Sample Average Approximation of Risk Functionals: Non-Asymptotic Error Bounds

Guanyu Jin Volker Krätschmer Roger J. A. Laeven

May 24, 2025

Abstract

We study Sample Average Approximation (SAA) for risk-averse stochastic programs, where the risk of an uncertain goal function is measured by a law-invariant convex risk measure. We provide non-asymptotic upper estimation on the probability of absolute deviation error between SAA and the true optimal value under more general assumptions for the goal function and risk measure, as compared to previous works by [1] and [14]. In particular, the estimation is based on conditions from empirical process theory, which allow the goal function to be non-linear, unbounded, and do not require any pathwise analytical properties such as continuity or convexity. Conditions on the risk measures are also readily verifiable in the coherent case, with prominent examples including distortion risk measures and expectiles. Furthermore, our non-asymptotic bound contains explicit constants that can be used to construct confidence regions for the true optimal value and can also be adapted to provide estimates for stochastic programs with inverse S-shaped distortion risk measures.

Keywords: Risk-averse stochastic program, Sample average approximation, Law-invariant risk measures

1. Introduction

Many decision problems in finance, microeconomics and operations research, assume that the decision maker has a preference which is expressed by a real-valued *risk measure* ρ, in the sense that a random loss X is preferred over Y, if and only if $\rho(X) \leq \rho(Y)$. An optimal decision can therefore be formulated as a solution of the following risk minimization problem:

$$\inf_{\theta\in\Theta} \rho\big(G(\theta, Z)\big), \tag{1.1}$$

where $\Theta \subset \mathbb{R}^m$ is a compact space of decisions, Z is a d-dimensional random vector with distribution $\mathbb{P}^Z$, and $G(\theta, Z)$ is a goal function. Depending on the choice of ρ, Θ, G, problem (1.1) encompasses a wide range of decision problems, including portfolio optimization, inventory management, as well as learning problems such as least squares regressions.

A simple example of ρ is the expected value, in which case the decision maker is assumed to be *risk-neutral*. This assumption is, however, too idealistic, and one is often more inclined to work with a ρ that accounts for *risk-aversion*. Besides expected value, other popular choices of ρ can be the expected utility evaluation, optimized certainty equivalent, or distortion risk measures. Therefore, we assume that ρ is a member of the class of convex *law-invariant* risk measures, which includes all aforementioned examples and more.

In general, the parameterized distribution of the goal function G is unknown, and the only available information must be inferred from an i.i.d. sample, formally expressed by a sequence $(Z_j)_{j\in\mathbb{N}}$ of independent d-dimensional random vectors on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ that are identically distributed as the d-dimensional random vector Z. By law-invariance, we may associate ρ with a functional $\mathcal{R}_\rho$ on sets of distribution functions. In this case, (1.1) reads as follows:

$$\inf_{\theta\in\Theta} \mathcal{R}_\rho(F_\theta),$$

where F_θ is the distribution function of $G(\theta, Z)$. Based on the i.i.d. sample $(Z_1, \cdots, Z_n)$, we may replace any distribution function F_θ with its empirical counterpart $\widehat{F}_{n,\theta}$, defined by,

$$\widehat{F}_{n,\theta}(t) := \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_{(-\infty,t]}\big(G(\theta, Z_j)\big).$$

Then the original optimization problem (1.1) may be approximated by the following one, which is generally known as the Sample Average Approximation (SAA):

$$\inf_{\theta\in\Theta} \mathcal{R}_\rho(\widehat{F}_{n,\theta}) \quad (n \in \mathbb{N}). \tag{1.2}$$

The interesting and most relevant question is certainly the following: how good is the SAA approximation? Since (1.2) depends on each realization of the i.i.d. samples, one way to answer this question is to examine the following deviation probability:

$$\mathbb{P}\Big(\Big\{\big|\inf_{\theta\in\Theta} \mathcal{R}_\rho\big(\widehat{F}_{n,\theta}\big) - \inf_{\theta\in\Theta} \mathcal{R}_\rho\big(F_\theta\big)\big| \geq \epsilon\Big\}\Big) \quad (n \in \mathbb{N}, \epsilon > 0). \tag{1.3}$$

To the best of our knowledge, there are currently no universal approaches on how to obtain explicit non-asymptotic upper bounds for the deviation probability (1.3), for general law-invariant convex risk measures and non-linear goal functions G (see Section 1.1 for a discussion on related literature that have studied (1.3) for only subclasses of ρ and G). From an application perspective, explicit non-asymptotic upper bounds for (1.3) are crucial for constructing confidence intervals to quantify the error of the sample average approximation of the optimal value. The main goal of this paper is to fill this gap in the literature on non-asymptotic upper bounds for (1.3). In summary, our contributions are as follows.

- We provide a universal approach for obtaining non-asymptotic upper estimation on the deviation probability (1.3) that (i) has explicit constants, holds for (ii) a general class of convex law-invariant risk measures, (iii) and allows for non-linear and unbounded goal functions. Moreover, our bound on (1.3) is sharper than the one provided in [1], and requires fewer assumptions.

- In particular, we show that our non-asymptotic upper bound on (1.3) can be made more explicit for the class of distortion risk measures and expectiles. Moreover, our estimation methodology can also be adapted to provide upper bound on (1.3) for rank-dependent utility risk measures, where the distortion function has an inverse S-shaped.

In the following subsection, we review some of the existing literature.

1.1. Related literature

Sample average approximation for risk-neutral stochastic programs has been well studied in the literature, where a general survey can be found in [15]. Consistency and asymptotic normality for both the optimal values and solutions are established in works such as [22, 23, 21]. Non-asymptotic lower and upper confidence bounds on the deviation probability (1.3) for risk-neutral stochastic programs have been provided by [9], under the assumption that the goal function G is convex in the decision variable and additional sub-Gaussian conditions are imposed on the distributions of $G(\theta, Z_1) - \mathbb{E}_{\mathbb{P}}[G(\theta, Z_1)]$. On the other hand, [14] was able to provide non-asymptotic upper estimates on (1.3) for more general goal functions G, but not for a lower estimate.

The most related works on non-asymptotic upper bounds for SAA of risk-averse stochastic programs are [1] and [14]. [1] provides non-asymptotic convergence rates for SAA of stochastic programs with general convex risk measures, both for the deviation probability and the expected deviation. However, their bounds do not have explicit constants, and the goal function G in the stochastic program is restricted to the linear cases, and uniformly bounded in $\theta \in \Theta$ when estimating the deviation probability. By contrast, the upper bounds provided by [14] do allow for nonlinear, unbounded goal functions and have explicit constants, but are only available for stochastic programs where the risk measures are expected loss, upper semi-deviations, and divergence risk measures. Our work significantly extends the results of these two papers by establishing non-asymptotic upper bounds for SAA of stochastic programs with (i) general risk functionals encompassing convex risk measures and rank-dependent utility functionals with inverse S-shaped probability weighting functions, where (ii) the goal functions may be nonlinear and unbounded, and (iii) the bounds feature explicit constants. In addition, when the goal functions G are uniformly bounded, our non-asymptotic upper bound yields convergence rates that are tighter than [1]. As an alternative to SAA, recent work by [4] also investigated the approximation of risk-averse stochastic programs using the stochastic gradient Langevin dynamics method. Under the assumption that the goal function is bounded, Lipschitz continuous, with a bounded gradient, they obtained explicit non-asymptotic upper bounds on the mean squared approximation error that decay with a rate of $1/n$ to the remaining positive constants that depend on the choice of the parameters. As compared to our work, this upper bound on the mean squared error is a more conservative moment bound for the deviation probability (1.3) in the sense that our bound decays exponentially with n to zero. However, we do note that their bound is independent of the dimension of the decision space Θ.

Asymptotic approximations of SAA for risk-averse stochastic programs have also been studied by [13], in the case of expected loss, upper semi-deviations, and divergence risk measures. A central limit theorem for risk-averse stochastic programs has also been established by [10], for risk measures that have a discrete Kusuoka [16] representation. A central limit theorem for law-invariant coherent risk measures in the non-optimization setting has been studied by [2] and [18].

1.2. Outline

The outline of the paper is organized as follows. Section 2 summarizes the main assumptions that are made in this work and introduces the necessary preliminaries. Section 3 presents the main result, which is an explicit non-asymptotic upper bound on the deviation probability (1.3) under the assumption that G is uniformly bounded. Moreover, we also discuss the assumption (A 4) imposed on the risk measures that is made throughout this paper and provide a more explicit characterization for it in some canonical examples of risk measures. Section 4 shows the application of our main result in Section 3 to the estimation of deviation probability for distortion risk measures with an inverse s-shaped distortion function, which is a major class of non-convex risk measures that are extensively used in behavioral economics (see [28]). Section 5 extends the main result, namely Theorem 3.1, to unbounded goal functions. Finally, the proofs of the main results are provided in Section 7.

2. Set up

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a fixed atomless complete probability space, and let $L^p(\Omega, \mathcal{F}, \mathbb{P})$ denote the usual L^p-space on $(\Omega, \mathcal{F}, \mathbb{P})$ $(p \in [1, \infty])$, where we tacitly identify random variables which are different on $\mathbb{P}$-null sets only. Furthermore, $\mathcal{X}$ stands for some $\mathbb{R}$-vector space of $\mathbb{P}$-integrable random variables, enclosing all $\mathbb{P}$-essentially bounded random variables. The vector space $\mathcal{X}$ will be equipped with a norm $\|\cdot\|_{\mathcal{X}}$ and the $\mathbb{P}$-a.s. order $\succeq_{\mathbb{P}}$ such that $(\mathcal{X}, \|\cdot\|_{\mathcal{X}}, \succeq_{\mathbb{Q}})$ is a Banach lattice which is solid, i.e., $X \in \mathcal{X}$ if $|Y| \geq |X|$ for some $Y \in \mathcal{X}$, and rearrangement invariant, meaning that a random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ belongs to $\mathcal{X}$ whenever it has the same distribution as some $Y \in \mathcal{X}$. In addition, we assume

$$\lim_{k \to \infty} \|X - X \wedge k\|_{\mathcal{X}} = 0, \quad \text{for nonnegative } X \in \mathcal{X}. \tag{2.1}$$

Prominent examples are provided by $L^p(\Omega, \mathcal{F}, \mathbb{P})$ with the usual L^p-norm for $p \in [1, \infty]$.

Next, let us fix any law-invariant convex risk measure $\rho : \mathcal{X} \to \mathbb{R}$ which by definition is a convex mapping satisfying in addition the following properties.

- **monotonicity:**

$$\rho(X) \leq \rho(Y) \quad \text{for } X, Y \in \mathcal{X} \text{ with } X \leq Y \ \mathbb{P}-\text{a.s.},$$

- **cash-invariance:**

$$\rho(X + c) = \rho(X) + c \quad \text{for } X \in \mathcal{X}, c \in \mathbb{R},$$

- **law-invariance:**

$$\rho(X) = \rho(Y) \quad \text{for indentically distributed } X, Y \in \mathcal{X}.$$

If in addition ρ sublinear, then ρ is called a law-invariant coherent risk measure. Throughout this paper we also impose the following additional property for ρ.

- **Lebesgue property**

$$\lim_{k \to \infty} \rho(X_k) = \rho(X)$$

  for any sequence $(X_k)_{k \in \mathbb{N}}$ in $\mathcal{X}$ which is uniformly bounded w.r.t. the $\mathbb{P}$-essential sup-norm and converges $\mathbb{P}$-a.s. to $X \in \mathcal{X}$.

The Lebesgue property is always fulfilled if $\mathcal{X}$ contains at least one random variable which is not $\mathbb{P}$-essentially bounded (see [6, Theorem 3]).

The outstanding example of a law-invariant convex risk measure with Lebesgue property is provided by the Average Value at Risk, which we recall now.

Example 2.1 *Let $\alpha \in (0,1)$. Then the mapping $\rho = AV@R_\alpha : L^1(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$, defined by*

$$AV@R_\alpha(X) = \frac{1}{1-\alpha} \int_{[\alpha,1)} F_X^{\leftarrow}(u) \, du,$$

where $F_X^{\leftarrow}$ denotes the left-continuous quantile function of the distribution function F_X, is called the Average Value at Risk w.r.t. α (e.g. [7], [23]). It has the following useful representation

$$AV@R_\alpha(X) = \inf_{x \in \mathbb{R}} \mathbb{E}\left[\frac{(X-x)^+}{1-\alpha} + x\right], \quad \textit{for } X \in L^1(\Omega, \mathcal{F}, \mathbb{P}),$$

(see e.g., [12]). In particular, $AV@R_\alpha$ may be identified as a law-invariant coherent risk measure.

The Average Value at Risk is the building block of law-invariant convex risk measures like ρ due to the *Kusuoka representation*, given by

$$\rho(X) = \sup_{\mu \in \mathcal{M}} \left(\int_{[0,1)} AV@R_\alpha(X) \, \mu(d\alpha) - \beta(\mu) \right), \tag{2.2}$$

where we define $AV@R_0[X] = \mathbb{E}[X]$, $\mathcal{M}$ is some set of Borel probability measures on $[0,1)$, and $\beta : \mathcal{M} \to (-\infty, \infty]$ denotes a penalty function with effective domain $\mathrm{dom}(\beta)$ (see Proposition A.1 in Appendix A). The Kusuoka representation reveals that the law-invariant convex risk measure ρ is a risk-averse functional in the sense that it is non-decreasing w.r.t. increasing convex order.

We shall restrict ourselves to mappings G that satisfy the following properties:

(A 1) $G(\theta, \cdot)$ is Borel measurable for every $\theta \in \Theta$.

(A 2) There exists a strictly positive Borel measurable mapping $\xi : \mathbb{R}^d \to \mathbb{R}$ such that $\xi(Z_1) \in \mathcal{X}$

$$\sup_{\theta \in \Theta} |G(\theta, z)| \leq \xi(z) \quad \text{for } z \in \mathbb{R}^d.$$

(A 3) There exists an at most countable subset $\bar{\Theta} \subset \Theta$ and $(\mathbb{P}^z)^n$-null sets N_n such that

$$\inf_{\theta \in \bar{\Theta}} \mathbb{E}\left[|G(\theta, Z_1) - G(\tilde{\theta}, Z_1)|\right] = \inf_{\theta \in \bar{\Theta}} \max_{j \in \{1,\dots,n\}} |G(\theta, z_j) - G(\tilde{\theta}, z_j)| = 0,$$

for $n \in \mathbb{N}$, $\tilde{\theta} \in \Theta$, and $(z_1, \dots, z_n) \in \mathbb{R}^{dn} \setminus N_n$.

(A 1) and (A 2) are standard conditions imposed on function classes. (A 3) is used for establishing measurability for (1.3) (see e.g., Lemma 7.1).

Convenient ways to find upper bounds of the deviation probabilities (1.3) may be provided by general devices from empirical process theory, which are based on covering numbers for classes of Borel measurable mappings from $\mathbb{R}^d$ into $\mathbb{R}$ w.r.t. L^p-norms. To recall these concepts adapted to our situation, let us fix any nonvoid set $\mathbb{F}$ of Borel measurable mappings from $\mathbb{R}^d$ into $\mathbb{R}$ and any probability measure $\mathbb{Q}$ on $\mathcal{B}(\mathbb{R}^d)$ with metric $d_{\mathbb{Q},p}$ induced by the L^p-norm $\|\cdot\|_{\mathbb{Q},p}$ for $p \in [1, \infty)$.

- *Covering numbers for* $\mathbb{F}$
  We use $N\big(\eta, \mathbb{F}, L^p(\mathbb{Q})\big)$ to denote the minimal number to cover $\mathbb{F}$ by closed $d_{\mathbb{Q},p}$-balls of radius $\eta > 0$ with centers in $\mathbb{F}$. We define $N\big(\eta, \mathbb{F}, L^p(\mathbb{Q})\big) := \infty$ if no finite cover is available.

- An *envelope* of $\mathbb{F}$ is defined as some Borel measurable mapping $C_{\mathbb{F}}$ from $\mathbb{R}^d$ into $\mathbb{R}$ satisfying $\sup_{h \in \mathbb{F}} |h| \leq C_{\mathbb{F}}$. If an envelope $C_{\mathbb{F}}$ has strictly positive outcomes, we shall speak of a *positive envelope*.

- $\mathcal{M}_{\text{fin}}$ denotes the set of all probability measures on $\mathcal{B}(\mathbb{R}^d)$ with finite support.

For abbreviation, let us introduce for a class $\mathbb{F}$ of Borel measurable functions from $\mathbb{R}^d$ into $\mathbb{R}$ with an arbitrary positive envelope $C_{\mathbb{F}}$ of $\mathbb{F}$ the following notation:

$$J(\mathbb{F}, C_{\mathbb{F}}, \delta) := \int_0^\delta \sup_{\mathbb{Q} \in \mathcal{M}_{\text{fin}}} \sqrt{\log\big(2N\big(\varepsilon\, \|C_{\mathbb{F}}\|_{\mathbb{Q},2}, \mathbb{F}, L^2(\mathbb{Q})\big)\big)}\, d\varepsilon. \tag{2.3}$$

For our purposes, the following function class is the most relevant one:

$$\mathbb{F}^\Theta := \{G(\theta, \cdot) \mid \theta \in \Theta\}. \tag{2.4}$$

Examples of function classes with explicit upper bounds on $J(\mathbb{F}, C_{\mathbb{F}}, \delta)$ are provided in Section 2 of [14]. We state here one of the examples, namely if $G(\theta, z)$ satisfies the Hölder condition for $\theta \in \mathbb{R}^m, z \in \mathbb{R}^d$: i.e., there exists a $\beta \in (0,1)$, and a square $\mathbb{P}^Z$-integrable strictly positive mapping $C : \mathbb{R}^d \to (0, \infty)$, such that for $z \in \mathbb{R}^d, \theta_1, \theta_2 \in \Theta$:

$$|G(\theta_1, z) - G(\theta_2, z)| \leq C(z)\|\theta_1 - \theta_2\|_2^\beta,$$

then, under some mild conditions imposed on G (see Proposition 2.6 of [14]), one has for $\delta \in (0, 1/2]$,

$$J(\mathbb{F}^\Theta, \xi, \delta) \leq 2\delta\sqrt{(3m+1)\ln(2) + \frac{m}{\beta}\ln(2/\delta)},$$

where $\xi \equiv C\Delta(\Theta)^\beta + |G(\bar{\theta}, .)|$ is an envelope function for $\mathbb{F}^\Theta$, $\Delta(\Theta)$ denotes the diameter of Θ, and $\bar{\theta} \in \Theta$ is a point where G is square $\mathbb{P}^Z$-integrable.

