \tag{73}$$

where c_1 and c_2 are σ-dependent constants. This completes the proof. ■

10.4.14 PROOF OF COROLLARY 39

Proof Let $n \geq 1$, $\epsilon > 0$ and τ_n be chosen as in the corollary. From Lemma 16, we know that CPT is (T2), with the parameters

$$L_1 = (K^+ + K^-)L, L_2 = LK^+, L_3 = LK^-, \gamma = 1-\alpha, \alpha_1 = \alpha_2 = \alpha_3 = \alpha, K_1 = \frac{k^+}{K^+}, \text{ and}$$

$K_2 = \frac{k^-}{K^-}$, where α and L are the exponent and Hölder constant, respectively, of the Hölder-continuous weight function in the definition of CPT.

Let ϵ' be as defined in (27) with $\tau = \tau_n$ and the other parameters as given above. Note that we may rewrite τ_n as $\tau_n = [1 + (\log n)^{\frac{1}{\beta}}] \max\{K_1^{-1}, K_2^{-1}\}$. As a result, we have $K_i \tau_n \geq (\log n)^{\frac{1}{\beta}}$ for $i = 1, 2$. Using these inequalities and substituting for various parameters in (27)as above, we get

$$\begin{aligned}
\epsilon' &\geq \epsilon - \frac{LK^+ \exp\left(-\alpha(1-\alpha)\log n\right)}{(K_1 \tau_n)^{\beta-1}\,\alpha(1-\alpha)(\beta-1)} - \frac{LK^- \exp\left(-\alpha(1-\alpha)\log n\right)}{(K_2 \tau_n)^{\beta-1}\,\alpha(1-\alpha)(\beta-1)} \\
&\geq \epsilon - \frac{L(K^+ + K^-)}{(\log n)^{\frac{\beta-1}{\beta}}\,\alpha(1-\alpha)(\beta-1) n^{\alpha(1-\alpha)}} = \epsilon - c_3(n).
\end{aligned}$$

Since $\epsilon > c_3(n)$, we have $\epsilon' > 0$, and Theorem 38 applies. Using the inequality $\epsilon' \geq \epsilon - c_3(n)$ in (28) and substituting for L_1, γ, α_1 and τ now yields the required bound in a straightforward fashion. ■

10.4.15 PROOF OF PROPOSITION 40

Proof Let $n \geq 1$, $\epsilon > 0$ and τ_n be chosen as in the proposition. From Lemma 16, we know that CPT is (T2), with the parameters

$$L_1 = (K^+ + K^-)L, L_2 = LK^+, L_3 = LK^-, \gamma = 1-\alpha, \alpha_1 = \alpha_2 = \alpha_3 = \alpha, K_1 = \frac{k^+}{K^+}, \text{ and}$$

$K_2 = \frac{k^-}{K^-}$, where α and L are the exponent and Hölder constant, respectively, of the Hölder-continuous weight function in the definition of CPT.

Note that X satisfies (C1) with $\beta = 2$. Following the steps leading to (72) and the equality in (73) in the proof of Theorem 38, we conclude that

$$\mathbb{P}\left(|C_n - C(X)| > \epsilon\right) \leq \mathbb{P}\left(W_1(F, F_n) > \left(\frac{\epsilon'}{L_1 \tau_n^\gamma}\right)^{\frac{1}{\alpha_1}}\right), \tag{74}$$

where ϵ' is as defined in (40) with $\tau = \tau_n$, $\beta = 2$ and the rest of the parameters as defined above.

