DISTRIBUTIONALLY ROBUST STOCHASTIC PROGRAMMING

ALEXANDER SHAPIRO*

School of Industrial and Systems Engineering,
Georgia Institute of Technology,
Atlanta, Georgia 30332-0205, USA
e-mail: ashapiro@isye.gatech.edu

Abstract. In this paper we study distributionally robust stochastic programming in a setting where there is a specified reference probability measure and the uncertainty set of probability measures consists of measures in some sense close to the reference measure. We discuss law invariance of the associated worst case functional and consider two basic constructions of such uncertainty sets. Finally we illustrate some implications of the property of law invariance.

Key words. Coherent risk measures, law invariance, Wasserstein distance, ϕ-divergence, sample average approximation, ambiguous chance constraints

AMS subject classifications. 90C15, 90C47, 91B30

1. Introduction.

Consider the following minimax stochastic optimization problem

$$\operatorname*{Min}_{x\in\mathcal{X}} \sup_{Q\in\mathfrak{M}} \mathbb{E}_Q[G(x,\xi(\omega))], \tag{1.1}$$

where $\mathcal{X} \subset \mathbb{R}^n$, $\xi : \Omega \to \Xi$ is a measurable mapping from Ω into $\Xi \subset \mathbb{R}^d$, $G : \mathbb{R}^n \times \Xi \to \mathbb{R}$ and $\mathfrak{M}$ is a nonempty set of probability measures (distributions), referred to as the uncertainty set, defined on a sample space $(\Omega, \mathcal{F})$. Such "worst case" (minimax) approach to stochastic optimization has a long history. It originated in John von Neumann's game theory and was applied in decision theory, game theory and statistics. In stochastic programming it goes back at least to Žáčková [25]. Recently the worst case approach attracted considerable attention and became known as distributionally robust stochastic optimization (DRSO).

Wide range of the uncertainty sets was suggested and analysed by various authors. If the uncertainty set consists of all probability distributions on Ξ, then the DRSO is reduced to a so-called robust optimization with respect to the worst realization of $\xi \in \Xi$ (we can refer to Ben-Tal, El Ghaoui and Nemirovski [5] for a thorough discussion of robust optimization). There are two natural, and somewhat different, approaches to constructing the uncertainty set of probability measures. One approach is to define $\mathfrak{M}$ by moment constraints. This is going back to a pioneering paper by Scarf [21] where it was applied to inventory modeling. In some, rather specific cases, this leads to computationally tractable DRSO problems (cf. [7],[11]).

Another approach is to assume that there is a reference probability measure P on $(\Omega, \mathcal{F})$ and the set $\mathfrak{M}$ consists of probability measures Q on $(\Omega, \mathcal{F})$ in some sense close to P. Of course this leaves a wide range of possible choices for quantifying the concept of closeness between probability measures. It also raises questions of practical relevance and computational tractability of obtained formulations. In that respect we can mention recent paper by Esfahani and Kuhn [12] where it is shown that, under mild assumptions, the DRSO problems over Wasserstein balls can be reformulated as finite convex programs – in some cases even as tractable linear programs.

* This research was partly supported by NSF grant 1633196 and DARPA EQUiPS program, grant SNL 014150709.

In this paper we deal with the second approach assuming existence of a specified reference probability measure P. With the set $\mathfrak{M}$ is associated the functional

$$\rho(Z) := \sup_{Q\in\mathfrak{M}} \mathbb{E}_Q[Z] = \sup_{Q\in\mathfrak{M}} \int_\Omega Z(\omega)dQ(\omega), \tag{1.2}$$

defined on an appropriate space $\mathcal{Z}$ of measurable functions $Z:\Omega\to\mathbb{R}$. We assume further that the probability measures Q are absolutely continuous with respect to P. By the Radon - Nikodym theorem, probability measure Q is absolutely continuous with respect to P iff $dQ=\zeta dP$ for some probability density function (pdf) $\zeta:\Omega\to\mathbb{R}_+$. That is with the set $\mathfrak{M}$ is associated set of probability density functions