3. Deviation probabilities

Throughout this section, we restrict ourselves to law-invariant convex risk measures ρ which are normalized, meaning $\rho(0) = 0$, and goal functions $G(\theta, z)$ that are uniformly bounded in θ, i.e., (A 2) is satisfied with $\xi \equiv B$ for some $0 < B < \infty$. By monotonicity of the Average Value at Risk, we may conclude the inequality $|AV@R_\alpha(X)| \leq \|X\|_\infty$ for any $\alpha \in (0,1)$ and every $\mathbb{P}$-essentially bounded random variable with essential sup-norm $\|X\|_\infty$. Hence, we may associate ρ with the function

$$\beta_\rho : \mathcal{M}\big([0,1)\big) \to \mathbb{R} \cup \{\infty\},\ \mu \mapsto \sup_{\substack{X \in L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \\ \rho(X) \leq 0}} \int_{[0,1)} AV@R_\alpha(X)\ \mu(d\alpha) \tag{3.1}$$

on the set $\mathcal{M}\big([0,1)\big)$ of all Borel probability measures on $[0,1)$, where $AV@R_0$ stands for the expectation. If ρ satisfies the Lebesgue property, then it has the following representation

$$\rho(X) = \sup_{\mu \in \mathcal{M}([0,1))} \left(\int_{[0,1)} AV@R_\alpha(X)\ \mu(d\alpha) - \beta_\rho(\mu) \right) \quad \text{for } X \in L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \tag{3.2}$$

(see [7, Theorem 4.62]). We impose the following condition on the representation (3.2).

(A 4) There exists some $q \in [1, \infty)$ such that

$$N_{q,b} := \sup_{t \in (0,1]} \sup_{\substack{\mu \in \mathcal{M}([0,1)) \\ \beta_\rho(\mu) \leq b}} \frac{\int_{[0,1)} \frac{t \wedge (1-\alpha)}{1-\alpha}\ \mu(d\alpha)}{t^{1/q}} < \infty \quad \text{for } b > 0.$$

We shall show the following result concerning the deviation probabilities (1.3).

Theorem 3.1 *Let ρ be a normalized law-invariant convex risk measure which satisfies the Lebesgue property, and let assumptions (A 1), (A 2), (A 3) and (A 4) be fulfilled with $q \in [1,\infty)$, $\xi \equiv B \in \mathbb{R}$, and constants $N_{q,b} \in \mathbb{R}$ ($b > 0$) as in (A 4). If $J(\mathbb{F}^{\Theta}, B, 1/8)$ is finite, then the mapping*

$$\inf_{\theta\in\Theta} \mathcal{R}_\rho(\widehat{F}_{n,\theta}) - \inf_{\theta\in\Theta} \mathcal{R}_\rho(F_\theta)$$

is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ such that

$$\begin{aligned}\mathbb{P}\left(\left\{\left|\inf_{\theta\in\Theta} \mathcal{R}_\rho(\widehat{F}_{n,\theta}) - \inf_{\theta\in\Theta} \mathcal{R}_\rho(F_\theta)\right| \geq \varepsilon\right\}\right) &\leq \mathbb{P}^*\left(\sup_{\theta\in\Theta}\left|\mathcal{R}_\rho(\widehat{F}_{n,\theta}) - \mathcal{R}_\rho(F_\theta)\right| \geq \epsilon\right) \\ &\leq \exp\left(-\frac{n\ t^2\ \varepsilon^{2q}}{2^{4q+1}\ (t+1)^2\ B^{2q}\ N_{q,(4B+\delta)}^{2q}}\right)\end{aligned}$$

holds for $t, \delta > 0$ whenever

$$\varepsilon > N_{q,(4B+\delta)} \cdot B\ [2^{2q-1}\ 128\ (t+1)\ \sqrt{2}]^{1/q}\ \left[J\big(\mathbb{F}^{\Theta}, B, 1/8\big) + 3/2\right]^{1/q} / n^{1/(2q)}.$$

Here $\mathbb{P}^*$ denotes the outer probability w.r.t. $\mathbb{P}$. Moreover, the mapping

$$\sup_{\theta\in\Theta}\left|\mathcal{R}_\rho(\widehat{F}_{n,\theta}) - \mathcal{R}_\rho(F_\theta)\right|$$

is also a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ if in addition $G(\cdot, z)$ is lower semicontinuous for $z \in \mathbb{R}^d$.

The proof may be found in Subsection 7.1.

Condition (A 4) may be simplified if ρ is coherent. In this case, as an important tool, let us introduce the auxiliary mapping

$$\overline{h}_\rho : [0,1] \to \mathbb{R},\ t \mapsto \rho\big(\mathbb{1}_{([1-t],1)}(U)\big),$$

where U denotes any random variable on the atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$ which is uniformly distributed on $(0,1)$. It is known that for coherent ρ the function β_ρ vanishes on its effective domain due to positive homogeneity of ρ. Hence, (A 4) reads as follows.

(A 5) There exists some $q \in [1,\infty)$ such that

$$\overline{N}_q := \sup_{t\in(0,1]} \frac{\overline{h}_\rho(t)}{t^{1/q}} < \infty.$$

Under (A 5) we obtain immediately the following version of Theorem 3.1

Theorem 3.2 *Let ρ be a law-invariant coherent risk measure which satisfies the Lebesgue property, and let assumptions (A 1), (A 2), (A 3) and (A 5) be fulfilled with $q \in [1, \infty)$, $\xi \equiv B \in \mathbb{R}$, and constant $\overline{N}_q \in \mathbb{R}$ introduced in (A 5). If $J(\mathbb{F}^\Theta, B, 1/8)$ is finite, then,*

$$\mathbb{P}\left(\left\{\left|\inf_{\theta\in\Theta} \mathcal{R}_\rho(\widehat{F}_{n,\theta}) - \inf_{\theta\in\Theta} \mathcal{R}_\rho(F_\theta)\right| \geq \varepsilon\right\}\right) \leq \mathbb{P}^*\left(\sup_{\theta\in\Theta}\left|\mathcal{R}_\rho(\widehat{F}_{n,\theta}) - \mathcal{R}_\rho(F_\theta)\right| \geq \epsilon\right)$$
$$\leq \exp\left(-\frac{n\ t^2\ \varepsilon^{2q}}{2^{4q-1}\ (t+1)^2\ B^{2q}\ \overline{N}_q^{2q}}\right)$$

holds for $t > 0$ if

$$\varepsilon > \overline{N}_q \cdot B\ [2^{2q-1}\ 128\ (t+1)\ \sqrt{2}]^{1/q}\ \left[J\big(\mathbb{F}^\Theta, B, 1/8\big) + 3/2\right]^{1/q}/n^{1/(2q)}.$$

Here $\mathbb{P}^*$ denotes the outer probability w.r.t. $\mathbb{P}$.

We want to illustrate condition (A 5) with some examples. As a first example, we shall consider the class of concave distortion risk measures.

Example 3.3 *A concave distortion function is a concave non-decreasing mapping $h : [0,1] \to [0,1]$ satisfying $h(0) = 0$ and $h(1) = 1$. Every such mapping h induces via*

$$\rho_h(X) := \int_{-\infty}^{0} \big[h(\mathbb{P}(X > x)) - 1\big]\ dx + \int_0^\infty h(\mathbb{P}(X > x))\ dx$$

a coherent risk measure on $\mathcal{X}$ consisting of all random variables X on $(\Omega, \mathcal{F}, \mathbb{P})$ satisfying

$$\|X\|_h := \int_0^\infty h(\mathbb{P}(|X| > x))\ dx < \infty \tag{3.3}$$

(see e.g. [5] or [7]). This is because ρ_h may be viewed as a Choquet integral w.r.t. the submodular set function $h(\mathbb{P})$.

The set $\mathcal{X}$ is a vector space, and the mapping $\|\cdot\|_h$, defined by (3.3)*, is a norm on $\mathcal{X}$ (see [5, Proposition 9.4]). Moreover if h is continuous, then $\mathcal{X}$, endowed with $\|\cdot\|_h$ and the $\mathbb{P}$-a.s. order $\succeq_{\mathbb{P}}$ is a solid, rearrangement invariant Banach lattice meeting property* (2.1) *(see [5, Proposition 9.5 with Theorem 8.9]). Continuity of h also implies the Lebesgue property (see [5, Theorem 8.9]).*

Easy calculation reveals that the auxiliary function $\overline{h}_{\rho_h}$ coincides with the distortion function h. Hence (A 5) is fulfilled iff $\sup_{t\in(0,1]} h(t)/t^{1/q} < \infty$ for some $q \in [1, \infty)$. In Table 1, we provide examples of distortion functions with upper bounds for the terms $\overline{N}_q$ and the range $rg(q)$ of q such that (A 5) holds.

Next, we shall focus on expectiles, genuinely introduced in the paper [27].

Example 3.4 *For $\alpha \in (0,1)$, the mapping*

$$\rho^\alpha : L^2(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R},\ X \mapsto \operatorname*{argmin}_{x\in\mathbb{R}} \left[\alpha\|(X-x)^+\|_2^2 + (1-\alpha)\|(X-x)^-\|_2^2\right]$$

Distortion Family	$h(p),\ p \in [0,1]$	$\overline{N}_q$	rg(q)
Proportional Hazard	$p^a,\ a \in (0,1)$	1	$q \geq 1/a$
Gini Principles	$(1+a)p - ap^2,\ a \in (0,1)$	1 + a	$q \geq 1$
Dual Moments	$1-(1-p)^k,\ k > 1$	k	$q \geq 1$
MAXMINVAR	$(1-(1-p)^n)^{1/k},\ k > 1$	$k^{1/k}$	$q \geq k$
LB-transform	$p^a(1-\log(p^a)),\ a \in (0,1)$	$\exp\left(\frac{-a+1/q}{a(aq-1)}\right)\frac{aq}{aq-1}$	$q > 1/a$

Table 1: Examples of distortion functions with upper bounds on $\overline{N}_q$ and the range $rg(q)$ of q such that (A 5) holds.

is well-defined and known as the expectile w.r.t. α. It has been shown in [3] that it is a law-invariant coherent risk measure for any $\alpha \in [1/2, 1)$. Fixing $\alpha \in [1/2, 1)$, the associated auxiliary function $\overline{h}_{\rho^\alpha}$ of ρ^α satisfies

$$\overline{h}_{\rho^\alpha}(t) = \frac{\alpha t}{1-\alpha+t(2\alpha-1)}.$$

In particular, since $\alpha \geq 1/2$,

$$\overline{N}_q := \sup_{t\in(0,1]} \frac{\overline{h}_{\rho^\alpha}(t)}{t^{1/q}} \leq \sup_{t\in(0,1]} \frac{\alpha}{1-\alpha+t(2\alpha-1)} = \frac{\alpha}{1-\alpha}.$$

Let us now comment on assumption (A 4) for general convex ρ.

Remark 3.5 *Let $q \in [1,\infty)$ such that representation* (3.2) *satisfies*

$$\overline{M}_q := \sup_{\mu\in\mathcal{M}([0,1))} \left(\int_{[0,1)} \frac{1}{(1-\alpha)^{1/q}}\ \mu(d\alpha) - \beta_\rho(\mu)\right) < \infty. \tag{3.4}$$

Then for any μ from the effective domain of β_ρ and every $t \in [0,1]$

$$\int_{[0,1)} \frac{t\wedge(1-\alpha)}{1-\alpha}\ \mu(d\alpha) \leq t^{1/q} \int_{[0,1)} \frac{1}{(1-\alpha)^{1/q}}\ \mu(d\alpha) = t^{1/q}\ (\overline{M}_q + \beta_\rho(\mu)).$$

Hence (A 4) is satisfied with $N_{q,b} = \overline{M}_q + b$ for $b > 0$.

A criterion to ensure (3.4) *for $q \in (1,\infty)$ is provided by the property of q-regularity, which was introduced in [1]. By definition, ρ is called q-regular with $q \in (1,\infty)$ if for any random variable X_q on $(\Omega,\mathcal{F},\mathbb{P})$ which is Pareto-distributed with location parameter* 1 *and scale parameter q*

$$\widehat{M}_q := \sup_{k\in\mathbb{N}} \rho(X_q \wedge k) < \infty.$$

In terms of the representation (3.2)*, the application of the monotone convergence theorem yields*

$$\begin{aligned}
\widehat{M}_q &= \sup_{\mu\in\mathcal{M}([0,1))}\left(\int_{[0,1)} AV@R_\alpha(X_q)\ \mu(d\alpha)-\beta_\rho(\mu)\right)\\
&= \sup_{\mu\in\mathcal{M}([0,1))}\left(\int_{[0,1)} \frac{q}{q-1}\ \frac{1}{(1-\alpha)^{1/q}}\ \mu(d\alpha)-\beta_\rho(\mu)\right)\\
&= \frac{q}{q-1}\ \sup_{\mu\in\mathcal{M}([0,1))}\left(\int_{[0,1)} \frac{1}{(1-\alpha)^{1/q}}\ \mu(d\alpha)-\frac{q-1}{q}\ \beta_\rho(\mu)\right).
\end{aligned}$$

Finally, $0=\rho(0)=\sup_{\mu\in\mathcal{M}([0,1))}[-\beta_\rho(\mu)]$ *and therefore* β_ρ *is nonnegative. Hence* (3.4) *is fulfilled with* $\overline{M}_q\leq(q-1)\ \widehat{M}_q/q$*.*

As an application of Remark 3.5, we shall look at utility-based shortfall risk measures.

Example 3.6 *Let* $l:\mathbb{R}\to\mathbb{R}$ *denote a non-constant, non-decreasing convex function, and* $x_0:=l(0)$ *be an element from the topological interior of the range of* l*. Furthermore, it is assumed that there are* $N_{(l)}>0$ *and* $p\in[1,\infty)$ *such that* $l(x)\leq N_{(l)}\ (x^+)^p$ *holds for* $x\in\mathbb{R}$*. Then*

$$\rho_{l,x_0}:L^p(\Omega,\mathcal{F},\mathbb{P})\to\mathbb{R},\ X\mapsto\inf\left\{x\in\mathbb{R}\mid\mathbb{E}[l(X-x)]\leq x_0\right\}$$

defines a normalized law-invariant convex risk measure which satisfies the Lebesgue property. As a convex risk measure on a Banach lattice it is also continuous w.r.t. the L^p*-norm (see [20, Proposition 3.1]).*

For any $q\in(p,\infty)$ *every random variable* X_q *on* $(\Omega,\mathcal{F},\mathbb{P})$ *which is Pareto-distributed with location parameter* 1 *and scale parameter* q *belongs to* $L^p(\Omega,\mathcal{F},\mathbb{P})$*. In particular,* $\sup_{k\in\mathbb{N}}\rho_{l,x_0}(X_q\wedge k)=\rho_{l,x_0}(X_q)$ *for* $q\in(p,\infty)$ *due to* L^p*-norm continuity of* ρ_{l,x_0}*. According to Remark 3.5, condition (A 4) is fulfilled for every* $q\in(p,\infty)$*, where the corresponding constants* $\overline{M}_q$ *satisfy the following upper estimation*

$$\overline{M}_q\leq\frac{q-1}{q}\ \rho_{l,x_0}(X_q)\leq\frac{q-1}{q}\ \inf\Big\{x\in\mathbb{R}\mid\mathbb{E}\big[\big((X_q-x)^+\big)^p\big]\leq N_{(l)}\ x_0\Big\}.$$

As a consequence of Theorems 3.1, 3.2, we may also provide the following simple criterion for tightness rates of the sequence

$$\left(\inf_{\theta\in\Theta}\mathcal{R}_\rho(\widehat{F}_{n,\theta})-\inf_{\theta\in\Theta}\mathcal{R}_\rho(F_\theta)\right)_{n\in\mathbb{N}}. \tag{3.5}$$

Theorem 3.7 *Let* ρ *be a law-invariant convex risk measure which satisfies the Lebesgue property, and let assumptions (A 1), (A 2), (A 3) be fulfilled with* $\xi\equiv B\in\mathbb{R}$ *and* $J(\mathbb{F}^\Theta,B,1/8)<\infty$*. If* ρ *satisfies (A 4)* $q\in[1,\infty)$*, or if* ρ *is coherent and meets (A 5) with the same* q*, then the sequence*

$$\left(n^{1/(2q)}\left[\inf_{\theta\in\Theta}\mathcal{R}_\rho(\widehat{F}_{n,\theta})-\inf_{\theta\in\Theta}\mathcal{R}_\rho(F_\theta)\right]\right)_{n\in\mathbb{N}}$$

is uniformly tight.

Proof Let $q \in [1, \infty)$ such that ρ satisfies (A 4) with q, or (A 5) with q if ρ is coherent. Then applying Theorem 3.1 or Theorem 3.2 both with $t = 1$, there is some $\widehat{N}_q > 0$ such that

$$\mathbb{P}\left(\left\{\left|\inf_{\theta\in\Theta} \mathcal{R}_\rho(\widehat{F}_{n,\theta}) - \inf_{\theta\in\Theta} \mathcal{R}_\rho(F_\theta)\right| \geq \frac{\varepsilon}{n^{1/(2q)}}\right\}\right) \leq \exp\left(-\frac{\varepsilon^{2q}}{2^{4q-1}\ 4\ B^{2q}\ \widehat{N}_q^{2q}}\right)$$

holds for every $\varepsilon > \widehat{N}_q \cdot B\ [2^{2q-1}\ 256\ \sqrt{2}]^{1/q}\ \left[J\big(\mathbb{F}^\Theta, B, 1/8\big) + 3/2\right]^{1/q}$ and any $n \in \mathbb{N}$. In particular

$$\lim_{\varepsilon\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\left\{n^{1/(2q)}\left|\inf_{\theta\in\Theta} \mathcal{R}_\rho(\widehat{F}_{n,\theta}) - \inf_{\theta\in\Theta} \mathcal{R}_\rho(F_\theta)\right| \geq \varepsilon\right\}\right) = 0$$

which completes the proof. □

Theorem 3.7 shows a clear improvement in convergence rate made in Theorem 3.1 as compared to Theorem 5.1 in [1], which only holds for linear goal functions and $q > 1$. Indeed, Theorem 3.1 allows for $q = 1$, which by Theorem 3.7 shows that uniform tightness can be achieved at a $n^{1/2}$ rate for risk measures such as the Dual Moments distortion risk measures as in Table 1. This improvement is not only for risk measures that satisfy (A 4) for $q = 1$. For example, the MAXMINVAR distortion risk measure of order $k > 1$ in Table 1 satisfies the condition of Theorem 5.1 in [1] only for $q > k$, whereas Table 1 shows that Theorem 3.2 also holds for MAXMINVAR with $q = k$. As we elaborated in Remark 3.5, (A 4) is a more general condition than the "q-regularity" imposed by [1].

Until now, we have derived upper bounds on the deviation probability (1.3) under (A 4), which in the coherent case is equivalent to (A 5), namely that the auxiliary function $\overline{h}_\rho(t)$ must approach zero at a rate of $t^{1/q}$ for some $q \geq 1$. For coherent risk measures that do not satisfy (A 5), we can still derive an upper bound on the deviation probability by modifying the estimation technique in the proof of Theorem 3.1. This is stated in the following Theorem.