As in the proof of Corollary 39, we can show that $\epsilon' \geq \epsilon - c_3(n)$. As a result, (74) yields

$$\mathbb{P}\left(|C_n - C(X)| > \epsilon\right) \leq \mathbb{P}\left(W_1(F, F_n) > \left(\frac{\epsilon - c_3(n)}{L_1 \tau_n^\gamma}\right)^{\frac{1}{\alpha_1}}\right). \tag{75}$$

On substituting for L_1, τ_n and γ, we also recognise that $L_1\tau_n^\gamma = c_4(n)$. The condition on ϵ given in the proposition implies that the assumptions of Lemma 8 hold with ϵ in that lemma replaced by $[(\epsilon - c_3(n))/c_4(n)]^{\frac{1}{\alpha}}$. Applying Lemma 8 to bound the right hand side in (75) yields the final result. ■

10.5 Proofs of the claims in Section 7

10.5.1 Proof of Theorem 41

Proof The proof is identical to the proof of Theorem 25 except that Lemma 10 is used in place of Lemma 3 there. ■

10.5.2 Proof of Theorem 45

Proof The proof follows the same development as in the proof of Theorem 38, except that the sub-exponential tail bound (4) is invoked instead of the sub-Gaussian tail bound (3) for bounding the integrals I_2 and I_3 in (69). More precisely, for $z > 0$, the sub-exponential tail bound (4) gives

$$1 - F(z) = \mathbb{P}\left(X > z\right) \leq \exp(-cz).$$

Using the bound allows us to verify through direct integration that

$$I_2 \leq \frac{L_2}{c\alpha_2} \exp(-\alpha_2 c K_1 \tau).$$

Similarly, for $z < 0$, the tail bound (4) gives

$$F(z) = \mathbb{P}\left(X < z\right) \leq \exp(cz),$$

which yields

$$I_3 \leq \frac{L_3}{c\alpha_3} \exp(-\alpha_3 c K_2 \tau).$$

These bounds on I_2 and I_3 lead to (72) with ϵ' defined as in Theorem 45. The equality in (73) immediately follows. Applying Lemma 10 to bound the probability in (73) completes the proof. ■

10.5.3 Proof of Corollary 46

Proof Let $n \geq 1$, $\epsilon > 0$ and τ_n be chosen as in the corollary. From Lemma 16, we know that CPT is (T2), with the parameters $L_1 = (K^+ + K^-)L, L_2 = LK^+, L_3 = LK^-, \gamma = 1 - \alpha, \alpha_1 =$

$\alpha_2 = \alpha_3 = \alpha, K_1 = \frac{k^+}{K^+}$ and $K_2 = \frac{k^-}{K^-}$, where α and L are the exponent and Hölder constant, respectively, of the Hölder-continuous weight function in the definition of CPT.

Let ϵ' be as defined in Theorem 45 with $\tau = \tau_n$ and the other parameters as given above. Note that we may rewrite τ_n as $\tau_n = [1 + c^{-1} \log n] \max\{K_1^{-1}, K_2^{-1}\}$. As a result, we have $cK_i\tau_n \geq \log n$ for $i = 1, 2$. Using these inequalities and substituting for various parameters in the expression for ϵ' in Theorem 45, we get

$$\epsilon' \geq \epsilon - \frac{LK^+}{c\alpha} \exp(-\alpha \log n) - \frac{LK^-}{c\alpha} \exp(-\alpha \log n) \geq \epsilon - \frac{L(K^+ + K^-)}{c\alpha n^\alpha}.$$

It follows from the above inequality and our assumptions on ϵ that ϵ' satisfies the conditions in Theorem 45. Applying Theorem 45 gives $\mathbb{P}(|C_n - C(X)| > \epsilon) \leq g(\epsilon')$, where g represents the right hand side in (29). It is a simple matter to verify that $g(\cdot)$ is decreasing in its argument. Consequently, $g(\epsilon') \leq g\left(\epsilon - \frac{L(K^+ + K^-)}{c\alpha n^\alpha}\right)$. Substituting $\tau = \tau_n$ and the parameters L_1, γ and α_1 in the expression for g immediately yields the required bound. ■

10.6 Proofs of the claims in Section 8

10.6.1 Proof of Theorem 47

Proof The proof is identical to the proof of Theorem 25 except that Lemma 11 is used in place of Lemma 3 there. ■

11. Proof of Section 9

11.1 Proof of Theorem 51

Proof The proof follows by using arguments analogous to that in the proof of Theorem 1 in (Auer et al., 2002).