$$\mathfrak{A} := \{\zeta = dQ/dP : Q\in\mathfrak{M}\}. \tag{1.3}$$

We work with the space $\mathcal{Z}:=L_p(\Omega,\mathcal{F},P)$, $p\in[1,\infty)$, of random variables $Z:\Omega\to\mathbb{R}$ having finite p-th order moments, and its dual space $\mathcal{Z}^*=L_q(\Omega,\mathcal{F},P)$, $q\in(1,\infty]$, $1/p+1/q=1$. For $Z\in\mathcal{Z}$ and $\zeta\in\mathcal{Z}^*$ their scalar product is defined as

$$\langle\zeta,Z\rangle := \int_\Omega \zeta Z dP.$$

For $p\in(1,\infty)$ both spaces $\mathcal{Z}$ and $\mathcal{Z}^*$ are reflexive, and the weak* topology of $\mathcal{Z}^*$ coincides with its weak topology. We also consider space $\mathcal{Z}=L_\infty(\Omega,\mathcal{F},P)$ and pair it with the space $L_1(\Omega,\mathcal{F},P)$ by equipping $L_1(\Omega,\mathcal{F},P)$ with its weak topology and $L_\infty(\Omega,\mathcal{F},P)$ with the weak* topology. We assume that $Z_x(\omega):=G(x,\xi(\omega))$ belongs to the space $\mathcal{Z}$ for all $x\in\mathcal{X}$.

Suppose that $\mathfrak{A}$ is a subset of the dual (paired) space $\mathcal{Z}^*$. Then the corresponding functional ρ can be written as

$$\rho(Z) = \sup_{\zeta\in\mathfrak{A}}\langle\zeta,Z\rangle. \tag{1.4}$$

This is the dual form of so-called coherent risk measures (Artzner et al [3]). We will refer to the set $\mathfrak{A}$ as the *uncertainty set* associated with ρ, and use notation $\rho=\rho_{\mathfrak{A}}$ for the corresponding functional. In the terminology of convex analysis, $\rho_{\mathfrak{A}}(\cdot)$ is the support function of the set $\mathfrak{A}$. If the set $\mathfrak{A}\subset\mathcal{Z}^*$ is bounded (in the norm topology of $\mathcal{Z}^*$), then $\rho_{\mathfrak{A}}:\mathcal{Z}\to\mathbb{R}$ is finite valued.

This paper is organized as follows. In the next section we discuss the basic concept of law invariance of risk functional ρ and its relation to the corresponding uncertainty set $\mathfrak{A}$. Section 3 is devoted to study of two generic approaches to construction of the uncertainty sets. In section 4 we consider applications of the law invariance to the Sample Average Approximation method and Chance Constrained problems.

We will use the following notation throughout the paper. By saying that Z is a random variable we mean that $Z:\Omega\to\mathbb{R}$ is a measurable function. For a random variable Z we denote by $F_Z(z):=P(Z\le z)$ its cumulative distribution function (cdf), and by $F_Z^{-1}(\tau):=\inf\{z: F_Z(z)\ge\tau\}$ the corresponding left-site τ-quantile. The notation $\zeta\succeq 0$ means that $\zeta(\omega)\ge 0$ for a.e. $\omega\in\Omega$. By $\mathfrak{D}$ we denote the set of probability density functions, i.e., a measurable $\zeta:\Omega\to\mathbb{R}_+$ belongs to $\mathfrak{D}$ if $\int_\Omega\zeta dP=1$. Note that $\mathfrak{D}\subset L_1(\Omega,\mathcal{F},P)$. We also use

$$\mathfrak{D}^* := \mathcal{Z}^*\cap\mathfrak{D} \tag{1.5}$$

to denote the set of probability density functions in the dual space $\mathcal{Z}^*$. By $\mathbb{I}_A(\cdot)$ we denote the indicator function of set A, that is $\mathbb{I}_A(x)=0$ if $x\in A$ and $\mathbb{I}_A(x)=+\infty$

otherwise. We also use characteristic function $\mathbf{1}_A(\cdot)$, defined as $\mathbf{1}_A(x) = 1$ if $x \in A$ and $\mathbf{1}_A(x) = 0$ otherwise. By $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$ we denote the extended real line. The functional $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$, defined in (1.4), can take value $+\infty$, but not $-\infty$ since the set $\mathfrak{A}$ is assumed to be nonempty.