Theorem 3.8 *Let ρ be a normalized law-invariant coherent risk measure that satisfies the Lebesgue property, and assumptions (A 1), (A 2), (A 3) be fulfilled with $\xi \equiv B \in \mathbb{R}$ and finite $J(\mathbb{F}^\Theta, B, 1/8)$. Then $K_\varepsilon := \inf\left\{K \in \mathbb{N} \mid 4B\ \overline{h}_\rho(2^{-K+1}) < \varepsilon\right\}$ is well-defined for any $\epsilon > 0$, and*

$$\begin{aligned}\mathbb{P}\left(\left\{\left|\inf_{\theta\in\Theta} \mathcal{R}_\rho(\widehat{F}_{n,\theta}) - \inf_{\theta\in\Theta} \mathcal{R}_\rho(F_\theta)\right| \geq \varepsilon\right\}\right) &\leq \mathbb{P}^*\left(\sup_{\theta\in\Theta}\left|\mathcal{R}_\rho(\widehat{F}_{n,\theta}) - \mathcal{R}_\rho(F_\theta)\right| \geq \epsilon\right) \\ &\leq \frac{1}{2304\ t^2\ \ln(4)}\ \exp\left(-\frac{n\ t^2\ \varepsilon^2}{8\ 4^{K_\varepsilon}\ (t+1)^2\ B^2}\right),\end{aligned}$$

holds for $t > 0$ and

$$\varepsilon > 2^{K_\varepsilon - 1}\ 128\ (t+1)\ \sqrt{2}\ B\big[J\big(\mathbb{F}^\Theta, B, 1/8\big) + 3/2\big]/\sqrt{n}$$

Here $\mathbb{P}^*$ denotes the outer probability w.r.t. $\mathbb{P}$.

The proof may be found in Subsection 7.2. We note that Theorem 3.8 holds under very general conditions, namely that K_ϵ always exists due to the Lebesgue property. As a tradeoff, the bound can be conservative for small ϵ, since it depends exponentially on K_ϵ. An application of Theorem 3.8 will be given in Section 4, when we estimate the deviation probability for inverse S-shaped distortion risk measures.

4. Distortion Risk Measures with Inverse S-Shaped Distortion Function

Let $h : [0,1] \to [0,1]$ be a non-decreasing distortion function such that $h(0) = 0$, $h(1) = 1$. We say that h is inverse S-shaped, if there exists a $p_0 \in (0,1)$ such that h is concave on $[0, p_0]$, but convex on $[p_0, 1]$.

For any distortion function h, we also define the dual function $\bar{h}(p) := 1 - h(1-p)$. For an inverse S-shaped distortion function, we also define $h_0(p) = \min\{h(p), h(p_0)\}$ and $\bar{h}_0(p) := \min\{\bar{h}(p), 1-h(p_0)\}$, for $p \in [0,1]$. Note that both h_0 and $\bar{h}_0$ are non-decreasing concave functions on $[0,1]$, and we denote their normalized versions respectively as $h_{cc} := h_0/h(p_0)$ and $h_{cv} := \bar{h}_0/(1-h(p_0))$.

Lemma 4.1 *Let h be an inverse S-shaped distortion function. Then, we have that*

$$\rho_h(X) = h(p_0)\rho_{h_{cc}}(X) - (1-h(p_0))\rho_{h_{cv}}(-X).$$

Therefore, we see that upper bounds on the deviation probability for distortion risk measures with inverse S-shaped distortion functions can be provided by bounding their concave and convex parts separately. We state this in the following theorem, where $\overline{N}_{q,cc}, \overline{N}_{q,cv}$ denote the constants in (A 5) for h_{cc} and h_{cv} (recall that by Example 3.3, $\overline{h}_{\rho_h}(t) = h(t)$).

Theorem 4.2 *Let ρ_h be a distortion risk measure with an inverse S-shaped distortion function. Let (A 1), (A 2) with $\xi \equiv B \in \mathbb{R}$, (A 3), ((A 5)) be fulfilled with $q_1, q_2 \in [1,\infty)$ such that $\overline{N}_{q_1,cc}, \overline{N}_{q_2,cv} < \infty$. Suppose that $J(\mathbb{F}^\Theta, B, 1/8)$ is finite. Then,*

$$\begin{aligned}
&\mathbb{P}^*\Big(\Big\{\sup_{\theta\in\Theta}\big|\mathcal{R}_{\rho_h}\big(\widehat{F}_{n,\theta}\big) - \mathcal{R}_{\rho_h}\big(F_\theta\big)\big| > \epsilon\Big\}\Big)\\
&\le \exp\left(-\frac{n\ t^2\ \varepsilon^{2q_1}}{2^{6q_1-1}\ (t+1)^2\ B^{2q_1}\ h(p_0)^{2q_1}\overline{N}^{2q_1}_{q_1,cc}}\right)\\
&\qquad + \exp\left(-\frac{n\ t^2\ \varepsilon^{2q_2}}{2^{6q_2-1}\ (t+1)^2\ B^{2q_2}\ (1-h(p_0))^{2q_2}\overline{N}^{2q_2}_{q_2,cv}}\right),
\end{aligned}$$

holds for $t > 0$ if

$$\begin{aligned}
\varepsilon > &\max_{(q,c)\in\{(q_1,cc),(q_2,cv)\}} \overline{N}_{q,c}\cdot B\ [2^{2q-1}\ 128\ (t+1)\ \sqrt{2}]^{1/q}\ \big[J\big(\mathbb{F}^\Theta, B, 1/8\big) + 3/2\big]^{1/q}/n^{1/(2q)}\\
&\cdot \max\{h(p_0), 1-h(p_0)\}.
\end{aligned}$$

Here, $\mathbb{P}^*$ denotes the outer probability w.r.t. $\mathbb{P}$.

Example 4.3 *Let h_1, h_2 be two distortion functions that satisfy (A 5) with $q_1, q_2 \geq 1$ (such as the examples in Table 1). Let $p_0 \in (0,1)$. Then, the function*

$$h^S(p) = \begin{cases} h_1(p) & p \leq p_0 \\ 1 - \frac{1-h_1(p_0)}{h_2(1-p_0)} h_2(1-p) & p \geq p_0, \end{cases}$$

is an inverse S-shaped distortion function, such that the corresponding h^S_{cc}, h^S_{cv} satisfies $\overline{N}_{q_1,cc}, \overline{N}_{q_2,cv} < \infty$.

Example 4.4 *Consider the Kahneman-Tversky function (see [24]), which is defined as*

$$h(p) = 1 - \frac{(1-p)^\beta}{\left((1-p)^\beta + p^\beta\right)^{1/\beta}}, \;\; 0 < \beta < 1.$$

One can show that the Kahneman-Tversky function satisfies $\overline{N}_{q,cc}, \overline{N}_{q,cv} < \infty$, for $q = 1/\beta$. We provide the details in Subsection 8.2.

However, there are inverse S-shaped distortion functions for which their convex part h_{cv} does not satisfy $\overline{N}_{q,cv} < \infty$ for any $q \geq 1$. One prominent example of this is Prelec's distortion function (see [19]), defined as

$$h(p) = 1 - \exp\left(-(-\log(1-p))^\alpha\right), \;\; 0 < \alpha < 1.$$

Indeed, for Prelec's distortion function, we have that for $p \leq 1/e$,

$$h_{cv}(p) = \exp\left(-(-\log(p))^\alpha\right) \cdot \frac{1}{1 - h(1/e)}, \;\; 0 < \alpha < 1.$$

However, since for any $q \geq 1$,

$$\lim_{p \downarrow 0} \frac{\exp\left(-(-\log(p))^\alpha\right)}{p^{1/q}} = \lim_{p \downarrow 0} \exp\left(\frac{1}{q} \log\left(\frac{1}{p}\right)\left(1 - \log^{\alpha-1}\left(\frac{1}{p}\right)\right)\right) = \infty,$$

we have that $\overline{N}_{q,cv} = \infty$ for any $q \geq 1$. Hence, Theorem 4.2 cannot be applied to Prelec's function. Instead, one should use Theorem 3.8 to estimate the deviation probability for the convex part h_{cv} of Prelec's function.

5. Unbounded Case

In this section, we provide an extension of Theorem 3.1 to goal functions G with unbounded support. This requires several technical adjustments to the proof of Theorem 3.1, namely

1. Instead of a uniform bound B as the envelope function, we assume that ξ is integrable up to some order $r \geq 2$ (i.e., $\|\xi\|_{\mathbb{P}^Z,r} < \infty$).

2. The deviation quantity $\mathcal{R}_\rho(\widehat{F}_{n,\theta}) - \mathcal{R}_\rho(F_\theta)$ is decomposed into a part with bounded goal functions for the application of Theorem 3.1, and a remaining part that requires further estimation.

3. A new upper bound b is established on the penalty function $\beta_\rho(\mu)$ in the definition of $N_{q,b}$ as in (A 4), which can no longer be taken as $4B + \delta$, as in Theorem 3.1.

Given an envelope function ξ with $\|\xi\|_{\mathbb{P}^Z,r} < \infty$, the decomposition step of the aforementioned technical adjustment is carried out by introducing an auxiliary function $\phi_{r,n} : \mathbb{R} \to \mathbb{R}$ defined by $\phi_{r,n}(x) := (x \wedge w_{r,n}) \vee (-w_{r,n})$, where for $n \in \mathbb{N}$, we set $w_{r,n} := (2n)^{1/r} \, \|\xi\|_{\mathbb{P}^Z,r}$. Note that the image of $\phi_{r,n}$ is always contained in the interval $[-w_{r,n}, w_{r,n}]$. The purpose of $\phi_{r,n}$ is solely to restrict the objective function G on a compact support, as we will consider the class of functions $\phi_{r,n}(G(\theta, Z))$ for $\theta \in \Theta$. We let $F_\theta^{\phi_{r,n}}$ to denote the distribution function of $\phi_{r,n}(G(\theta, Z))$, and $\widehat{F}_\theta^{\phi_{r,n}}$ the corresponding empirical distribution function based on $Z_1, \ldots, Z_n$.

To make the third technical adjustment, we impose the following condition on the Kusuoka representation, which by Remark 3.5, also provides a sufficient condition on (A 4).

(A 6) There exists $r, q \in [1, \infty)$ with $r \geq q \vee 2$, such that $\|\xi\|_{\mathbb{P}^Z,r} < \infty$ and

$$\overline{M}_{r,q}^{\xi} := \sup_{\mu \in \mathrm{dom}(\beta)} \left(\int_{[0,1)} \frac{(4 \, \|\xi\|_{\mathbb{P}^Z,r}) \vee 1}{(1-\alpha)^{1/q}} \, \mu(d\alpha) - \beta(\mu) \right) < \infty.$$

We note that in a similar way as in Remark 3.5, an upper bound on $\overline{M}_{r,q}^{\xi}$ can also be provided by the q-regularity condition introduced in [1], namely one has

$$\overline{M}_{r,q}^{\xi} \leq \frac{q-1}{q} \sup_{k \in \mathbb{N}} \rho(X^*_{[4\|\xi\|_{\mathbb{P}^Z,r}] \vee 1, q} \wedge k),$$

where $X^*_{\lambda,q}$ is the Pareto distribution with scale parameter $\lambda > 0$ and shape parameter q.

Finally, besides the auxiliary function $\phi_{r,n}$, we also introduce the auxiliary events,

$$B_{n,r}^{\xi} := \left\{ \frac{1}{n} \sum_{j=1}^{n} \xi(Z_j)^r \leq 2\mathbb{E}[\xi(Z_1)^r] \right\} \quad (n \in \mathbb{N}). \tag{5.1}$$

After introducing all the necessary notations, we can state the following theorem on the deviation probability (3), when G has unbounded support.

Theorem 5.1 *Let ρ be a normalized law-invariant convex risk measure which satisfies the Lebesgue property, and let assumptions (A 1), (A 2), (A 3) and (A 6) be fulfilled for* *some $q \in [1,\infty)$ and $r \geq q \vee 2$. Denote $b_{r,\delta} := 2\overline{M}^{\xi}_{r.q} + \delta$. If $J(\mathbb{F}^{\Theta}, \xi, 1/4) < \infty$, then for any $t, \delta > 0$,*

$$\begin{aligned}
&\mathbb{P}\left(\left|\inf_{\theta\in\Theta}\mathcal{R}_\rho(\widehat{F}_{n,\theta}) - \inf_{\theta\in\Theta}\mathcal{R}_\rho(F_\theta)\right| \geq \epsilon\right) \\
&\leq \mathbb{P}^*\left(\sup_{\theta\in\Theta}\left|\mathcal{R}_\rho(\widehat{F}_{n,\theta}) - \mathcal{R}_\rho(F_\theta)\right| \geq \epsilon\right) \\
&\leq \exp\left(-\frac{n^{1-2q/r}\ t^2\ \varepsilon^{2q}}{2^{6q-1}\ (t+1)^2\ \overline{N}^{2q}_{q,b_{r,\delta}}\ \|\xi\|^{2q}_{\mathbb{P}_Z,r}}\right) + \mathbb{P}\big(\Omega \setminus B^{\xi}_{n,r}\big)
\end{aligned}$$

whenever $t, \delta > 0$ and

$$\begin{aligned}
\varepsilon > &\ \frac{\big(64\sqrt{2}(t+1)\ [J(\mathbb{F}^{\Theta}, \xi, 1/8) + 3/2]\big)^{1/q} 8\overline{N}_{q,b_{r,\delta}} \|\xi\|^{1-1/q}_{\mathbb{P}^Z,r}}{n^{1/(2q)+(1-q)/(rq)}} \\
&+ \frac{4[3\overline{M}^{\xi}_{r,q} + \delta]\ \|\xi\|_{\mathbb{P}^Z,r} n^{(2-r)/(2rq)}}{n^{1/(2q)+(1-q)/(rq)}}.
\end{aligned}$$

The proof may be found in the Section 7.5. Let us discuss the further estimation of the probability $\mathbb{P}(\Omega \setminus B^{\xi}_{n,r})$, which is similar to the case of $r = 2$ as discussed in Remark 3.2 of [14], using the substitution $\tilde{\xi} = \xi^{r/2}$. Following that, we state an analogous of Remark 3.2 in [14] in the following remark.

Remark 5.2 *We have the following estimations on $\mathbb{P}(\Omega \setminus B^{\xi}_{n,r})$.*

(1) If $\xi(Z_1)$ is integrable of order $2r$, then Chebyshev's inequality implies

$$\mathbb{P}(\Omega \setminus B^{\xi}_{n,r}) \leq \frac{\mathrm{Var}[\xi^r(Z_1)]}{n\mathbb{E}[\xi^r(Z_1)]^2}.$$

(2) If $\mathbb{E}[\exp(\lambda\xi^r(Z_1))] < \infty$ for some $\lambda > 0$, then by Remark 3.2 of [14], we have that

$$\mathbb{P}(\Omega \setminus B^{\xi}_{n,r}) \leq \exp\left(-n\mathbb{E}[\xi^r(Z_1)]^2/(8\delta_0^2)\right) \vee \exp\left(-n\mathbb{E}[\xi^r(Z_1)]/(4\delta_0)\right),$$

for $n \in \mathbb{N}$ and

$$\delta_0 \geq \sup_{k\in\mathbb{N}, k\geq 2}\left(\left|\mathbb{E}\left[(\xi^r(Z_1) - \mathbb{E}[\xi^r(Z_1)])^k\right]\right|/k!\right)^{1/k}.$$

Finally, let us comment on the bound in Theorem 5.1. Note that if G is uniformly bounded in θ, then we may take $\xi \equiv B < \infty$, which means $r = \infty$ and $\mathbb{P}(\Omega \setminus B^{\xi}_{n,r}) = 0$. Then, the rate in Theorem 5.1 recovers back to $n^{-1/(2q)}$, which is the same as in Theorem 3.1 for the bounded case. Moreover, combining the estimation of $\mathbb{P}(\Omega \setminus B^{\xi}_{n,r})$ in Remark 5.2, the bound in Theorem 5.1 yields a smaller confidence region for the true optimal value $\inf_{\theta\in\Theta}\mathcal{R}_\rho(F_\theta)$ than the moment bound provided in [1], which through Markov's inequality decays in the order of $1/n^{1/(q(2-1/r))-1/(r-2)}$, which is always slower than $1/\sqrt{n}$.

6. Conclusion and Discussion

By utilizing the Kusuoka representation of law-invariant convex risk measures and tools from empirical process theory, we obtain non-asymptotic upper estimations on the deviation probability (1.3) under more general conditions imposed on the goal functions and the risk measures as compared to [1] and [14]. As a direction of future research, it would be valuable to further investigate the tightness of the rate $n^{-1/(2q)}$ in Theorem 3.1 for $q > 1$.

7. Proofs

Lemma 7.1 *Let assumptions (A 1) - (A 3) be fulfilled. If ρ satisfies the Lebesgue property, then the mapping*

$$\inf_{\theta\in\Theta} \mathcal{R}_\rho(F_\theta) - \inf_{\theta\in\Theta} \mathcal{R}_\rho(\widehat{F}_{n,\theta})$$

is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ for every $n \in \mathbb{N}$. If in addition G is lower semicontinuous in θ, then the mapping

$$\sup_{\theta\in\Theta} \left|\mathcal{R}_\rho(F_\theta) - \mathcal{R}_\rho(\widehat{F}_{n,\theta})\right|$$

is also a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$.

Proof First of all $\{\mathcal{R}_\rho(F_\theta) \mid \theta \in \Theta\}$ is bounded from below due to (A 2). Fix any $n \in \mathbb{N}$. Furthermore let ξ be from (A 2), and let $\overline{\Theta}$ be some at most countable subset of Θ as in assumption (A 3). The left-continuous quantile function of a distribution function F will be denoted by $F^{\leftarrow}$. We may select some random variable U on the atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$ which is uniformly distributed on $(0, 1)$.

By assumption (A 3), there is some $A_n \in \mathcal{F}$ with $\mathbb{P}(A_n) = 1$ such that

$$\inf_{\vartheta\in\overline{\Theta}} \max_{j\in\{1,\dots,n\}} \left\{\left|G\big(\theta, Z_j(\omega)\big) - G\big(\vartheta, Z_j(\omega)\big)\right|\right\} = 0 \quad \text{for } \omega \in A_n, \theta \in \Theta.$$

Let $\omega \in A_n$ and $\theta \in \Theta$ be arbitrary. We may find some sequence $(\vartheta_k)_{k\in\mathbb{N}}$ in $\overline{\Theta}$ such that $\widehat{F}_{n,\vartheta_k|\omega}(x) \to \widehat{F}_{n,\theta|\omega}(x)$ at every continuity point x of $\widehat{F}_{n,\theta|\omega}$. Then $\widehat{F}^{\leftarrow}_{n,\vartheta_k|\omega}(\alpha) \to \widehat{F}^{\leftarrow}_{n,\theta|\omega}(\alpha)$ at every continuity point α of $\widehat{F}^{\leftarrow}_{n,\theta|\omega}(\alpha)$ (see e.g. [25, Lemma 21.2]). In particular, $\widehat{F}^{\leftarrow}_{n,\vartheta_k|\omega}(U) \to \widehat{F}^{\leftarrow}_{n,\theta|\omega}(U)$ $\mathbb{P}$-a.s.. Since in addition the sequence $\big(\widehat{F}^{\leftarrow}_{n,\vartheta_k|\omega}(U)\big)_{k\in\mathbb{N}}$ is uniformly bounded by $\max_{j\in\{1,\dots,n\}} \xi\big(Z_j(\omega)\big)$, Lebesgue property of ρ implies

$$\rho\big(\widehat{F}^{\leftarrow}_{n,\vartheta_k|\omega}(U)\big) \to \rho\big(\widehat{F}^{\leftarrow}_{n,\theta|\omega}(U)\big). \tag{7.1}$$

Finally, note that $\widehat{F}^{\leftarrow}_{n,\vartheta|\omega}(U)$ is distributed according to the empirical distribution based on $G\big(\vartheta, Z_1(\omega)\big), \dots, G\big(\vartheta, Z_n(\omega)\big)$ for $\vartheta \in \Theta$ and ω. Thus, we have shown

$$\inf_{\vartheta\in\overline{\Theta}} \mathcal{R}_\rho(\widehat{F}_{n,\vartheta|\omega}) = \inf_{\theta\in\Theta} \mathcal{R}_\rho(\widehat{F}_{n,\theta|\omega}) \quad \text{for } \omega \in A_n.$$

Since $\overline{\Theta}$ is at most countable, and $(\Omega, \mathcal{F}, \mathbb{P})$ is complete, it remains to show that $\mathcal{R}_\rho(\widehat{F}_{n,\theta})$ is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ for every $\theta \in \Theta$.