Let 1 denote the optimal arm, without loss of generality. Also, we abuse notation slightly by denoting $\rho(P_i)$ by $\rho(i)$. First, we bound the number of pulls $T_i(n)$ of any suboptimal arm $i \neq 1$. Fix a round $t \in \{1, \ldots, n\}$ and suppose that a sub-optimal arm i is pulled in this round. Then, we have

$$\rho_{i,T_i(t-1)} - w_{i,T_i(t-1)} \leq \rho_{1,T_1(t-1)} - w_{1,T_1(t-1)}. \tag{76}$$

The LCB-value of arm i can be larger than that of 1 *only if* one of the following three conditions holds:

(1) $\rho_{1,T_1(t-1)}$ is outside the confidence interval, that is,

$$\rho_{1,T_1(t-1)} - w_{1,T_1(t-1)} \geq \rho(1), \tag{77}$$

(2) $\rho_{i,T_i(t-1)}$ is outside the confidence interval, that is,

$$\rho_{i,T_i(t-1)} + w_{i,T_i(t-1)} \leq \rho(i), \tag{78}$$

(3) Gap Δ_i is small: If we negate both the two conditions above and use (76), then we obtain

$$\begin{aligned} &\rho(i) - 2w_{i,T_i(t-1)} \leq \rho_{i,T_i(t-1)} - w_{i,T_i(t-1)} \leq \rho_{1,T_1(t-1)} - w_{1,T_1(t-1)} \leq \rho(1) \\ &\Rightarrow \quad \Delta_i < \ 2w_{i,T_i(t-1)}. \end{aligned} \tag{79}$$

The last condition is equivalent to the following:

$$T_i(t-1) < \frac{(32\sqrt{\sigma^2 \mathsf{e} \log(t)} + 512\sigma)^2 (2L)^{\frac{2}{\kappa}}}{\Delta_i^{\frac{2}{\kappa}}}.$$

Let $u = \dfrac{(32\sqrt{\sigma^2 \mathsf{e} \log n} + 512\sigma)^2 (2L)^{\frac{2}{\kappa}}}{\Delta_i^{\frac{2}{\kappa}}} + 1$. When $T_i(t-1) \geq u$, i.e., when the condition in (79) does not hold, then either (i) arm i is not pulled at time t, or (ii) (77) or (78) occurs. Thus, we have

$$\begin{aligned} T_i(n) &= 1 + \sum_{t=K+1}^{n} \mathbb{I}\left\{I_t = i\right\} \leq u + \sum_{t=u+1}^{n} \mathbb{I}\left\{I_t = i; T_i(t-1) \geq u\right\} \\ &\leq u + \sum_{t=u+1}^{n} \mathbb{I}\left\{\rho_{i,T_i(t-1)} - w_{i,T_i(t-1)} \leq \rho_{1,T_1(t-1)} - w_{1,T_1(t-1)};\ T_i(t-1) \geq u\right\} \\ &\leq u + \sum_{t=1}^{\infty} \sum_{s=1}^{t-1} \sum_{s_i=u}^{t-1} \mathbb{I}\left\{\rho_{i,s_i} - w_{i,s_i} \leq \rho_{1,s} - w_{1,s}\right\} \\ &\leq u + \sum_{t=1}^{\infty} \sum_{s=1}^{t-1} \sum_{s_i=u}^{t-1} \mathbb{I}\left\{\left(\rho(1) < \rho_{1,s} - w_{1,s}\right) \text{ or } \left(\rho(i) > \rho_{i,s_i} + w_{i,s_i}\right) \text{ occurs}\right\}. \end{aligned}$$

Using Theorem 27, we can bound the probability of occurrence of each of the two events inside the indicator on the RHS of the final display above as follows:

$$\begin{aligned} &\mathbb{P}\left(\rho(1) < \rho_{1,s} - w_{1,s}\right) \leq \frac{8}{t^4}, \text{ and} \\ &\mathbb{P}\left(\rho(i) > \rho_{i,s_i} + w_{i,s_i}\right) \leq \frac{8}{t^4}. \end{aligned}$$