For this purpose, let $F^{\leftarrow}_{(x_1,\ldots,x_n)}$ denote the left-continuous quantile function of the empirical distribution function based on $(x_1, \ldots x_n) \in \mathbb{R}^n$. In the same way as we derived convergence (7.1), we may use again the Lebesgue property to verify the continuity of the mapping

$$\varphi : \mathbb{R}^n \to \mathbb{R},\ (x_1, \ldots, x_n) \mapsto \rho\big(F^{\leftarrow}_{(x_1,\ldots,x_n)}(U)\big).$$

Thus $\mathcal{R}_\rho(\widehat{F}_{n,\theta}) = \varphi \circ \big(G(\theta, Z_1), \ldots, G(\theta, Z_n)\big)$ is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ for every $\theta \in \Theta$.

For the remaining part of the proof, let us assume that $G(\cdot, z)$ is lower semicontinuous for every $z \in \mathbb{R}^d$. Note that $G(\cdot, z)$ is bounded due to (A 2), and by (A 3) we may find some $A \in \mathcal{F}$ with $\mathbb{P}(A) = 1$ such that

$$\inf_{\theta \in \Theta} \big(G\big(\theta, Z_1(\omega)\big) - \psi(\theta)\big) = \inf_{\theta \in \overline{\Theta}} \big(G\big(\theta, Z_1(\omega)\big) - \psi(\theta)\big) \quad \text{for } \omega \in A.$$

for any continuous mapping $\psi : \Theta \to \mathbb{R}$. This implies that

$$\inf_{\theta \in \Theta} \big(G\big(\theta, Z_1\big) - \psi(\theta)\big)$$

is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ for any continuous mapping $\psi : \Theta \to \mathbb{R}$, because $\overline{\Theta}$ is countable and $(\Omega, \mathcal{F}, \mathbb{P})$ is complete. Since Θ is compactly metrizable, we may draw on Theorem 2.1 and Remark 1.1 both from [17] to conclude that G is measurable w.r.t. the product σ-algebra $\mathcal{B}(\Theta) \otimes \mathcal{B}(\mathbb{R}^d)$ of the Borel σ-algebra $\mathcal{B}(\Theta)$ on Θ and the Borel σ-algebra $\mathcal{B}(\mathbb{R}^d)$ on $\mathbb{R}^d$. As a further consequence, $\varphi \circ \big(G(\cdot, Z_1), \ldots, G(\cdot, Z_n)\big)$ is also $\mathcal{B}(\Theta) \otimes \mathcal{B}(\mathbb{R}^d)$ measurable. Moreover, in view of Corollary A.3 in Appendix A along with assumption (A 2), the mapping

$$\Theta \to \mathbb{R},\ \theta \mapsto \rho\big(G(\theta, Z_1)\big) \tag{7.2}$$

is lower semicontinuous. Therefore, the mapping

$$\Theta \times \Omega \to \mathbb{R},\ (\theta, \omega) \mapsto \varphi \circ \Big(G\big(\theta, Z_1(\omega)\big), \ldots, G\big(\theta, Z_n(\omega)\big)\Big) - \rho\big(G(\theta, Z_1)\big),$$

is $\mathcal{B}(\Theta) \otimes \mathcal{B}(\mathbb{R}^d)$ measurable, and it is also bounded in θ for every $\omega \in \Omega$ due to (A 2) along with the monotonicity of ρ. Hence, completeness of $(\Omega, \mathcal{F}, \mathbb{P})$ implies that

$$\sup_{\theta \in \Theta} \Big|\mathcal{R}_\rho(\widehat{F}_{n,\theta}) - \mathcal{R}_\rho(F_\theta)\Big| = \sup_{\theta \in \Theta} \big|\varphi \circ \big(G(\theta, Z_1), \ldots, G(\theta, Z_n)\big) - \rho\big(G(\theta, Z_1)\big)\big|$$

is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ (see [26, Example 1.7.5]). This completes the proof. □

In the following proofs, we shall make use of the usual notation from empirical process theory

$$(\mathbb{P}_n - \mathbb{P})(f) := \frac{1}{n} \sum_{j=1}^{n} \Big(f(Z_j) - \mathbb{E}[f(Z_j)]\Big) \tag{7.3}$$

for $\mathbb{P}^Z$-integrable mappings $f : \mathbb{R}^d \to \mathbb{R}$.

7.1. Proof of Theorem 3.1

First of all, the mapping

$$\inf_{\theta\in\Theta}\mathcal{R}_\rho(\widehat{F}_{n,\theta})-\inf_{\theta\in\Theta}\mathcal{R}_\rho(F_\theta)$$

is a random variable on $(\Omega,\mathcal{F},\mathbb{P})$ due to Lemma 7.1, and if G is lower semicontinuous in θ, then Lemma 7.1 tells us that

$$\sup_{\theta\in\Theta}\left|\mathcal{R}_\rho(\widehat{F}_{n,\theta})-\mathcal{R}_\rho(F_\theta)\right|,$$

is also a random variable on $(\Omega,\mathcal{F},\mathbb{P})$. Note that the following inequality is valid for any $\epsilon>0$:

$$\mathbb{P}\left(\left\{\left|\inf_{\theta\in\Theta}\mathcal{R}_\rho(\widehat{F}_{n,\theta})-\inf_{\theta\in\Theta}\mathcal{R}_\rho(F_\theta)\right|\geq\varepsilon\right\}\right)\leq\mathbb{P}^*\left(\sup_{\theta\in\Theta}\left|\mathcal{R}_\rho(\widehat{F}_{n,\theta})-\mathcal{R}_\rho(F_\theta)\right|\geq\epsilon\right).$$

Recall that G is uniformly bounded by B. Hence, for every, $\theta\in\Theta$,

$$\begin{aligned}\mathcal{R}_\alpha(F_\theta)&=\inf_{|x|\leq B}\mathbb{E}\big[\big(G(\theta,Z_1)-x\big)^++(1-\alpha)x\big]/(1-\alpha),\\ \mathcal{R}_\alpha(\widehat{F}_{n,\theta})&=\inf_{|x|\leq B}\frac{1}{n}\sum_{j=1}^n\Big[\Big(G\big(\theta,Z_j\big)-x\Big)^++(1-\alpha)x\Big]/(1-\alpha).\end{aligned}$$

Define the function class $\mathbb{F}_B^\Theta:=\{[G(\theta,\cdot)-x]^+\mid\theta\in\Theta,x\in[-B,B]\}$, and furthermore $\Delta_n:=\sup_{f\in\mathbb{F}_B^\Theta}|(\mathbb{P}_n-\mathbb{P})(f)|$. We obtain

$$\begin{aligned}\left|\mathcal{R}_\alpha(F_\theta)-\mathcal{R}_\alpha(\widehat{F}_{n,\theta})\right|&\leq\frac{1}{1-\alpha}\sup_{|x|\leq B}\left|(\mathbb{P}_n-\mathbb{P})\big(\big[G(\theta,\cdot)-x\big]^+\big)\right|\\ &\leq\frac{1}{1-\alpha}\sup_{f\in\mathbb{F}_B^\Theta}|(\mathbb{P}_n-\mathbb{P})(f)|=\frac{\Delta_n}{1-\alpha}\quad\text{for }\alpha\in[0,1),\theta\in\Theta.\end{aligned}$$

Let $\mathcal{M}_{B,\delta}:=\{\mu\in\mathcal{M}\mid\beta(\mu)\leq(4B+\delta)\}$, for $\delta>0$. We obtain from Proposition A.2 in Appendix A, that

$$\begin{aligned}\mathcal{R}_\alpha(F_\theta)&=\sup_{\mu\in\mathcal{M}_{B,\delta}}\left(\int_{[0,1)}\mathcal{R}_\alpha(F_\theta)\ \mu(d\alpha)-\beta(\mu)\right),\\ \mathcal{R}_\alpha(\widehat{F}_{n,\theta})&=\sup_{\mu\in\mathcal{M}_{B,\delta}}\left(\int_{[0,1)}\mathcal{R}_\alpha(\widehat{F}_{n,\theta})\ \mu(d\alpha)-\beta(\mu)\right).\end{aligned}$$

Since the class $\mathbb{F}_B^\Theta$ is uniformly bounded by the constant $2B$, we have for any θ and every $\delta>0$:

$$\begin{aligned}\left|\mathcal{R}_\rho(\widehat{F}_{n,\theta})-\mathcal{R}_\rho(F_\theta)\right|&\leq\sup_{\mu\in\mathcal{M}_{B,\delta}}\int_{[0,1)}\left|\mathcal{R}_\alpha(F_\theta)-\mathcal{R}_\alpha(\widehat{F}_{n,\theta})\right|\mu(d\alpha)\\ &\leq\sup_{\mu\in\mathcal{M}_{B,\delta}}\int_{[0,1)}\left(\frac{\Delta_n}{1-\alpha}\wedge 2B\right)\mu(d\alpha).\end{aligned}\tag{7.4}$$

Next, note that $\Delta_n \le 2B$ holds pointwise so that in view of (7.4)

$$\left\{\sup_{\theta\in\Theta}\left|\mathcal{R}_\rho(\widehat{F}_{n,\theta}) - \mathcal{R}_\rho(F_\theta)\right| > 0\right\} \subseteq \{0 < \Delta_n \le 2B\} = \bigcup_{K=0}^{\infty} \Omega_{nK}, \tag{7.5}$$

where $\Omega_{nK} := \big\{2^{K-1}\Delta_n < 2B \le 2^K\Delta_n\big\}$. For $K \in \mathbb{N}$ and $\omega \in \Omega_{nK}$ we obtain

$$\frac{\Delta_n}{1-\alpha} \wedge 2B \begin{cases} = \dfrac{\Delta_n}{1-\alpha} & \text{for } \alpha < 1-2^{1-K} \\ \le 2B & \text{for } \alpha \ge 1-2^{1-K}. \end{cases}$$

This implies by (7.4),

$$\begin{aligned}
&\left|\mathcal{R}_\rho(\widehat{F}_{n,\theta|_\omega}) - \mathcal{R}_\rho(F_\theta)\right| \\
&\le \sup_{\mu\in\mathcal{M}_{B,\delta}}\left(\Delta_n(\omega)\int_{[0,1-2^{1-K})}\frac{1}{1-\alpha}\,\mu(d\alpha) + 2B\,\mu\big([1-2^{1-K},1)\big)\right) \quad \text{for } \theta\in\Theta, \delta>0.
\end{aligned}$$

Furthermore, for any Borel probability measure μ on $[0,1)$ and every $t\in[0,1]$,

$$\int_{[0,1)}\frac{t\wedge(1-\alpha)}{1-\alpha}\,\mu(d\alpha) = \mu\big([1-t,1)\big) + t\int_{[0,1-t)}\frac{1}{1-\alpha}\,\mu(d\alpha).$$

Hence, recalling $\Delta_n(\omega) < 2^{2-K}\,B$ and $2B \le 2^K\Delta_n(\omega)$,

$$\begin{aligned}
&\left|\mathcal{R}_\rho(\widehat{F}_{n,\theta|_\omega}) - \mathcal{R}_\rho(F_\theta)\right| \\
&\le \sup_{\mu\in\mathcal{M}_{B,\delta}}\left(\int_{[0,1-2^{1-K})}\frac{\Delta_n(\omega)-2^{2-K}B}{1-\alpha}\,\mu(d\alpha) + 2B\int_{[0,1)}\frac{2^{1-K}\wedge(1-\alpha)}{1-\alpha}\,\mu(d\alpha)\right) \\
&\le 2B\sup_{\mu\in\mathcal{M}_{B,\delta}}\int_{[0,1)}\frac{2^{1-K}\wedge(1-\alpha)}{1-\alpha}\,\mu(d\alpha) \\
&\le 2^K\Delta_n(\omega)\sup_{\mu\in\mathcal{M}_{B,\delta}}\int_{[0,1)}\frac{2^{1-K}\wedge(1-\alpha)}{1-\alpha}\,\mu(d\alpha) \quad \text{for } \theta\in\Theta, \delta>0.
\end{aligned} \tag{7.6}$$

Moreover, for $\omega\in\Omega$ with $2^{-1}\Delta_n(\omega) < 2B \le \Delta_n(\omega)$, we have $2B \le \Delta_n(\omega)/(1-\alpha)$ for every $\alpha\in[0,1)$. Therefore,

$$\left|\mathcal{R}_\rho(\widehat{F}_{n,\theta|_\omega}) - \mathcal{R}_\rho(F_\theta)\right| \le 2B \le \Delta_n(\omega). \tag{7.7}$$

Now, let $\varepsilon > 0$. Putting together (7.5) with (7.6) and (7.7), we may observe

$$\begin{aligned}
&\sup_{\theta\in\Theta}\left|\mathcal{R}_\rho(\widehat{F}_{n,\theta}) - \mathcal{R}_\rho(F_\theta)\right| \\
&\le \mathbb{1}_{\Omega_{n0}}\cdot\Delta_n + \sum_{K=1}^{\infty}\mathbb{1}_{\Omega_{nK}}\cdot 2^K\Delta_n\sup_{\mu\in\mathcal{M}_{B,\delta}}\int_{[0,1)}\frac{2^{1-K}\wedge(1-\alpha)}{1-\alpha}\,\mu(d\alpha)
\end{aligned} \tag{7.8}$$

for every $\delta > 0$. Invoking assumption (A 4) with $q \in [1, \infty)$ and the constants $N_{q,b}$ we have

$$2^K \sup_{\mu \in \mathcal{M}_{B,\delta}} \int_{[0,1)} \frac{2^{1-K} \wedge (1-\alpha)}{1-\alpha}\ \mu(d\alpha) \leq 2^K\ N_{q,(4B+\delta)}\ 2^{(1-K)/q}, \quad \text{for } \delta > 0, K \in \mathbb{N}.$$

This implies

$$\begin{aligned}
&\mathbb{1}_{\Omega_{nK}} \cdot 2^K \Delta_n \sup_{\mu \in \mathcal{M}_{B,\delta}} \int_{[0,1)} \frac{2^{1-K} \wedge (1-\alpha)}{1-\alpha}\ \mu(d\alpha) \\
&\leq \mathbb{1}_{\Omega_{nK}} \cdot 2 \cdot 2^{(K-1)(q-1)/q}\ N_{q,(4B+\delta)}\ \Delta_n \\
&= \mathbb{1}_{\Omega_{nK}} \cdot N_{q,(4B+\delta)}\ 2 \cdot 2^{(K-1)(q-1)/q}\ \Delta_n^{(q-1)/q}\ \Delta_n^{1/q} \\
&\leq \mathbb{1}_{\Omega_{nK}} \cdot N_{q,(4B+\delta)}\ 2 \cdot 2^{(K-1)(q-1)/q}\ 2^{(2-K)(q-1)/q}\ B^{(q-1)/q}\ \Delta_n^{1/q} \\
&\leq \mathbb{1}_{\Omega_{nK}} \cdot N_{q,(4B+\delta)} \cdot 2^{(2q-1)/q}\ B^{(q-1)/q}\ \Delta_n^{1/q} \quad \text{for } \delta > 0, K \in \mathbb{N}.
\end{aligned}$$

Since $N_{q,(4B+\delta)} \geq 1$ for $\delta > 0$, and since $\Delta_n \leq 2B$ we may also observe

$$\begin{aligned}
\mathbb{1}_{\Omega_{n0}} \cdot \Delta_n = \mathbb{1}_{\Omega_{n0}} \cdot \Delta_n^{(q-1)/q}\ \Delta_n^{1/q} &= \mathbb{1}_{\Omega_{nK}} \cdot 2^{(q-1)/q}\ B^{(q-1)/q}\ \Delta_n^{1/q} \\
&\leq \mathbb{1}_{\Omega_{n0}} \cdot N_{q,(4B+\delta)} \cdot 2^{(2q-1)/q}\ B^{(q-1)/q}\ \Delta_n^{1/q} \quad \text{for } \delta > 0.
\end{aligned}$$

Hence, combining (7.8) with (7.5), we end up with

$$\begin{aligned}
&\mathbb{P}^*\left(\left\{\sup_{\theta \in \Theta} \left|\mathcal{R}_\rho(\widehat{F}_{n,\theta}) - \mathcal{R}_\rho(F_\theta)\right| \geq \varepsilon\right\}\right) \\
&\leq \mathbb{P}^*\left(\left\{\sum_{K=0}^{\infty} \mathbb{1}_{\Omega_{nK}} \cdot N_{q,(4B+\delta)} \cdot 2^{(2q-1)/q}\ B^{(q-1)/q}\ \Delta_n^{1/q} \geq \varepsilon\right\}\right) \\
&\leq \mathbb{P}^*\left(\left\{\ \Delta_n \geq \frac{\varepsilon^q}{N_{q,(4B+\delta)}^q \cdot 2^{(2q-1)}\ B^{(q-1)}}\right\}\right) \quad \text{for } \delta > 0. \qquad (7.9)
\end{aligned}$$

Recall the at most countable subset $\overline{\Theta} \subseteq \Theta$ from assumption (A 3). In view of Lemma 7.2 we may further conclude from (7.9)

$$\begin{aligned}
&\mathbb{P}^*\left(\left\{\sup_{\theta \in \Theta} \left|\mathcal{R}_\rho(\widehat{F}_{n,\theta}) - \mathcal{R}_\rho(F_\theta)\right| \geq \varepsilon\right\}\right) \\
&\leq \mathbb{P}\left(\left\{\sup_{f \in \widehat{\mathbb{F}}_I} \left|(\mathbb{P}_n - \mathbb{P})(f)\right| \geq \frac{\varepsilon^q}{N_{q,(4B+\delta)}^q \cdot 2^{(2q-1)}\ B^{(q-1)}}\right\}\right), \qquad (7.10)
\end{aligned}$$

where $\widehat{F}_I := \{[G(\theta, \cdot) - x]^+ \mid \theta \in \overline{\Theta}, x \in I\}$ for some countable dense subset $I \subseteq [-B, B]$. We want to apply the concentration inequality in Theorem B.1 from Appendix B to upper estimate the deviation probabilities on the right-hand of the inequality (7.10).