Plugging the bounds on the events above and taking expectations on the inequality for $T_i(n)$ derived above, we obtain

$$\mathbb{E}[T_i(n)] \leq u + \sum_{t=1}^{\infty} \sum_{s=1}^{t-1} \sum_{s_i=u}^{t-1} \frac{16}{t^4} \leq u + 16 \sum_{t=1}^{\infty} \frac{1}{t^2} \leq u + \frac{8\pi^2}{3}. \tag{80}$$

The preceding analysis together with the fact that $\mathbb{E}(R_n) = \sum_{i=1}^{K} \Delta_i \mathbb{E}[T_i(n)]$ leads to the first regret bound presented in the theorem.

For inferring the second bound on the regret, i.e., the bound that does not scale inversely with the gaps, observe that

$$\mathbb{E}(R_n) = \sum_i \Delta_i \ \mathbb{E}[T_i(n)]$$

$$
\begin{aligned}
&= \sum_i \left(\Delta_i \, \mathbb{E}[T_i(n)]^{\frac{\kappa}{2}}\right) \left(\mathbb{E}[T_i(n)]^{1-\frac{\kappa}{2}}\right) \\
&\le \left(\sum_i \Delta_i^{2/\kappa} \, \mathbb{E}[T_i(n)]\right)^{\frac{\kappa}{2}} \left(\sum_i \mathbb{E}[T_i(n)]\right)^{1-\frac{\kappa}{2}} && (81)\\
&\le \left(\sum_i \Delta_i^{2/\kappa} \left(\frac{(32\sqrt{\sigma^2 \mathsf{e} \log n} + 512\sigma)^2 (2L)^{\frac{2}{\kappa}}}{\Delta_i^{\frac{2}{\kappa}}} + 1 + \frac{8\pi^2}{3}\right)\right)^{\frac{\kappa}{2}} n^{\frac{2-\kappa}{2}} && (82)\\
&\le \left(K(32\sqrt{\sigma^2 \mathsf{e} \log n} + 512\sigma)^2 (2L)^{\frac{2}{\kappa}} + \sum_i \Delta_i^{2/\kappa}\left(1 + \frac{8\pi^2}{3}\right)\right)^{\frac{\kappa}{2}} n^{\frac{2-\kappa}{2}},
\end{aligned}
$$

where the inequality in (81) follows by applying Hölder's inequality with the conjugate exponents $2/\kappa$ and $2/(2-\kappa)$, and the inequality in (82) follows from (80) and the fact that $\sum_i \mathbb{E}[T_i(n)] = n$. ■

11.2 Proof of Theorem 52

Proof The initial passage of the proof of Theorem 51 upto (79) holds even in the case of Risk-LCB with CPT as the risk measure. However, there are deviations in simplifying the condition $\Delta_i < 2w_{i,T_i(t-1)}$ in (79), and we specify this condition for the case of CPT risk measure below.

$$
\begin{aligned}
\Delta_i &< L(K^+ + K^-) \left[\max\left\{\frac{K^+}{k^+}, \frac{K^-}{k^-}\right\} \left(\sqrt{\log T_i(t-1)} + 1\right)\right]^{1-\alpha} \\
&\quad \times \left[\frac{\sigma(32\sqrt{\mathsf{e}\log t} + 512)}{\sqrt{T_i(t-1)}}\right]^{\alpha} + \frac{2(K^+ + K^-)L}{\alpha(1-\alpha) T_i(t-1)^{\alpha(1-\alpha)}} \\
&\le \frac{L(K^+ + K^-)}{T_i(t-1)^{\alpha \min\left\{\frac{1}{2}, 1-\alpha\right\}}} \left[\max\left\{\frac{K^+}{k^+}, \frac{K^-}{k^-}\right\} \left(\sqrt{\log T_i(t-1)} + 1\right)\right]^{1-\alpha} \\
&\quad \times \left[\sigma(32\sqrt{\mathsf{e}\log t} + 512)\right]^{\alpha} + \frac{2(K^+ + K^-)L}{\alpha(1-\alpha)}. && (83)
\end{aligned}
$$

With $\tilde{c}_3$ and $\tilde{c}_4$ as defined in the theorem statement, the inequality in (83) is equivalent to the following:

$$
\begin{aligned}
&(T_i(t-1))^{\min\left\{\frac{1}{2}, 1-\alpha\right\}} \le \\
&\frac{\left(L(K^+ + K^-) \left[\max\left\{\frac{K^+}{k^+}, \frac{K^-}{k^-}\right\} \left(\sqrt{\log T_i(t-1)} + 1\right)\right]^{1-\alpha} \left[(32\sqrt{\mathsf{e}\log t} + 512)\sigma\right]^{\alpha} + \tilde{c}_3\right)^{\frac{1}{\alpha}}}{\Delta_i^{\frac{1}{\alpha}}}.
\end{aligned}
$$

Then, if a sub-optimal arm i is pulled at time t and (79) is negated, then either (77) or (78) must hold.

Setting $u = \frac{\left[\tilde{c}_4[(32\sqrt{\mathsf{e}\log n}+512)\sigma]^\alpha + \tilde{c}_3\right]^{\frac{1}{\alpha\min\left\{\frac{1}{2},1-\alpha\right\}}}}{\Delta_i^{\frac{1}{\alpha\min\left\{\frac{1}{2},1-\alpha\right\}}}} + 1$, and following the steps leading to (80) in the proof of Theorem 51, we obtain

$$\mathbb{E}[T_i(n)] \leq u + \frac{8\pi^2}{3}. \tag{84}$$

The regret bound in (35) follows by recalling that $\mathbb{E}(R_n) = \sum_{i=1}^K \Delta_i \mathbb{E}[T_i(n)]$.

Using a technique similar to that employed in the proof of the second bound in Theorem 51, we now prove the regret bound in (36) that does not scale inversely with the gaps.

$$\begin{aligned}
\mathbb{E}(R_n) &= \sum_i \Delta_i \,\mathbb{E}[T_i(n)] \\
&= \sum_i \left(\Delta_i \,\mathbb{E}[T_i(n)]^{\alpha\min\left\{\frac{1}{2},1-\alpha\right\}}\right)\left(\mathbb{E}[T_i(n)]^{1-\alpha\min\left\{\frac{1}{2},1-\alpha\right\}}\right) \\
&\leq \left(\sum_i \Delta_i^{\frac{1}{\alpha\min\left\{\frac{1}{2},1-\alpha\right\}}}\,\mathbb{E}[T_i(n)]\right)^{\alpha\min\left\{\frac{1}{2},1-\alpha\right\}}\left(\sum_i \mathbb{E}[T_i(n)]\right)^{1-\alpha\min\left\{\frac{1}{2},1-\alpha\right\}} \qquad (85)\\
&\leq \left[\sum_i \Delta_i^{\frac{1}{\alpha\min\left\{\frac{1}{2},1-\alpha\right\}}}\left[\frac{\left[\tilde{c}_4[(32\sqrt{\mathsf{e}\log n}+512)\sigma]^\alpha + \tilde{c}_3\right]^{\frac{1}{\alpha\min\left\{\frac{1}{2},1-\alpha\right\}}}}{\Delta_i^{\frac{1}{\alpha\min\left\{\frac{1}{2},1-\alpha\right\}}}} + 1 + \frac{8\pi^2}{3}\right]\right]^{\alpha\min\left\{\frac{1}{2},1-\alpha\right\}} \\
&\quad \times n^{1-\alpha\min\left\{\frac{1}{2},1-\alpha\right\}} \qquad (86)\\
&\leq \Bigg[K\left[\tilde{c}_4[(32\sqrt{\mathsf{e}\log n}+512)\sigma]^\alpha + \tilde{c}_3\right]^{\frac{1}{\alpha\min\left\{\frac{1}{2},1-\alpha\right\}}} \\
&\qquad + \sum_i \Delta_i^{\frac{1}{\alpha\min\left\{\frac{1}{2},1-\alpha\right\}}}\left[1 + \frac{8\pi^2}{3}\right]\Bigg]^{\alpha\min\left\{\frac{1}{2},1-\alpha\right\}} \times n^{1-\alpha\min\left\{\frac{1}{2},1-\alpha\right\}},
\end{aligned}$$