For this purpose, note that the constant function $2B$ is a positive envelope of $\widehat{\mathbb{F}}_I$, and recall from Lemma 7.2

$$J\big(\widehat{\mathbb{F}}_I, 2B, 1/2\big) \leq 4\ J(\mathbb{F}^\Theta, B, 1/8\big) + 6.$$

Since in addition, the function class $\widehat{\mathbb{F}}_I$ is at most countable, we may conclude from Theorem 2.1 in [14]

$$\mathbb{E}\left[\sup_{f\in\widehat{\mathbb{F}}_I}\big|(\mathbb{P}_n-\mathbb{P})(f)\big|\right] \leq \frac{32\ \sqrt{2}\ B}{\sqrt{n}}\ J\big(\widehat{\mathbb{F}}_I, 2B, 1/2\big) \leq \frac{128\ \sqrt{2}\ B\big[J\big(\mathbb{F}^\Theta, B, 1/8\big)+3/2\big]}{\sqrt{n}}.$$

If $t > 0$ and $\varepsilon > N_{q,(4B+\delta)}\cdot B\ [2^{2q-1}\ 128\ (t+1)\ \sqrt{2}]^{1/q}\ \big[J\big(\mathbb{F}^\Theta, B, 1/8\big)+3/2\big]^{1/q}/n^{1/(2q)}$, then

$$\begin{aligned}
&\frac{\varepsilon^q}{N^q_{q,(4B+\delta)}\cdot 2^{(2q-1)}\ B^{(q-1)}}\\
&= \frac{t\ \varepsilon^q}{(t+1)\ N^q_{q,(4B+\delta)}\cdot 2^{(2q-1)}\ B^{(q-1)}} + \frac{\varepsilon^q}{(t+1)\ N^q_{q,(4B+\delta)}\cdot 2^{(2q-1)}\ B^{(q-1)}}\\
&> \frac{t\ \varepsilon^q}{(t+1)\ N^q_{q,(4B+\delta)}\cdot 2^{(2q-1)}\ B^{(q-1)}} + \mathbb{E}\left[\sup_{f\in\widehat{\mathbb{F}}_I}\big|(\mathbb{P}_n-\mathbb{P})(f)\big|\right] \quad \text{for } \delta > 0.
\end{aligned}$$

Therefore in this case, by Theorem B.1 from Appendix B

$$\begin{aligned}
&\mathbb{P}\left(\left\{\sup_{f\in\widehat{\mathbb{F}}_I}\big|(\mathbb{P}_n-\mathbb{P})(f)\big| \geq \frac{\varepsilon^q}{N^q_{q,(4B+\delta)}\cdot 2^{(2q-1)}\ B^{(q-1)}}\right\}\right)\\
&\leq \exp\left(-\frac{n\ t^2\ \varepsilon^{2q}}{8\ (t+1)^2\ B^2\ N^{2q}_{q,(4B+\delta)}\cdot 4^{(2q-1)}\ B^{2(q-1)}}\right)\\
&\leq \exp\left(-\frac{n\ t^2\ \varepsilon^{2q}}{2^{4q+1}\ (t+1)^2\ B^{2q}\ N^{2q}_{q,(4B+\delta)}}\right) \quad \text{for } \delta > 0.
\end{aligned}$$

This completes the proof due to (7.10). □

7.2. Proof of Theorem 3.8

We re-examine the proof of Theorem 3.1. Let $\{\Omega_{nK}\}_{K\geq 0}$ be the disjoint partitions introduced in the proof of Theorem 3.1, where $\Omega_{nK} := \big\{2^{K-1}\Delta_n < 2B \leq 2^K\Delta_n\big\}$. Using the fact that ρ is coherent (hence $\beta_\rho(\mu) = 0$), we have:

$$\begin{aligned}
&\sup_{\theta\in\Theta}\Big|\mathcal{R}_\rho(\widehat{F}_{n,\theta}) - \mathcal{R}_\rho(F_\theta)\Big|\\
&\leq \mathbb{1}_{\Omega_{n0}}\cdot\Delta_n + \sum_{K=1}^\infty \mathbb{1}_{\Omega_{nK}}\cdot 2^K\Delta_n \sup_{\mu\in\mathcal{M},\beta_\rho(\mu)=0}\int_{[0,1)}\frac{2^{1-K}\wedge(1-\alpha)}{1-\alpha}\ \mu(d\alpha)\\
&= \mathbb{1}_{\Omega_{n0}}\cdot\Delta_n + \sum_{K=1}^\infty \mathbb{1}_{\Omega_{nK}}\cdot 2^K\Delta_n\ \overline{h}_\rho(2^{1-K}).
\end{aligned}$$

For $K \geq 1$, and $\omega \in \Omega_{nK}$, we have that

$$2^K \Delta_n \ \overline{h}_\rho(2^{1-K}) \leq 4B\overline{h}_\rho(2^{1-K}).$$

For a given $\epsilon > 0$, let $K_\epsilon := \min\{K \in \mathbb{N} \mid 4B\overline{h}_\rho(2^{1-K}) < \epsilon\}$. Then, for $K \geq K_\epsilon$, we have that

$$\sup_{\theta\in\Theta} \left| \mathcal{R}_\rho(\widehat{F}_{n,\theta}) - \mathcal{R}_\rho(F_\theta) \right| \leq 4B\overline{h}_\rho(2^{1-K}) < \epsilon.$$

Hence,

$$\left\{ \sup_{\theta\in\Theta} \left| \mathcal{R}_\rho(\widehat{F}_{n,\theta}) - \mathcal{R}_\rho(F_\theta) \right| > \epsilon \right\} \cap \bigcup_{K \geq K_\epsilon} \Omega_{nK} = \emptyset.$$

Therefore, using $\overline{h}_\rho(t) \leq 1, \forall t \in [0,1]$, we have

$$\begin{aligned}
\mathbb{P}^* \left(\sup_{\theta\in\Theta} \left| \mathcal{R}_\rho(\widehat{F}_{n,\theta}) - \mathcal{R}_\rho(F_\theta) \right| > \epsilon \right) &\leq \mathbb{P}^* \left(\mathbb{1}_{\Omega_{n0}} \cdot \Delta_n + \sum_{K=1}^{K_\epsilon - 1} \mathbb{1}_{\Omega_{nK}} \cdot 2^K \Delta_n \ \overline{h}_\rho(2^{1-K}) > \epsilon \right) \\
&\leq \mathbb{P}^* \left(\mathbb{1}_{\Omega_{n0}} \cdot \Delta_n + \sum_{K=1}^{K_\epsilon - 1} \mathbb{1}_{\Omega_{nK}} \cdot 2^K \Delta_n > \epsilon \right) \\
&\leq \sum_{K=0}^{K_\epsilon - 1} \mathbb{P}^* \left(\Delta_n > \frac{\epsilon}{2^K} \right).
\end{aligned}$$

Now, following the same arguments as in the proof of Theorem 3.1, we may use Lemma 7.2 to obtain the further estimation:

$$\begin{aligned}
&\mathbb{P}^* \left(\left\{ \sup_{\theta\in\Theta} \left| \mathcal{R}_\rho(\widehat{F}_{n,\theta}) - \mathcal{R}_\rho(F_\theta) \right| \geq \varepsilon \right\} \right) \\
&\leq \sum_{K=0}^{K_\varepsilon - 1} \mathbb{P} \left(\left\{ \sup_{f \in \widehat{\mathbb{F}}_I} \left| (\mathbb{P}_n - \mathbb{P})(f) \right| \geq \frac{\varepsilon}{2^K} \right\} \right),
\end{aligned} \tag{7.11}$$

where $\widehat{F}_I := \{[G(\theta,\cdot) - x]^+ \mid \theta \in \overline{\Theta}, x \in I\}$ for some countable dense subset $I \subseteq [-B, B]$. To apply the concentration inequality in Theorem B.1, we note that the constant function $2B$ is a positive envelope of $\widehat{\mathbb{F}}_I$, and recall from Lemma 7.2

$$J\big(\widehat{\mathbb{F}}_I, 2B, 1/2\big) \leq 4 \ J\big(\mathbb{F}^\Theta, B, 1/8\big) + 6.$$

Since in addition the function class $\widehat{\mathbb{F}}_I$ is at most countable, we may conclude from Theorem 2.1 in [14]

$$\mathbb{E} \left[\sup_{f \in \widehat{\mathbb{F}}_I} \left| (\mathbb{P}_n - \mathbb{P})(f) \right| \right] \leq \frac{32 \sqrt{2} \ B}{\sqrt{n}} \ J\big(\widehat{\mathbb{F}}_I, 2B, 1/2\big) \leq \frac{128 \sqrt{2} \ B \big[J\big(\mathbb{F}^\Theta, B, 1/8\big) + 3/2 \big]}{\sqrt{n}}.$$

If $t > 0$ and $\varepsilon > 2^{K_\varepsilon - 1}\ 128\ (t+1)\ \sqrt{2}\ B\big[J\big(\mathbb{F}^\Theta, B, 1/8\big) + 3/2\big]/\sqrt{n}$, then for arbitrary $K \in \{0, \ldots, K_\varepsilon - 1\}$,

$$\frac{\varepsilon}{2^K} = \frac{t\ \varepsilon}{(t+1)\ 2^K} + \frac{\varepsilon}{(t+1)\ 2^K} > \frac{t\ \varepsilon}{(t+1)\ 2^K} + \mathbb{E}\left[\sup_{f \in \widehat{\mathbb{F}}_I} \big|(\mathbb{P}_n - \mathbb{P})(f)\big|\right].$$

Therefore in this case, by Theorem B.1 from Appendix B

$$\mathbb{P}\left(\left\{\sup_{f \in \widehat{\mathbb{F}}_I} \big|(\mathbb{P}_n - \mathbb{P})(f)\big| \geq \frac{\varepsilon}{2^K}\right\}\right) \leq \exp\left(-\frac{n\ t^2\ \varepsilon^2}{2\ 4^{K+1}\ (t+1)^2\ B^2}\right),$$

for $K \in \{0, \ldots, K_\varepsilon - 1\}$. Therefore,

$$\begin{aligned}
\mathbb{P}^*\left(\left\{\sup_{\theta \in \Theta} \Big|\mathcal{R}_\rho(\widehat{F}_{n,\theta}) - \mathcal{R}_\rho(F_\theta)\Big| \geq \varepsilon\right\}\right) &\leq \sum_{k=0}^{K_\varepsilon - 1} \exp\left(-\frac{n\ t^2\ \varepsilon^2}{2\ 4^{k+1}\ (t+1)^2\ B^2}\right) \\
&\leq \sum_{k=0}^{K_\varepsilon - 1} \int_k^{k+1} \exp\left(-\frac{n\ t^2\ \varepsilon^2}{4^u\ 8\ (t+1)^2\ B^2}\right) du \\
&= \int_0^{K_\epsilon} \exp\left(-\frac{n\ t^2\ \varepsilon^2}{4^u\ 8\ (t+1)^2\ B^2}\right) du.
\end{aligned}$$

Invoking the change of variable formula, we may further proceed

$$\begin{aligned}
&\int_0^{K_\epsilon} \exp\left(-\frac{n\ t^2\ \varepsilon^2}{4^u\ 8\ (t+1)^2\ B^2}\right) du \\
&\leq \int_{4^{-K_\epsilon}}^1 \exp\left(-y\ \frac{n\ t^2\ \varepsilon^2}{8\ (t+1)^2\ B^2}\right)\ \frac{1}{y\ \ln(4)} dy \\
&\leq \frac{4^{K_\varepsilon}}{\ln(4)}\ \int_{4^{-K_\varepsilon}}^1 \exp\left(-y\ \frac{n\ t^2\ \varepsilon^2}{8\ (t+1)^2\ B^2}\right) dy \\
&\leq \frac{1}{2304\ t^2\ \ln(4)}\ \exp\left(-\frac{n\ t^2\ \varepsilon^2}{8\ 4^{K_\varepsilon}\ (t+1)^2\ B^2}\right),
\end{aligned}$$

where in the last step assumption $\varepsilon > 2^{K_\varepsilon - 1}\ 128\ (t+1)\ \sqrt{2}\ B\big[J\big(\mathbb{F}^\Theta, B, 1/8\big) + 3/2\big]/\sqrt{n}$ has been taken into account. The proof is now complete.

7.3. Auxilary Lemma for Proofs of Theorems 3.1, 3.8 and 5.1

The following lemma is repeatedly used in proofs of Theorems 3.1, 3.8, and 5.1. In Theorems 3.1 and 3.8, the goal functions are assumed to be uniformly bounded by some $B \in \mathbb{R}$, whereas in Theorem 5.1, goal functions are allowed to have unbounded support. Therefore, to state the following lemma for both the bounded and unbounded cases, we adopt the following rules for interpreting the notations $\phi_{r,n}$ and $w_{r,n}$. Namely, for the proofs of Theorems 3.1 and 3.8, Lemma 7.2 is applied with $w_{r,n} = B$ and $\phi_{r,n}(x) = x$, whereas for the proof of Theorem 5.1, Lemma 7.2 is applied with $w_{r,n} = (2n)^{1/r}\|\xi\|_{\mathbb{P}^Z,r}$ and $\phi_{r,n}(x) = (x \wedge w_{r,n}) \vee (-w_{r,n})$. Moreover, we denote the function class

$$\mathbb{F}^{\Theta,\phi_{r,n}}_{w_{r,n}} := \Big\{ \big[\phi_{r,n}\big(G(\theta,\cdot)\big) - x\big]^+ \mid \theta \in \Theta, x \in [-w_{r,n}, w_{r,n}] \Big\}.$$

Lemma 7.2 *Let (A 1) - (A 3) be fulfilled. Furthermore, let $\overline{\Theta}$ be the at most countable subset of Θ from (A 3). Then for any $n \in \mathbb{N}$, the function class $\mathbb{F}^{\Theta,\phi_{r,n}}_{w_{r,n}}$ satisfies the following properties.*

1) For countable dense subset I of $[-w_{r,n}, w_{r,n}]$.

$$\sup_{f \in \mathbb{F}^{\Theta,\phi_{r,n}}_{w_{r,n}}} \big|(\mathbb{P}_n - \mathbb{P})(f)\big| = \sup_{(\theta,x) \in \overline{\Theta} \times I} \Big|(\mathbb{P}_n - \mathbb{P})\Big(\big(\phi_{r,n}/ \circ G(\theta,\cdot) - x\big)^+\Big)\Big| \ \mathbb{P} - a.s..$$

2) If I is some countable dense subset of $[-w_{r,n}, w_{r,n}]$, then a positive envelope of $\overline{\mathbb{F}}^{r,n}_I := \big\{ \big(\phi_{r,n} \circ G(\theta,\cdot) - x\big)^+ \mid \theta \in \overline{\Theta}, x \in I \big\}$ is given by $\xi + w_{r,n}$, and

$$J(\overline{\mathbb{F}}^{r,n}_I, \xi + w_{r,n}, \delta) \leq 4\ J(\mathbb{F}^{\Theta}, \xi, \delta/4) + 6 \quad \textit{for } \delta \in (0,4].$$

Proof Statement 1) follows immediately from assumption (A 3) along with the continuity of $\phi_{r,n}$ because

$$\begin{aligned}
&\big|\big(\phi_{r,n} \circ G(\theta,z) - x\big)^+ - \big(\phi_{r,n} \circ G(\vartheta,z) - y\big)^+\big| \\
&\leq |\phi_{r,n} \circ G(\theta,z) - \phi_{r,n} \circ G(\vartheta,z)| + |x-y| \\
&\leq |G(\theta,z) - G(\vartheta,z)| + |x-y|, \quad \text{for } \theta,\vartheta \in \Theta \text{ and } x,y \in \mathbb{R}.
\end{aligned}$$

Concerning statement 2), let I be any countable dense subset of $[-w_{r,n}, w_{r,n}]$. Condition (A 2) ensures that $\xi + w_{r,n}$ is a positive envelope of the function class $\overline{\mathbb{F}}^{r,n}_I$ defined in the display of statement 2).