where the inequality in (85) follows by applying Hölder's inequality with the conjugate exponents $\frac{1}{\alpha\min\left\{\frac{1}{2},1-\alpha\right\}}$ and $\frac{1}{1-\alpha\min\left\{\frac{1}{2},1-\alpha\right\}}$, and the inequality in (86) follows from (84) and the fact that $\sum_i \mathbb{E}[T_i(n)] = n$. ■

12. Conclusions

We presented a unified approach to derive concentration bounds for empirical estimates of risk measures. Our approach for deriving concentration bounds involves relating the estimation error to the Wasserstein distance between the true CDF and the EDF formed from an i.i.d. sample, and then applying recent concentration bounds for the latter. This approach yields concentration bounds for two general categories of risk measures and two estimation schemes, for a class of distributions which includes sub-Gaussian, sub-exponential, and heavy-tailed distributions. The two categories of risk measures covered by our results contain well known risk measures such as OCE (with CVaR as a special case), spectral risk measures, UBSR, CPT value and RDEU as special cases, while the estimators given in the literature for these risk measures form specific examples of the two estimation schemes covered by our results. Our bounds extend and, in some cases, improve existing bounds for specific risk measures. More importantly, our unified approach contrasts with the case-by-case approaches tried in the literature. We illustrate the usefulness of our bounds by providing an algorithm and the corresponding regret bounds for a stochastic bandit problem involving a risk measure of each category.

References

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

M. Allais. Le comportement de l'homme rationel devant le risque: Critique des postulats et axioms de l'ecole americaine. *Econometrica*, 21:503–546, 1953.

T. M. Apostol. *Mathematical Analysis*. Addison-Wesley, 2nd edition, 1974.

P. Artzner, F. Delbaen, J. Eber, and D. Heath. Coherent measures of risk. *Mathematical finance*, 9 (3):203–228, 1999.

P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2-3):235–256, 2002.

N. C. Barberis. Thirty years of prospect theory in economics: A review and assessment. *Journal of Economic Perspectives*, pages 173–196, 2013.

D. Baudry, R. Gautron, E. Kaufmann, and O. Maillard. Optimal Thompson Sampling strategies for support-aware CVaR bandits. In *International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 716–726, Jul 2021.

A. Ben-Tal and M. Teboulle. Expected utility, penalty functions, and duality in stochastic nonlinear programming. *Management Science*, 32(11):1445–1466, November 1986. ISSN 0025-1909.

A. Ben-Tal and M. Teboulle. An old-new concept of convex risk measures: The optimized certainty equivalent. *Mathematical Finance*, 17:449–476, 02 2007.

D. S. Bernstein. *Scalar, Vector, and Matrix Mathematics: Theory, Facts, and Formulas*. Princeton University Press, 2018. ISBN 9780691151205.

S. P. Bhat and L. A. Prashanth. Concentration of risk measures: A Wasserstein distance approach. In *Advances in Neural Information Processing Systems*, pages 11739–11748, 2019.

E. Boissard. Simple bounds for the convergence of empirical and occupation measures in 1-Wasserstein distance. *Electronic Journal of Probability*, 16:2296–2333, 2011.

F. Bolley, A. Guillin, and C. Villani. Quantitative concentration inequalities for empirical measures on non-compact spaces. *Probability Theory and Related Fields*, 137(3-4):541–593, 2007.

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

A. Cassel, S. Mannor, and A. Zeevi. A general approach to multi-armed bandits under risk criteria. In *Proceedings of the 31st Conference On Learning Theory*, pages 1295–1306, 2018.

J. Cheng, L. A. Prashanth, M. C. Fu, S. I. Marcus, and C. Szepesvári. Stochastic optimization in a cumulative prospect theory framework. *IEEE Transactions on Automatic Control*, 63(9): 2867–2882, 2018.