For $\theta \in \Theta, x \in \mathbb{R}$, we set $F(\theta,x,\cdot) := \big(G(\theta,\cdot) - x\big)^+$, so that by definition the class $\overline{\mathbb{F}}^{r,n}_I$ consists of all functions $F(\theta,x,\cdot)$ with $(\theta,x) \in \overline{\Theta} \times I$. In the next step, we want to derive upper estimates for the uniform entropy integrals $J\big(\overline{\mathbb{F}}^{r,n}_I, \xi + w_{r,n}, \delta\big)$. Therefore, let us fix $\eta > 0$ and $\mathbb{Q} \in \mathcal{M}_{\text{fin}}$ with support $supp(\mathbb{Q})$. Consider any $(\theta,x), (\vartheta,y) \in \overline{\Theta} \times I$ such that the inequalities $\|G(\theta,\cdot) - G(\vartheta,\cdot)\|_{\mathbb{Q},2} \leq \eta\ \|\xi\|_{\mathbb{P}^Z,2}/2$ and $|x-y| \leq \eta\ w_{r,n}/2$ are valid. Since $\phi_{r,n}$ is 1-Lipschitz continuous, then

$$\begin{aligned}
\|F(\theta,x,\cdot) - F(\vartheta,y,\cdot)\|^2_{\mathbb{Q},2} &= \sum_{z \in supp(\mathbb{Q})} \mathbb{Q}(\{z\})\ \ |F(\theta,x,z) - F(\vartheta,y,z)|^2 \\
&\leq 4 \sum_{z \in supp(\mathbb{Q})} \mathbb{Q}(\{z\})\ \ \big[|G(\theta,z) - G(\vartheta,z)|^2 + |x-y|^2\big] \\
&\leq 4\ \|G(\theta,\cdot) - G(\vartheta,\cdot)\|^2_{\mathbb{Q},2} + 4\ |x-y|^2 \\
&\leq \eta^2\ \|\xi + w_{r,n}\|^2_{\mathbb{Q},2}.
\end{aligned}$$

Hence, setting $\mathbb{F}^{\overline{\Theta}} := \{G(\theta,\cdot) \mid \theta \in \overline{\Theta}\}$,

$$
\begin{aligned}
&N\big(\eta\|\xi + w_{r,n}\|_{\mathbb{Q},2}, \overline{\mathbb{F}}^{r,n}, L^2(\mathbb{Q})\big)\\
&\leq N\big(\eta\|\xi\|_{\mathbb{Q},2}/2, \mathbb{F}^{\overline{\Theta}}, L^2(\mathbb{Q})\big) \cdot N\big(\eta\ w_{r,n}/2, I, |\cdot|\big)\\
&\leq N\big(\eta\|\xi\|_{\mathbb{Q},2}/4, \mathbb{F}^{\Theta}, L^2(\mathbb{Q})\big) \cdot N\big(\eta\ w_{r,n}/4, [-w_{r,n}, w_{r,n}], |\cdot|\big),
\end{aligned}
$$

where for nonvoid bounded $J \subseteq \mathbb{R}$ and $\overline{\eta} > 0$, we denote by $N\big(\overline{\eta}, J, |\cdot|\big)$ the minimal number to cover J by intervals of the form $[x - \overline{\eta}, x + \overline{\eta}]$ with $x \in J$. Note that

$$
N\big(\eta\ w_{r,n}/4, [-w_{r,n}, w_{r,n}], |\cdot|\big) \leq 8/\eta,
$$

holds. Then using change of variable formula we may conclude for any $\delta \in (0, 4]$

$$
\begin{aligned}
&J(\overline{\mathbb{F}}_I^{r,n}, \xi + (w_{r,n} + B), \delta)\\
&= \int_0^\delta \sup_{\mathbb{Q}\in\mathcal{M}_{\text{fin}}} \sqrt{\log\big(2\ N(\eta\ \|\xi + (w_{r,n} + B)\|_{\mathbb{Q}}, \overline{\mathbb{F}}_I^{r,n}, L^2(\mathbb{Q})\big)}\mathrm{d}\eta\\
&\leq \int_0^\delta \sup_{\mathbb{Q}\in\mathcal{M}_{\text{fin}}} \sqrt{\log\left(2N(\eta\|\xi\|_{\mathbb{Q},2}/4, \mathbb{F}^{\Theta}, L^2(\mathbb{Q})\right) + \log(8/\eta)}\mathrm{d}\eta\\
&\leq \int_0^\delta \sup_{\mathbb{Q}\in\mathcal{M}_{\text{fin}}} \sqrt{\log\left(2N(\eta\|\xi\|_{\mathbb{Q},2}/4, \mathbb{F}^{\Theta}, L^2(\mathbb{Q})\right)}\mathrm{d}\eta + \int_0^\delta \sqrt{\log(8/\eta)}\mathrm{d}\eta\\
&= 4\ J(\mathbb{F}^{\Theta}, \xi, \delta/4) + 4\ \int_0^{\delta/4} \sqrt{\log(2/u)}\ du.
\end{aligned}
$$

Furthermore, we may use change of variable formula along with integration by parts to obtain

$$
\begin{aligned}
\int_0^{\delta/4} \sqrt{\log(2/\epsilon)}\mathrm{d}\epsilon \leq \int_0^1 \sqrt{\log(e/\epsilon)}\mathrm{d}\epsilon &= \int_0^\infty \sqrt{1+t}e^{-t}\mathrm{dt}\\
&\leq \int_0^\infty (1 + t/2)e^{-t}dt = 3/2.
\end{aligned}
$$

This shows statement 2) and completes the proof. □

7.4. Proofs of results from Section 4

We start with the derivation of Lemma 4.1.

Proof of Lemma 4.1 By definition, we have

$$
\rho_h(X) = \int_0^\infty h\left(\mathbb{P}[X > t]\right) dt + \int_{-\infty}^0 \left(h\left(\mathbb{P}[X > t]\right) - 1\right) dt.
$$

Let $F^{\leftarrow}$ be the quantile function of X, and p_0 the separation point between concave and convex part of h. We note that for any $t > F^{\leftarrow}(1-p_0)$, we have that $\mathbb{P}(X > t) \leq p_0$. Suppose first that $F^{\leftarrow}(1-p_0) \geq 0$. Then, we have that

$$\begin{aligned}
&\int_0^{\infty} h\left(\mathbb{P}[X > t]\right) dt \\
&= \int_0^{F^{\leftarrow}(1-p_0)} h\left(\mathbb{P}[X > t]\right) dt + \int_{F^{\leftarrow}(1-p_0)}^{\infty} h\left(\mathbb{P}[X > t]\right) dt \\
&= \int_0^{F^{\leftarrow}(1-p_0)} 1 - \bar{h}\left(\mathbb{P}[X \leq t]\right) dt + \int_{F^{\leftarrow}(1-p_0)}^{\infty} h_0\left(\mathbb{P}[X > t]\right) dt \\
&= \int_0^{F^{\leftarrow}(1-p_0)} 1 - h(p_0) - \bar{h}\left(\mathbb{P}[X \leq t]\right) dt + \int_0^{\infty} h_0\left(\mathbb{P}[X > t]\right) dt.
\end{aligned}$$

We also note that $t \geq -F^{\leftarrow}(1-p_0) \Leftrightarrow \mathbb{P}(-X \geq t) \leq 1-p_0$. Hence, the first integral is also equal to:

$$\begin{aligned}
&\int_0^{F^{\leftarrow}(1-p_0)} 1 - h(p_0) - \bar{h}\left(\mathbb{P}[-X \geq -t]\right) dt \\
&= \int_{-F^{\leftarrow}(1-p_0)}^{0} 1 - h(p_0) - \bar{h}\left(\mathbb{P}[-X \geq s]\right) ds \\
&= \int_{-\infty}^{0} 1 - h(p_0) - \bar{h}_0\left(\mathbb{P}[-X \geq s]\right) ds \\
&= \int_{-\infty}^{0} 1 - h(p_0) - \bar{h}_0\left(\mathbb{P}[-X > s]\right) ds.
\end{aligned}$$

Furthermore, we have

$$\begin{aligned}
\int_{-\infty}^{0} \left(h\left(\mathbb{P}[X > t]\right) - 1\right) dt &= -\int_{-\infty}^{0} \bar{h}(\mathbb{P}(X \leq t)) dt \\
&= -\int_0^{\infty} \bar{h}(\mathbb{P}(-X \geq s)) ds \\
&= -\int_0^{\infty} \bar{h}_0(\mathbb{P}(-X > s)) ds.
\end{aligned}$$

Moreover,

$$\int_{-\infty}^{0} h_0(\mathbb{P}(X > t)) - h(p_0) dt = 0,$$

since for $t \leq F^{\leftarrow}(1-p_0)$, $h_0(\mathbb{P}(X > t)) = h(p_0)$. Therefore, assembling all equalities yields that

$$\begin{aligned}
\rho_h(X) &= \int_0^\infty h_0\left(\mathbb{P}[X > t]\right) dt + \int_{-\infty}^0 h_0(\mathbb{P}(X > t)) - h(p_0) dt \\
&\quad - \left(\int_0^\infty \bar{h}_0(\mathbb{P}(-X > s)) ds + \int_{-\infty}^0 \bar{h}_0\left(\mathbb{P}[-X > s]\right) - (1 - h(p_0)) ds \right) \\
&= h(p_0)\rho_{h_0/h(p_0)}(X) - (1 - h(p_0))\rho_{\bar{h}_0/(1-h(p_0))}(-X).
\end{aligned}$$

Now, suppose $F^{\leftarrow}(1-p_0) < 0$. Then,

$$\begin{aligned}
&\int_{-\infty}^0 \left(h\left(\mathbb{P}[X > t]\right) - 1\right) dt \\
&= \int_{-\infty}^{F^{\leftarrow}(1-p_0)} \left(h\left(\mathbb{P}[X > t]\right) - 1\right) dt + \int_{F^{\leftarrow}(1-p_0)}^0 \left(h\left(\mathbb{P}[X > t]\right) - 1\right) dt \\
&= - \int_{-\infty}^{F^{\leftarrow}(1-p_0)} \bar{h}\left(\mathbb{P}[X \leq t]\right) dt + \int_{F^{\leftarrow}(1-p_0)}^0 \left(h_0\left(\mathbb{P}[X > t]\right) - 1\right) dt \\
&= - \int_{-F^{\leftarrow}(1-p_0)}^\infty \bar{h}_0\left(\mathbb{P}[-X \geq s]\right) ds + \int_{F^{\leftarrow}(1-p_0)}^0 \left(h_0\left(\mathbb{P}[X > t]\right) - 1\right) dt.
\end{aligned}$$

We note that

$$\begin{aligned}
&\int_{-F^{\leftarrow}(1-p_0)}^\infty \bar{h}_0\left(\mathbb{P}[-X \geq s]\right) ds \\
&= \int_0^\infty \bar{h}_0\left(\mathbb{P}[-X \geq s]\right) ds + (1 - h(p_0))F^{\leftarrow}(1-p_0),
\end{aligned}$$

and

$$\begin{aligned}
&\int_{F^{\leftarrow}(1-p_0)}^0 \left(h_0\left(\mathbb{P}[X > t]\right) - 1\right) dt \\
&= \int_{-\infty}^0 \left(h_0\left(\mathbb{P}[X > t]\right) - h(p_0)\right) dt + (1 - h(p_0))F^{\leftarrow}(1-p_0).
\end{aligned}$$

Hence,

$$\begin{aligned}
&\int_{-\infty}^0 \left(h\left(\mathbb{P}[X > t]\right) - 1\right) dt \\
&= \int_{-\infty}^0 \left(h_0\left(\mathbb{P}[X > t]\right) - h(p_0)\right) dt - \int_0^\infty \bar{h}_0\left(\mathbb{P}[-X \geq s]\right) ds.
\end{aligned}$$

Moreover,

$$\int_0^\infty h(\mathbb{P}(X > t))dt = \int_0^\infty h_0(\mathbb{P}(X > t))dt,$$

and

$$\int_{-\infty}^{0} \bar{h}_0(\mathbb{P}(-X > s)) - (1 - h(p_0))ds = 0.$$

Hence, we again have

$$\begin{aligned}
\rho_h(X) = &\int_0^\infty h_0(\mathbb{P}(X > t))dt + \int_{-\infty}^{0} \left(h_0\left(\mathbb{P}[X > t]\right) - h(p_0)\right) dt \\
&- \int_0^\infty \bar{h}_0\left(\mathbb{P}[-X \geq s]\right) ds - \int_{-\infty}^{0} \bar{h}_0(\mathbb{P}(-X > s)) - (1 - h(p_0))ds.
\end{aligned}$$

□

Let us turn over to the proof of Theorem 4.2

Proof of Theorem 4.2 Let $\theta \in \Theta$ and $F_\theta^-, \widehat{F}_{n,\theta}^-$ denote respectively the true and empirical distribution functions of $-G(\theta, Z)$. Then, we have that

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\theta\in\Theta}\left|\rho_h(\widehat{F}_{n,\theta}) - \rho_h(F_\theta)\right| > \epsilon\right) \\
&\leq \mathbb{P}\left(\sup_{\theta\in\Theta}\left|\rho_{h_{cc}}(\widehat{F}_{n,\theta}) - \rho_{h_{cc}}(F_\theta)\right| \geq \epsilon/(2h(p_0))\right) \\
&\quad + \mathbb{P}\left(\sup_{\theta\in\Theta}\left|\rho_{h_{cv}}(\widehat{F}_{n,\theta}^-) - \rho_{h_{cv}}(F_\theta^-)\right| \geq \epsilon/(2(1 - h(p_0)))\right).
\end{aligned}$$

Note that the class $\{-G(\theta, Z) : \theta \in \Theta\}$ is also uniformly bounded by B and has the same covering numbers as the class $\mathbb{F}^\Theta$. Hence, applying Theorem 5.1 to each separate term gives the statement. □

7.5. Proof of Theorem 5.1

To begin the proof, let us first introduce the following notations:

$$M_{q,\delta} := \sup_{\substack{\beta\in\mathrm{dom}(\beta)\\ \beta(\mu)\leq b_\delta}} \int_{[0,1)} \frac{1}{(1-\alpha)^{1/q}}\,\mu(d\alpha), \quad b_\delta := 2\rho\big(2\xi(Z_1)\big) + \delta \quad (\delta > 0) \tag{7.12}$$

$$L(n, q, r, \delta) := \left(\frac{q}{r\ (2n)^{(r-q)/r}}\right)^{1/q}\ 2\ M_{q,\delta}\ \|\xi\|_{\mathbb{P}^Z, r}\ (\delta > 0). \tag{7.13}$$

The following proposition allows us to reduce the unbounded problem to a bounded one.

Proposition 7.3 *If $r \in [2, \infty) \cap (q, \infty)$, then $M_{q,\delta} \leq \overline{M}_{r,q}^\xi + b_\delta$ and the inequalities*

$$\begin{aligned}
&\mathbb{P}\Big(\Big\{\big|\inf_{\theta\in\Theta}\mathcal{R}_\rho\big(\widehat{F}_{n,\theta}\big) - \inf_{\theta\in\Theta}\mathcal{R}_\rho\big(F_\theta\big)\big| \geq \varepsilon\Big\}\ \cap\ B_{n,r}^\xi\Big) \\
&\leq \mathbb{P}^*\Big(\Big\{\sup_{\theta\in\Theta}\big|\mathcal{R}_\rho\big(\widehat{F}_{n,\theta}\big) - \mathcal{R}_\rho\big(F_\theta\big)\big| \geq \varepsilon\Big\}\ \cap\ B_{n,r}^\xi\Big) \\
&\leq \mathbb{P}^*\Big(\Big\{\sup_{\theta\in\Theta}\big|\mathcal{R}_\rho\big(\widehat{F}_{n,\theta}^{\phi_{r,n}}\big) - \mathcal{R}_\rho\big(F_\theta^{\phi_{r,n}}\big)\big| \geq \epsilon - L(n, r, q, \delta)\Big\}\ \cap\ B_{n,r}^\xi\Big)
\end{aligned}$$

hold for any $\delta, \varepsilon > 0$ and every $n \in \mathbb{N}$. Here $\mathbb{P}^*$ denotes the outer probability w.r.t. $\mathbb{P}$.

The proof is delegated to Section 7.6.

In the next step we want to bound the penalty function in Kusuoka representation (2.2) for $\mathcal{R}_\rho(\widehat{F}^{\phi_{r,n}}_{n,\theta})$ and $\mathcal{R}_\rho(F^{r,n}_\theta)$ on the auxiliary event $B^\xi_{n,r}$.

Proposition 7.4 *Under assumptions (A 1), (A 2) and (A 6) with q, r as well as $\overline{M}^\xi_{q,r}$ as in (A 6), we obtain the following Kusuoka representation for $\theta \in \Theta, n \in \mathbb{N}$ and $\delta > 0$:*

$$\mathcal{R}_\rho(\widehat{F}^{\phi_{r,n}}_{n,\theta|_\omega}) = \sup_{\substack{\mu \in \mathrm{dom}(\beta) \\ \beta(\mu) \leq 2\overline{M}^\xi_{r,q} + \delta}} \Big(\int_{[0,1)} \mathcal{R}_\alpha(\widehat{F}^{\phi,r}_{n,\theta|_\omega})\ \mu(d\alpha) - \beta(\mu) \Big) \quad \text{if } \omega \in B^\xi_{n,r},$$

$$\mathcal{R}_\rho(F^{\phi_{r,n}}_\theta) = \sup_{\substack{\mu \in \mathrm{dom}(\beta) \\ \beta(\mu) \leq 2\overline{M}^\xi_{r,q} + \delta}} \Big(\int_{[0,1)} \mathcal{R}_\alpha(F^{\phi_{r,n}}_\theta)\ \mu(d\alpha) - \beta(\mu) \Big),$$

where $\mathcal{R}_\alpha$ stands for the functional associated with $AV@R_\alpha$.

Proof Let $\theta \in \Theta, n \in \mathbb{N}$ and $\delta > 0$. We fix any $\omega \in B^\xi_{n,r}$ and denote by $\widehat{F}_{n,\xi}(\cdot, \omega)$ the empirical distribution function based on $\xi\big(Z_1(\omega)\big), \ldots, \xi\big(Z_n(\omega)\big)$.

Since $(\Omega, \mathcal{F}, \mathbb{P})$ is atomless, we may find random variables $X^{(\omega)}, Y^{(\omega)}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ which have respectively $\widehat{F}_{n,\xi}(\cdot, \omega)$ and $\widehat{F}^{\phi_{r,n}}_{n,\theta|_\omega}$ as distribution functions. Both random variables are $\mathbb{P}$-essentially bounded so that they belong to $\mathcal{X}$. Furthermore, the distribution functions $F_{-X^{(\omega)}}, F_{Y^{(\omega)}}, F_{X^{(\omega)}}$ of $-X^{(\omega)}, Y^{(\omega)}$ and $X^{(\omega)}$ respectively, satisfy pointwise the inequalities $F_{X^{(\omega)}} \leq F_{Y^{(\omega)}} \leq F_{-X^{(\omega)}}$ due to (A 2). Moreover, by (A 2) again, we also have $-\xi(Z_1) \leq \phi_{r,n}(G(\theta, Z_1)) \leq \xi(Z_1)$. Hence, by Proposition A.2 (see Appendix A)

$$\mathcal{R}_\rho(\widehat{F}^{\phi_{r,n}}_{n,\theta|_\omega}) = \sup_{\substack{\mu \in \mathrm{dom}(\beta) \\ \beta(\mu) \leq 2\rho(2X^{(\omega)}) + \delta}} \Big(\int_{[0,1)} \mathcal{R}_\alpha(\widehat{F}^{\phi_{r,n}}_{n,\theta|_\omega})\ \mu(d\alpha) - \beta(\mu) \Big) \quad \text{if } \omega \in B^\xi_{n,r},$$

$$\mathcal{R}_\rho(F^{\phi_{r,n}}_\theta) = \sup_{\substack{\mu \in \mathrm{dom}(\beta) \\ \beta(\mu) \leq 2\rho(2\xi(Z_1)) + \delta}} \Big(\int_{[0,1)} \mathcal{R}_\alpha(F_\theta)\ \mu(d\alpha) - \beta(\mu) \Big),$$

Furthermore, by Lemma 4.3 in [1], and recalling $r \geq q$ as well as $\omega \in B^\xi_{n,r}$

$$AV@R_\alpha(2X^{(\omega)}) \leq \frac{2\left(\frac{1}{n}\sum_{j=1}^n \xi_j\big(Z_j(\omega)\big)^q\right)^{1/q}}{(1-\alpha)^{1/q}} \leq \frac{2\left(\frac{1}{n}\sum_{j=1}^n \xi_j\big(Z_j(\omega)\big)^r\right)^{1/r}}{(1-\alpha)^{1/q}}$$
$$\leq \frac{2^{1+1/r}\|\xi\|_{\mathbb{P}^Z,r}}{(1-\alpha)^{1/q}}.$$

Also, by Lemma 4.3 in [1]

$$AV@R_\alpha(2\xi) \leq \frac{\|\xi\|_{\mathbb{P}^Z,q}}{(1-\alpha)^{1/q}} \leq \frac{\|\xi\|_{\mathbb{P}^Z,r}}{(1-\alpha)^{1/q}}.$$

So we may conclude $2\rho(2X^{(\omega)}) \vee 2\rho(2\xi) \leq 2\overline{M}^{\xi}_{r,q}$ which completes the proof. $\square$

Now we are ready to show Theorem 5.1.

Proof of Theorem 5.1. Let us fix $n \in \mathbb{N}$ and $\delta, \varepsilon > 0$, and let I be any countable dense subset of $[-w_{r,n}, w_{r,n}]$, whereas $\overline{\Theta}$ is the at most countable subset of Θ from assumption (A 3).