J. Cotter and K. Dowd. Extreme spectral risk measures: an application to futures clearinghouse margin requirements. *Journal of Banking & Finance*, 30(12):3469–3485, 2006.

D. Denneberg. Distorted probabilities and insurance premiums. *Methods of Operations Research*, 63(3):3–5, 1990.

K. Dowd and D. Blake. After VaR: The theory, estimation and insurance applications of quantile-based risk measures. *The Journal of Risk and Insurance*, 73(2):193–229, 2006.

J Dunkel and S Weber. Stochastic root finding and efficient estimation of convex risk measures. *Operations Research*, 58(5):1505–1521, 2010.

D. A. Edwards. On the Kantorovich–Rubinstein Theorem. *Expositiones Mathematicae*, 29(4): 387–398, 2011.

D. Ellsberg. Risk, ambiguity and the Savage's axioms. *The Quarterly Journal of Economics*, 75(4): 643–669, 1961.

H. Föllmer and A. Schied. Convex measures of risk and trading constraints. *Finance and stochastics*, 6(4):429–447, 2002.

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

N. Fournier and A. Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3-4):707–738, 2015.

N. Galichet. *Contributions to multi-armed bandits: Risk-awareness and sub-sampling for linear contextual bandits*. PhD thesis, Universite Paris Sud Paris XI, 2015.

C. R. Givens and R. M. Shortt. A class of Wasserstein metrics for probability distributions. *Michigan Mathematical Journal*, 31(2):231–240, 1984.

A. Gopalan, L. A. Prashanth, M. C. Fu, and S. I. Marcus. Weighted bandits or: How bandits learn distorted values that are not expected. In *AAAI Conference on Artificial Intelligence*, pages 1941–1947, 2017.

Z. Hu and D. Zhang. Utility-based shortfall risk: Efficient computations via monte carlo. *Naval Research Logistics (NRL)*, 65(5):378–392, 2018.

A. Kagrecha, J. Nair, and K. Jagannathan. Distribution oblivious, risk-aware algorithms for multi-armed bandits with unbounded rewards. In *Advances in Neural Information Processing Systems*, pages 11269–11278, 2019.

D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, pages 263–291, 1979.

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

J. Lee, S. Park, and J. Shin. Learning bounds for risk-sensitive learning. In *Advances in Neural Information Processing Systems*, volume 33, pages 13867–13879, 2020.

J. Lei. Convergence and concentration of empirical measures under Wasserstein distance in unbounded functional spaces. *Bernoulli*, 26(1):767–798, 2020.

J. U. Nair. *Scheduling for heavy-tailed and light-tailed workloads in queueing systems*. PhD thesis, California Institute of Technology, 2012.

A. K. Pandey, L. A. Prashanth, and S. P. Bhat. Estimation of spectral risk measures. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(13):12166–12173, May 2021.

L. A. Prashanth, J. Cheng, M. C. Fu, S. I. Marcus, and C. Szepesvári. Cumulative prospect theory meets reinforcement learning: prediction and control. In *International Conference on Machine Learning*, pages 1406–1415, 2016.

L. A. Prashanth, K. Jagannathan, and R. K. Kolla. Concentration bounds for CVaR estimation: The cases of light-tailed and heavy-tailed distributions. In *International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 5577–5586, Jul 2020.

J. Quiggin. *Generalized Expected Utility Theory: The Rank-dependent Model*. Springer Science & Business Media, 2012.

R. T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2(3): 21–41, 2000.

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

A. Tversky and D. Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323, 1992.

S. S. Vallander. Calculation of the Wasserstein distance between probability distributions on the line. *Theory of Probability and its Applications*, 18(4):784–786, 1974.

R. Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

C. Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.

S. Wang. Insurance pricing and increased limits ratemaking by proportional hazards transforms. *Insurance: Mathematics and Economics*, 17(1):43–54, 1995.

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

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

L. A. Wasserman. *All of Nonparametric Statistics*. Springer, 2015.

J. Weed and F. Bach. Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *Bernoulli*, 25(4A):2620–2648, 2019.