Clearly, the inequality

$$\mathbb{P}\left(\left\{\left|\inf_{\theta\in\Theta}\mathcal{R}_\rho(F_\theta) - \inf_{\theta\in\Theta}\mathcal{R}_\rho(\widehat{F}_{n,\theta})\right| \geq \varepsilon\right\}\right) \leq \mathbb{P}^*\left(\left\{\sup_{\theta\in\Theta}\left|\mathcal{R}_\rho(F_\theta) - \mathcal{R}_\rho(\widehat{F}_{n,\theta})\right| \geq \varepsilon\right\}\right)$$

is valid. Denote $b_{r,\delta} := 2\overline{M}^{\xi}_{r.q} + \delta$ and $\mathcal{M}_{b_{r,\delta}} := \{\mu \in \mathcal{M} \mid \beta_\rho(\mu) \leq b_{r,\delta}\}$. Then, combining Proposition 7.3, Lemma 7.2 and mimicking the proof of Theorem 3.1 (by replacing $\mathcal{M}_{B,\delta}$ with $\mathcal{M}_{b_{r,\delta}}$ and $N_{q,4B+\delta}$ with $N_{q,b_{r,\delta}}$ (which is finite due to (A 6)), we have that

$$\begin{aligned}
&\mathbb{P}^*\left(\left\{\sup_{\theta\in\Theta}\left|\mathcal{R}_\rho(F_\theta) - \mathcal{R}_\rho(\widehat{F}_{n,\theta})\right| \geq \varepsilon\right\} \cap B^{\xi}_{n,r}\right) \\
&\leq \mathbb{P}^*\Big(\Big\{\sup_{\theta\in\Theta}\big|\mathcal{R}_\rho\big(\widehat{F}^{\phi_{r,n}}_{n,\theta}\big) - \mathcal{R}_\rho\big(F^{\phi_{r,n}}_{\theta}\big)\big| \geq \epsilon - L(n,r,q,\delta)\Big\} \ \cap \ B^{\xi}_{n,r}\Big) \\
&\leq \mathbb{P}^*\left(\left\{\sup_{f\in\overline{\mathbb{F}}^{r,n}_I}\big|(\mathbb{P}_n - \mathbb{P})(f)\big| \geq \frac{[\varepsilon - L(n,q,r,\delta)]^q}{N^q_{q,b_{r,\delta}} 2^{2q-1} w^{(q-1)}_{r,n}}\right\}\right), \qquad (7.14)
\end{aligned}$$

where $\overline{\mathbb{F}}^{r,n}_I := \big\{\big(\phi_{r,n} \circ G(\theta,\cdot) - x\big)^+ \mid (\theta, x) \in \overline{\Theta} \times I\big\}$, and $L(n,q,r,\delta)$ is as in (7.13). In order to find upper estimations for the deviation probability on the right-hand side of inequality (7.14), we want to apply Theorem B.1 in Appendix B to $\overline{F}^{r,n}_I$.

Firstly, every member of this class is Borel measurable, and $\xi + w_{r,n}$ is a positive envelope. Since in addition $\overline{F}^{r,n}_I$ is at most countable, we may apply Theorem 2.1 from [14] to conclude

$$\mathbb{E}\left[\sup_{f\in\overline{\mathbb{F}}^{r,n}_I}\big|(\mathbb{P}_n - \mathbb{P})(f)\big|\right] \leq \frac{16\sqrt{2}\,\|\xi\|_{\mathbb{P}^Z,2}}{\sqrt{n}}\ J\big(\overline{\mathbb{F}}^{r,n}_I, \xi + w_{r,n}, 1/2\big).$$

Therefore by Lemma 7.2

$$\mathbb{E}\left[\sup_{f\in\overline{\mathbb{F}}^{r,n}_I}\big|(\mathbb{P}_n - \mathbb{P})(f)\big|\right] \leq \frac{64\sqrt{2}\,\|\xi\|_{\mathbb{P}^Z,2}}{\sqrt{n}}\ \big[J\big(\mathbb{F}^{\Theta}, \xi, 1/8\big) + 3/2\big]. \qquad (7.15)$$

For abbreviation, we set

$$\overline{\varepsilon} := \frac{\varepsilon}{N_{q,b_{r,\delta}} 2^{(2q-1)/q} w^{(q-1)/q}_{r,n}}$$

$$\overline{L} := \frac{L(n,q,r,\delta)}{N_{q,b_{r,\delta}} 2^{(2q-1)/q} w^{(q-1)/q}_{r,n}}$$

Now, if ε satisfies the lower bound in the statement, then inequality (7.15) implies

$$(\overline{\varepsilon} - \overline{L})^q > \frac{\overline{\varepsilon}^q}{2^q} \vee \mathbb{E}\left[(t+1) \sup_{f \in \overline{\mathbb{F}}_I^{r,n}} \big|(\mathbb{P}_n - \mathbb{P})(f)\big|\right].$$

Hence

$$(\overline{\varepsilon} - \overline{L})^q = \frac{t}{t+1}\,(\overline{\varepsilon} - \overline{L})^q + \frac{(\overline{\varepsilon} - \overline{L})^q}{t+1} > \frac{t}{t+1}\,\frac{\overline{\varepsilon}^q}{2^q} + \mathbb{E}\left[\sup_{f \in \overline{\mathbb{F}}_I^{r,n}} \big|(\mathbb{P}_n - \mathbb{P})(f)\big|\right]$$

Finally, note that the mappings from $\overline{\mathbb{F}}_I^{r,n}$ are uniformly bounded by $w_{r,n}$. Thus, applying Theorem B.1 (see Appendix B) to $\overline{\mathbb{F}}_I^{r,n}$, we may derive from (7.14)

$$\begin{aligned}
&\mathbb{P}\left(\left\{\sup_{\theta \in \Theta} \Big|\mathcal{R}_\rho(F_\theta) - \mathcal{R}_\rho(\widehat{F}_{n,\theta})\Big| \geq \varepsilon\right\} \cap B^{\xi}_{n,r}\right) \\
&\leq \mathbb{P}\left(\left\{\sup_{f \in \overline{\mathbb{F}}_I^{r,n}} \big|(\mathbb{P}_n - \mathbb{P})(f)\big| \geq (\overline{\varepsilon} - \overline{L})^q\right\}\right) \\
&\leq \mathbb{P}\left(\left\{\sup_{f \in \overline{\mathbb{F}}_I^{r,n}} \big|(\mathbb{P}_n - \mathbb{P})(f)\big| \geq \frac{t}{t+1}\,\frac{\overline{\varepsilon}^q}{2^q} + \mathbb{E}\left[\sup_{f \in \overline{\mathbb{F}}_I^{r,n}} \big|(\mathbb{P}_n - \mathbb{P})(f)\big|\right]\right\}\right) \\
&\leq \exp\left(-\frac{n\, t^2\, \overline{\varepsilon}^{2q}}{2^{2q+1}\,(t+1)^2 w_{r,n}^2}\right).
\end{aligned}$$

□

7.6. Proof of Proposition 7.3

The starting point is the following auxiliary result.

Lemma 7.5 *Let ξ from (A 2) be $\mathbb{P}^Z$-integrable of order r for some $r \geq 2$. Then,*

$$\begin{aligned}
&\left\{\big|\inf_{\theta \in \Theta} \mathcal{R}_\rho\big(\widehat{F}_{n,\theta}\big) - \inf_{\theta \in \Theta} \mathcal{R}_\rho\big(F_\theta\big)\big| \geq \epsilon\right\} \ \cap\ B^{\xi}_{n,r} \\
&\subseteq \left\{\sup_{\theta \in \Theta} \big|\mathcal{R}_\rho\big(\widehat{F}_{n,\theta}\big) - \mathcal{R}_\rho\big(F_\theta\big)\big| \geq \epsilon\right\} \ \cap\ B^{\xi}_{n,r} \\
&\subseteq \left\{\sup_{\theta \in \Theta} \big|\mathcal{R}_\rho\big(\widehat{F}^{\phi_{r,n}}_{n,\theta}\big) - \mathcal{R}_\rho\big(F^{\phi_{r,n}}_{\theta}\big)\big| + \sup_{\theta \in \Theta} \big|\mathcal{R}_\rho\big(F^{\phi_{r,n}}_{\theta}\big) - \mathcal{R}_\rho\big(F_\theta\big)\big| \geq \epsilon\right\} \ \cap\ B^{\xi}_{n,r},
\end{aligned}$$

for $\epsilon > 0$ and $n \in \mathbb{N}$.

Proof This follows from the fact that on $B^{\xi}_{n,r}$, we have that

$$\frac{\xi(Z_j)^r}{n} \leq \frac{1}{n}\sum_{j=1}^{n} \xi(Z_j)^r \leq 2\mathbb{E}[\xi(Z_1)^r],$$

which implies that $\xi(Z_j) \leq (2n)^{1/r}\|\xi\|_{\mathbb{P}^Z,r}$. In particular, this means that $\widehat{F}^{\phi_{r,n}}_{n,\theta} = \widehat{F}_{n,\theta}$ on the event $B^{\xi}_{n,r}$. □

It remains to provide an upper estimation for

$$\sup_{\theta\in\Theta}\big|\mathcal{R}_\rho\big(F_\theta^{\phi_{r,n}}\big) - \mathcal{R}_\rho\big(F_\theta\big)\big|, \quad (n \in \mathbb{N}).$$

Proposition 7.6 *Let (A 1), (A 2) be fulfilled with ξ from (A 2), and let $r, q \in [1,\infty)$ and $\overline{M}^{\xi}_{r,q}$ be as in (A 6). If $r \in [2,\infty) \cap (q,\infty)$, then*

$$\sup_{\theta\in\Theta}\big|\mathcal{R}_\rho\big(F_\theta^{\phi_{r,n}}\big) - \mathcal{R}_\rho\big(F_\theta\big)\big| \leq L(n,r,q,\delta)\ \|\xi\|_{\mathbb{P}^Z,r}, \quad \textit{for } \delta > 0,\ n \in \mathbb{N}.$$

Proof Let $n \in \mathbb{N}$ and fix $\delta > 0$. We note that the following inequalities

$$-\xi(z) \leq G(\theta,z), \phi_{r,n}\big(G(\theta,z)\big) \leq \xi(z),$$

are valid for $\theta \in \Theta$ and $z \in \mathbb{R}^d$. Hence, by Proposition A.2 in Appendix A,

$$\begin{aligned}
&\big|\mathcal{R}_\rho\big(F_\theta^{\phi_{r,n}}\big) - \mathcal{R}_\rho\big(F_\theta\big)\big| \\
&\leq \sup_{\substack{\mu\in\mathrm{dom}(\beta)\\ \beta(\mu)\leq b_\delta}} \int_{[0,1)} \Big|AV@R_\alpha\Big(\phi_{r,n}\big(G(\theta,Z_1)\big)\Big) - AV@R_\alpha\big(G(\theta,Z_1)\big)\Big|\ \mu(d\alpha), \quad \text{for } \theta \in \Theta,
\end{aligned}$$

where $b_\delta := 2\rho\big(2\xi(Z_1)\big) + \delta$. Furthermore, by monotonicity and subadditivity of the Average Value at Risk, we obtain

$$\begin{aligned}
&\Big|AV@R_\alpha\Big(\phi_{r,n}\big(G(\theta,Z_1)\big)\Big) - AV@R_\alpha\big(G(\theta,Z_1)\big)\Big| \\
&\leq AV@R_\alpha\Big(\Big|\phi_{r,n}\big(G(\theta,Z_1)\big) - G(\theta,Z_1)\Big|\Big) \\
&= AV@R_\alpha\ \Big(\big(G(\theta,Z_1) + w_{r,n}\big)^- + \big(G(\theta,Z_1) - w_{r,n}\big)^+\Big) \\
&\leq AV@R_\alpha\ \Big(2\ \big(\xi(Z_1) - w_{r,n}\big)^+\Big),
\end{aligned}$$

for $\alpha \in [0,1)$. In addition, $2\ \big(\xi(Z_1) - w_{r,n}\big)^+$ is integrable of order $q < r$ by assumption on ξ, so that by the proof of Lemma 4.3 in [1]

$$AV@R_\alpha\ \Big(2\ \big(\xi(Z_1) - w_{r,n}\big)^+\Big) \leq \frac{2\ \big\|\big(\xi(Z_1) - w_{r,n}\big)^+\big\|_{\mathbb{P}^Z,q}}{\big(1-\alpha\big)^{1/q}}, \quad \text{for } \alpha \in [0,1).$$

Therefore, we may conclude

$$\begin{aligned}
&\sup_{\theta\in\Theta}\big|\mathcal{R}_\rho\big(F_\theta^{\phi_{r,n}}\big) - \mathcal{R}_\rho\big(F_\theta\big)\big| \\
&\leq \sup_{\substack{\mu\in\mathrm{dom}(\beta)\\ \beta(\mu)\leq b_\delta+\delta}} \int_{[0,1)} \frac{2\ \big\|\big(\xi(Z_1) - w_{r,n}\big)^+\big\|_{\mathbb{P}^Z,q}}{\big(1-\alpha\big)^{1/q}}\ \mu(d\alpha).
\end{aligned} \tag{7.16}$$

Since ξ is $\mathbb{P}^Z$-integrable of order $r > q$, and using in the first step the change of variable formula, we may further observe

$$\begin{aligned}\left\|\left(\xi - w_{r,n}\right)^+\right\|_{\mathbb{P}^Z,q}^q &= q\int_{w_{r,n}}^{\infty} (y - w_{r,n})^{q-1}\ \mathbb{P}\big(\{\xi(Z_1) > y\}\big)\ dy \\ &= \frac{q}{w_{r,n}^{r-q}} \int_{w_{r,n}}^{\infty} w_{r,n}^{r-q}(y - w_{r,n})^{q-1}\ \mathbb{P}\big(\{\xi(Z_1) > y\}\big)\ dy \\ &\leq \frac{q}{w_{r,n}^{r-q}} \int_{w_{r,n}}^{\infty} y^{r-1}\ \mathbb{P}\big(\{\xi(Z_1) > y\}\big)\ dy \\ &\leq \frac{q}{r\ w_{r,n}^{r-q}}\ \|\xi\|_{\mathbb{P}^Z,r}^r = \frac{q}{r\ (2n)^{(r-q)/r}}\ \|\xi\|_{\mathbb{P}^Z,r}^q. \end{aligned} \tag{7.17}$$

Finally,

$$\int_{[0,1)} \frac{1}{(1-\alpha)^{1/q}}\ \mu(d\alpha) \leq \int_{[0,1)} \frac{(4\|\xi\|_{\mathbb{P}^Z,q}) \vee 1}{(1-\alpha)^{1/q}}\ \mu(d\alpha) \leq \overline{M}_{r,q}^{\xi} + \beta(\mu), \quad \text{for } \mu \in \mathrm{dom}(\beta).$$

Now, the statement of Proposition 7.6 may be concluded easily from (7.16) along with (7.17). □

Now, Proposition 7.3 is an immediate consequence of Lemma 7.5 together with Proposition 7.6. □

8. Additional Details for Distortion Functions

8.1. Some Details on Table 1

- If $h(p) = (1-(1-p)^k)^{1/k}$, then

$$\lim_{p\downarrow 0} \frac{h(p)}{p^{1/k}} = \lim_{p\downarrow 0}\left(\frac{1-(1-p)^k}{p}\right)^{1/k} = \left(\lim_{p\downarrow 0}\frac{1-(1-p)^k}{p}\right)^{1/k} = k^{1/k}.$$

  Hence, $\overline{N}_q < \infty$ for $q \geq k$. Moreover, $\overline{N}_q \leq k^{1/k}$, for $q \geq k$, since $\overline{N}_k = k^{1/k}$, due to the fact that $\frac{1-(1-p)^k}{p}$ is decreasing in p (since it is the slope of the concave function $h(p) = 1-(1-p)^k$ between $[0,p]$).

- If $h(p) = p^a(1-\log(p^a))$ for $a \in (0,1)$, then $\lim_{p\downarrow 0} p^{a-1/q}(1-\log(p^a)) < \infty$ if and only if $q > 1/a$. To calculate $\overline{N}_q$, we examine the first order condition:

$$\begin{aligned} \frac{d}{dp}p^{a-1/q}(1-\log(p^a)) &= 0 \\ p^{a-1/q-1}(a-1/q)(1-\log(p^a)) - ap^{a-1/q}p^{-a+a-1} &= 0 \\ 1 - \frac{a}{a-1/q} &= a\log(p) \\ p &= \exp\left(\frac{-1}{a^2q-a}\right). \end{aligned}$$

Hence,

$$\begin{aligned}\overline{N}_q &= \exp\left(\frac{-a+1/q}{a^2q-a}\right)\left(1+\frac{1}{aq-1}\right)\\ &= \exp\left(\frac{-a+1/q}{a^2q-a}\right)\left(\frac{aq}{aq-1}\right).\end{aligned}$$

8.2. Extra Details On Kahneman-Tversky's Distortion Functions

Recall the Kahneman-Tversky's function

$$h(p) = 1 - \frac{(1-p)^\beta}{\left((1-p)^\beta + p^\beta\right)^{1/\beta}},\ 0<\beta<1.$$

Its dual function is given by

$$\bar{h}(p) = \frac{p^\beta}{\left(p^\beta + (1-p)^\beta\right)^{1/\beta}},\ 0<\beta<1.$$

To show $\overline{N}_{1/\beta,h_{cv}} < \infty$, it is sufficient to notice that

$$\lim_{p\downarrow 0}\frac{\bar{h}(p)}{p^\beta} = \lim_{p\downarrow 0}\frac{1}{\left(p^\beta + (1-p)^\beta\right)^{1/\beta}} = 1.$$

Now, to show $\overline{N}_{1/\beta,h_{cc}} < \infty$, we need to show $\lim_{p\downarrow 0} h(p)/p^\beta < \infty$. This is equal to

$$\begin{aligned}\lim_{p\downarrow 0}\frac{1-\frac{(1-p)^\beta}{\left((1-p)^\beta+p^\beta\right)^{1/\beta}}}{p^\beta} &= \lim_{p\downarrow 0}\frac{\left((1-p)^\beta+p^\beta\right)^{1/\beta}-(1-p)^\beta}{p^\beta}\cdot\frac{1}{\left((1-p)^\beta+p^\beta\right)^{1/\beta}}\\ &= \lim_{p\downarrow 0}\frac{\left((1-p)^\beta+p^\beta\right)^{1/\beta}-(1-p)^\beta}{p^\beta}\cdot\lim_{p\downarrow 0}\frac{1}{\left((1-p)^\beta+p^\beta\right)^{1/\beta}}\\ &= \lim_{p\downarrow 0}\frac{\left((1-p)^\beta+p^\beta\right)^{1/\beta}-(1-p)^\beta}{p^\beta},\end{aligned}$$

provided the last limit exists. Hence, to examine it, we use L'Hopital's rule and differentiate both the numerator and the denominator. This gives

$$\begin{aligned}&\lim_{p\downarrow 0}\frac{\frac{1}{\beta}\left((1-p)^\beta+p^\beta\right)^{1/\beta-1}\left(\beta p^{\beta-1}-\beta(1-p)^{\beta-1}\right)+\beta(1-p)^{\beta-1}}{\beta p^{\beta-1}}\\ &= \lim_{p\downarrow 0}\frac{1}{\beta}\left((1-p)^\beta+p^\beta\right)^{1/\beta-1}\left(1-(1-p)^{\beta-1}p^{1-\beta}\right)+(1-p)^{\beta-1}p^{1-\beta}\\ &= \frac{1}{\beta}.\end{aligned}$$

A. Kusuoka representation

Let $\rho : \mathcal{X} \to \mathbb{R}$ be the law-invariant convex risk measure on the $\mathbb{R}$-vector space $\mathcal{X}$ as introduced in Section 2. We want to verify the so-called Kusuoka representation for ρ.

Proposition A.1 *If ρ satisfies the Lebesgue property, then under the assumptions on $\mathcal{X}$ made in Section 2, there exists a set $\mathcal{M}$ of Borel probability measures on $[0,1)$ and a mapping $\beta : \mathcal{M} \to (-\infty, \infty]$ such that*

$$\rho(X) = \sup_{\mu \in \mathcal{M}} \left(\int_{[0,1)} AV@R_\alpha(X)\ \mu(d\alpha) - \beta(\mu) \right),$$

holds for every $X \in \mathcal{X}$. Moreover, if ρ is coherent, then $\beta(\mu) = 0$ for any μ from the effective domain of β.

Proof Since ρ fulfills the Lebesgue property, its restriction to $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ satisfies the following property

$$\rho(X_k) \nearrow \rho(X) \text{ whenever } X_k \nearrow X \text{ and } X_k, X \in L^\infty(\Omega, \mathcal{F}, \mathbb{P}). \tag{A.1}$$

Under this property, we may draw on Theorem 4.62 from [7] to find a set $\mathcal{M}$ of Borel probability measures on $[0,1)$ and a mapping $\beta : \mathcal{M} \to (-\infty, \infty]$ such that

$$\rho(X) = \sup_{\mu \in \mathcal{M}} \left(\int_{[0,1)} AV@R_\alpha(X)\ \mu(d\alpha) - \beta(\mu) \right), \tag{A.2}$$

for $X \in L^\infty(\Omega, \mathcal{F}, \mathbb{P})$. Moreover, $\beta(\mu) = 0$ for every μ with finite $\beta(\mu)$, in the case that ρ is even coherent.

Denoting the effective domain of β by $\mathrm{dom}(\beta)$, we may introduce via

$$\overline{\rho}(X) = \sup_{\mu \in \mathrm{dom}(\beta)} \left(\int_{[0,1)} AV@R_\alpha(X)\ \mu(d\alpha) - \beta(\mu) \right),$$

the mapping $\overline{\rho} : \mathcal{X} \to \mathbb{R} \cup \{\infty\}$. Recall that the mappings from $\mathcal{X}$ are assumed to be $\mathbb{P}$-integrable. We have

$$|AV@R_\alpha(X)| \leq AV@R_\alpha(|X|), \quad \text{for } \alpha \in (0,1) \text{ and } X \in \mathcal{X}, \tag{A.3}$$

because the Average Value at Risk satisfies monotonicity and sublinearity. Furthermore, the Average Value at Risk is continuous w.r.t. the L^1-norm (see [20, Proposition 3.1]). This means

$$\lim_{k \to \infty} AV@R_\alpha(|X| \wedge k) = AV@R_\alpha(|X|), \quad \text{for } \alpha \in (0,1), X \in \mathcal{X}. \tag{A.4}$$

Combining (A.3) and (A.4), we may conclude by monotone convergence that

$$\begin{aligned}
&\int_{[0,1)} |AV@R_\alpha(X)| \; \mu(d\alpha) - \beta(\mu) \\
&= |\mathbb{E}[X]| \; \mu(\{0\}) + \int_{(0,1)} |AV@R_\alpha(X)| \; \mu(d\alpha) - \beta(\mu) \\
&\leq \mathbb{E}[|X|] \; \mu(\{0\}) + \int_{(0,1)} AV@R_\alpha(|X|) \; \mu(d\alpha) - \beta(\mu) \\
&= \lim_{k\to\infty} \left(\mathbb{E}[|X| \wedge k]] + \int_{(0,1)} AV@R_\alpha(|X| \wedge k) \; \mu(d\alpha) - \beta(\mu) \right).
\end{aligned}$$

holds for any $\mu \in \operatorname{dom}(\beta)$ and every $X \in \mathcal{X}$. Moreover, as a law-invariant convex risk measure on a Banach lattice, ρ is norm-continuous (e.g., [20, Proposition 3.1]), and $\big\| |X| - |X| \wedge k \big\|_{\mathcal{X}} \to 0$ for $k \to \infty$, if $X \in \mathcal{X}$ due to (2.1). Hence, in view of (A.2), we end up with

$$|\mathbb{E}[X] \; \mu(\{0\})| + \int_{(0,1)} |AV@R_\alpha(X)| \; \mu(d\alpha) - \beta(\mu) \leq \lim_{k\to\infty} \rho(|X| \wedge k) = \rho(|X|),$$

for $\mu \in \operatorname{dom}(\beta)$ and $X \in \mathcal{X}$. In particular

$$\overline{\rho}(X) \leq \rho(|X|) < \infty, \quad \text{for } X \in \mathcal{X}.$$

Then $\overline{\rho}$ may be verified easily as a law-invariant convex risk measure so that it is also continuous w.r.t. $\|\cdot\|_{\mathcal{X}}$. Since ρ and $\overline{\rho}$ coincide on $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ by (A.2), and since $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ is a dense subset of $\mathcal{X}$ w.r.t. $\|\cdot\|_{\mathcal{X}}$ by (2.1), the norm continuity of both mappings yields that they also coincide on the entire space $\mathcal{X}$. This completes the proof. □

For any $X \in \mathcal{X}$, we denote its distribution function by F_X. If we restrict ρ to subsets $\{X \in \mathcal{X} \mid F_Y \leq F_X \leq F_{-Y} \text{ pointwise}\}$ for some nonnegative $Y \in \mathcal{X}$, we may bound the penalty function β in the Kusuoka representation.

Proposition A.2 *Let $Y \in \mathcal{X}$ with $Y \geq 0$ $\mathbb{P}$-a.s.. Then, with Kusuoka representation from Proposition A.1, we may find some $b \in \mathbb{R}$ such that*

$$\rho(X) = \sup_{\substack{\mu \in \operatorname{dom}(\beta) \\ \beta(\mu) \leq b+\delta}} \left(\int_{[0,1)} AV@R_\alpha(X) \; \mu(d\alpha) - \beta(\mu) \right),$$

holds for any $\delta > 0$, and arbitrary $X \in \mathcal{X}$ satisfying $F_Y \leq F_X \leq F_{-Y}$ pointwise. Here $\operatorname{dom}(\beta)$ *denotes the effective domain of β. Sufficient choices are $b = \rho(2\,Y) - 2\,\rho(-Y)$ or $b = 2\rho(2Y) - 3\rho(0)$.*

Proof Fix $\delta > 0$ and $X \in \mathcal{X}$ with $F_Y \leq F_X \leq F_{-Y}$ pointwise. This implies $F_{2Y} \leq F_{2X}$. We shall denote the expectation by $AV@R_0$. In view of the Kusuoka representation derived in Proposition A.1, we may restrict the supremum to all $\mu \in \mathcal{M}$ such that

$$\rho(X) - \delta/2 < \int_{[0,1)} AV@R_\alpha(X) \; \mu(d\alpha) - \beta(\mu).$$

Then, by positive homogeneity of expectation and Average Value at Risk along with the Kusuoka representation from Proposition A,

$$
\begin{aligned}
&\beta(\mu) - \delta/2 \\
&< \left[\int_{[0,1)} AV@R_\alpha(2X)\ \mu(d\alpha) - \beta(\mu)\right] - \rho(X) - \left[\int_{[0,1)} AV@R_\alpha(X)\ \mu(d\alpha) - \beta(\mu)\right] \\
&\leq \rho(2X) - \rho(X) - \rho(X) + \delta/2 = \rho(2X) - 2\rho(X) + \delta/2.
\end{aligned}
$$

Since the expectation as well as the Average Value at Risk are monotone, ρ satisfies this property as well by the Kusuoka representation. This means in particular, that $\rho(-Y) \leq \rho(X)$ due to $F_X \leq F_{-Y}$ pointwise, and $AV@R_\alpha(2X) \leq AV@R_\alpha(2Y)$ for every $\alpha \in [0,1)$ because $F_Y \leq F_X$ pointwise and thus $F_{2Y} \leq F_{2X}$ pointwise. Hence

$$\beta(\mu) \leq \rho(2Y) - 2\rho(-Y) + \delta,$$

and the first statement follows with $b := \rho(2Y) - 2\rho(-Y)$. Moreover, convexity of ρ implies $-\rho(-Y) \leq \rho(Y) - 2\rho(0)$. In addition, by Kusuoka representation $\rho(0) + \beta(\mu) \geq 0$ holds for $\mu \in \mathrm{dom}(\beta)$ so that by positive homogeneity of the Average Value at Risk we may conclude

$$
\begin{aligned}
2\rho(Y) &= \sup_{\mu \in \mathrm{dom}(\beta)} \left(\int_{[0,1)} AV@R_\alpha(2Y)\ \mu(d\alpha) - 2\beta(\mu)\right) \\
&\leq \sup_{\mu \in \mathrm{dom}(\beta)} \left(\int_{[0,1)} AV@R_\alpha(2Y)\ \mu(d\alpha) - \beta(\mu)\right) + \rho(0) = \rho(2Y) + \rho(0).
\end{aligned}
$$

So we end up with $b \leq 2\rho(2Y) - 3\rho(0)$ which completes the proof. □

As a consequence of the representation result Theorem A.1 we may verify that the following type of Fatou property.

Corollary A.3 *Let ρ fulfill the Lebesgue property, and let $\{X_k, X, Y \mid k \in \mathbb{N}\} \subset \mathcal{X}$ such that $\liminf_{k\to\infty} X_k \geq X$ $\mathbb{P}$-a.s. and $\sup_{k\in\mathbb{N}} |X_k| \leq Y$ $\mathbb{P}$-a.s.. Then*

$$\liminf_{k\to\infty} \rho(X_k) \geq \rho(X).$$

Proof By assumption $\overline{X}_k := \inf_{l\in\mathbb{N}, l\geq k} X_l$ and $\overline{X} := \sup_{k\in\mathbb{N}} \overline{X}_k$ are well-defined random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ satisfying $\overline{X}_k \leq Y$ $\mathbb{P}$-a.s. for $k \in \mathbb{N}$ and $\overline{X} \leq Y$ $\mathbb{P}$-a.s.. These random variables belong to $\mathcal{X}$ because $\mathcal{X}$ is solid. Furthermore we may find by Proposition A.2 a set $\mathcal{M}$ of Borel probability measures on $[0,1)$ and a mapping $\beta : \mathcal{M} \to (-\infty, \infty]$ such that

$$\rho(\overline{Y}) = \sup_{\mu\in\mathcal{M}} \left(\int_{[0,1)} AV@R_\alpha(\overline{Y})\ \mu(d\alpha) - \beta(\mu)\right), \tag{A.5}$$

for $\overline{Y} \in \{X_k, \overline{X}, X, \overline{X}, Y \mid k \in \mathbb{N}\}$, where $AV@R_0$ stands for the expectation. By the dominated convergence theorem the sequence $(\overline{X}_k)_{k\in\mathbb{N}}$ converges to $\overline{X}$ w.r.t. the L^1-norm. Hence, by L^1-norm continuity of the expectation and the Average Value at Risk we obtain $AV@R_\alpha(\overline{X}_k) \to AV@R_\alpha(\overline{X})$. Moreover, $X_k \geq \overline{X}_k$ $\mathbb{P}$-a.s. for $k \in \mathbb{N}$, and $\overline{X} \geq X$ $\mathbb{P}$-a.s.. Then, by monotonicity of the expectation and the Average Value at Risk

$$\liminf_{k\to\infty} AV@R_\alpha(X_k) \geq \liminf_{k\to\infty} AV@R_\alpha(\overline{X}_k) \geq AV@R_\alpha(\overline{X}) \geq AV@R_\alpha(X),$$

$\alpha \in [0,1)$. Finally, observe $AV@R_\alpha(X_k) \geq AV@R_\alpha(-Y)$ for $\alpha \in [0,1)$ due to the monotonicity of expectation and Average Value at Risk again. Then the application of the Fatou lemma yields

$$\liminf_{k\to\infty} \int_{[0,1)} AV@R_\alpha(X_k)\ \mu(d\alpha) \geq \int_{[0,1)} AV@R_\alpha(X)\ \mu(d\alpha).$$

Now, the statement of Corollary A.3 may be derived easily from the representation A.5. □

B. A concentration inequality

Let $(Z_j)_{j\in\mathbb{N}}$ be a sequence of independent, identically distributed d-dimensional random vectors on $(\Omega, \mathcal{F}, \mathbb{P})$. The main tool for the proof of Theorem 5.1 is the following concentration inequality.

Theorem B.1 *Let $\mathbb{F}$ be some at most countable set of real-valued Borel measurable mappings on $\mathbb{R}^d$ that are uniformly bounded by some positive constant b. Then for every $n \in \mathbb{N}$ and any $\varepsilon > 0$*

$$\mathbb{P}\big(\{S_n \geq \mathbb{E}[S_n] + \varepsilon\}\big) \leq \exp\left(\frac{-\varepsilon^2}{2nb^2}\right),$$

where $S_n := \sup_{f\in\mathbb{F}} \big|\sum_{j=1}^n f(Z_j) - n\ \mathbb{E}\big[f(Z_1)\big]\big|$.

Proof Since the mappings from $\mathbb{F}$ are uniformly bounded, and since $\mathbb{F}$ is at most countable, the function

$$h : \mathbb{R}^{dn} \to \mathbb{R},\ (z_1, \ldots, z_n) \mapsto \sup_{f\in\mathbb{F}} \left|\sum_{j=1}^n f(z_j) - n\ \mathbb{E}\big[f(Z_1)\big]\right|,$$

is Borel measurable and bounded. It furthermore satisfies for $z_1, \ldots, z_d, z \in \mathbb{R}^d$ and any $i \in \{1, \ldots, n\}$ the inequality

$$|h(z_1, \ldots, z_n) - h(z_1^i, \ldots, z_n^i)| \leq \sup_{f\in\mathbb{F}} \big|f(z_i) - f(z_i^i)\big| \leq 2b,$$

where $z_i^i := z$ and $z_j^i := z_j$ if $j \neq i$. Hence, we may conclude immediately the statement from McDiarmid's bounded differences inequality (see e.g. [8, Theorem 3.3.14]). □

References

[1] Bartl, D. and Tangpi, L. (2022). *Non-asymptotic rates for the estimation of risk measures*. Math. Oper. Res., **48**(4), 1811-2382.

[2] Belomestny, D. and Krätschmer, V. (2012). *Central limit theorems for law-invariant coherent risk measures*. Journal of Applied Probability, **49**, 1–21.

[3] Bellini, F., Klar, B., Müller, A., & Rosazza Gianin, E. (2014). *Generalized quantiles as risk measures*. Insurance: Mathematics and Economics, **54**, 41–48.

[4] Chu, J., & Tangpi, L. (2024). *Non-asymptotic estimation of risk measures using stochastic gradient Langevin dynamics*. SIAM J. Finan. Math. **15**(2), 503–536.

[5] Denneberg, D. (1994). *Non-additive measure and integral*. Kluwer, Dordrecht.

[6] Delbaen, F. (2009). Risk measures for non-integrable random variables, *Mathematical Finance* **19**, 329–333.

[7] Föllmer, H. and A. Schied (2011). *Stochastic Finance*. de Gruyter, Berlin, New York (3rd ed.).

[8] Gine, E. and Nickl, R. (2016). *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge University Press, Cambridge.

[9] Guigues, V., Juditsky, A. and Nemirovski, A. (2017). *Non-asymptotic confidence bounds for optimal value of a stochastic program*. Optimization Methods and Software **32**, 1033–1058.

[10] Guigues, V., Krätschmer, V. and Shapiro, A. (2018). *A central limit theorem and hypotheses testing for risk-averse stochastic programs*. SIAM J. OPTIM. **28**, 1337–1366.

[11] Jouini, E., Schachermayer, W. and Touzi, N. (2006). *Law invariant risk measures have the Fatou property*. Advances in Mathematical Economics **9,** 49–71.

[12] Kaina, M. and Rüschendorf, L. (2009). *On convex risk measures on L^p-spaces*. Math. Methods Oper. Res. **69**, 475 – 495.

[13] Krätschmer, V. (2024). *First order asymptotics of the sample average approximation method to solve risk averse stochastic programs*. Mathematical Programming. **208**, 209–242.

[14] Krätschmer, V. (2024). *Nonasymptotic upper estimates for errors of the sample average approximation method to solve risk-averse stochastic programs*. SIAM J. Opt. **34**, 1264–1294.

[15] Kim, S., Pasupathy, R., and Henderson, S. G. (2015). *A Guide to Sample-Average Approximation*, *Handbook of Simulation Optimization* (Chapter 8, pp. 207–243). Springer.

[16] Kusuoka, S. (2001). *On law invariant coherent risk measures.* Adv. Math. Econ **3**, 83–95.

[17] Levin, V. L. (1994). A characterization theorem for normal integrands with applications to descriptive function theory, functional analysis and nonconvex optimization. *Set-Valued Analysis.* **2**, 395–414.

[18] Pflug, G. Ch. and Wozabal, D. (2010). *Asymptotic distribution of law-invariant risk functionals.* Finance and Stochastics **14**, 397–418.

[19] Prelec, D. (1998). *The probability weighting function.* Econometrica, **66**(3), 497–527.

[20] Ruszczynski, A. and Shapiro, A. (2006). *Optimization of convex risk functions.* Mathematics of Operations Research **31**, 433–451.

[21] Shapiro, A. (2013). *Consistency of sample estimates of risk avers stochastic program.* Journal of Applied Probability **50**, 533–541.

[22] Shapiro, A. (1993). *Asymptotic behavior of optimal solutions in stochastic programming.* Mathematics of Operations Research, **18**, 829–845.

[23] Shapiro, A., Dentcheva, D. and Ruszczynski, A. (2014). *Lectures on stochastic programming.* MOS-SIAM Ser. Optim., Philadelphia (2nd ed.).

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

[25] van der Vaart, A.W. (1998). *Asymptotic statistics.* Cambridge University Press, Cambridge.

[26] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak convergence and empirical processes.* Springer, New York.

[27] Newey, W. and Powell, J. (1987). *Asymmetric least squares estimation and testing.* Econometrica **55**, 819–847.

[28] Wakker, P. P. (2010). *Prospect Theory for Risk and Rationality.* Cambridge University Press.