2. Law invariance.

We say that two random variables $Z, Z' : \Omega \to \mathbb{R}$ are *distributionally equivalent*, denoted $Z \stackrel{\mathcal{D}}{\sim} Z'$, if they have the same distribution with respect to the reference probability measure P, i.e., $P(Z \le z) = P(Z' \le z)$ for all $z \in \mathbb{R}$. In other words two random variables are distributionally equivalent if their cumulative distributions functions are equal to each other.

Definition 1. *It is said that a functional $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ is* law invariant *(with respect to the reference probability measure P) if for all $Z, Z' \in \mathcal{Z}$ the implication $Z \stackrel{\mathcal{D}}{\sim} Z' \Rightarrow \rho(Z) = \rho(Z')$ holds.*

We discuss now a relation between law invariance of the functional ρ, given in the form (1.4), and law invariance of the corresponding uncertainty set $\mathfrak{A}$ of density functions. Note that the uncertainty set $\mathfrak{A}$ is not defined uniquely by the relation (1.4). That is, the maximum in the right hand side of (1.4) is not changed if the set $\mathfrak{A}$ is replaced by the weak* topological closure of the convex hull of $\mathfrak{A}$. Therefore it is natural to assume that the uncertainty set $\mathfrak{A}$ is convex and closed in the weak* topology of the space $\mathcal{Z}^*$.

Definition 2. *We say that the uncertainty set $\mathfrak{A}$ is law invariant if $\zeta \in \mathfrak{A}$ and $\zeta' \stackrel{\mathcal{D}}{\sim} \zeta$ imply that $\zeta' \in \mathfrak{A}$.*

The relation "$\stackrel{\mathcal{D}}{\sim}$" defines an equivalence relation on the set of random variables. That is, for any random variables $X, Y, Z : \Omega \to \mathbb{R}$ we have that: (i) $X \stackrel{\mathcal{D}}{\sim} X$, (ii) if $X \stackrel{\mathcal{D}}{\sim} Y$ then $Y \stackrel{\mathcal{D}}{\sim} X$, (iii) if $X \stackrel{\mathcal{D}}{\sim} Y$ and $Y \stackrel{\mathcal{D}}{\sim} Z$, then $X \stackrel{\mathcal{D}}{\sim} Z$. It follows that the set of random variables is the union of disjoint classes of distributionally equivalent random variables. We denote

$$\mathcal{O}(Z) := \{Y : Y \stackrel{\mathcal{D}}{\sim} Z\}$$

the corresponding class of distributionally equivalent random variables, referred to as the *orbit* of random variable Z. The set $\mathfrak{A}$ is law invariant iff the following implication holds: $\zeta \in \mathfrak{A} \Rightarrow \mathcal{O}(\zeta) \subset \mathfrak{A}$. Consequently if the set $\mathfrak{A}$ is law invariant, then it can be represented as the union of disjoint classes $\mathcal{O}(\zeta)$, $\zeta \in \mathfrak{A}$.

Following is the main result of this section.

Theorem 3. (i) *If the uncertainty set $\mathfrak{A}$ is law invariant, then the corresponding functional $\rho = \rho_{\mathfrak{A}}$ is law invariant.* (ii) *Conversely, if the functional $\rho = \rho_{\mathfrak{A}}$ is law invariant and the set $\mathfrak{A}$ is convex and weakly* closed, then $\mathfrak{A}$ is law invariant.*

We give a proof of this theorem in several steps. For $\zeta \in \mathcal{Z}^*$ consider the following functional

$$\varrho_\zeta(Z) := \sup_{\eta \in \mathcal{O}(\zeta)} \langle \eta, Z \rangle, \quad Z \in \mathcal{Z}. \tag{2.1}$$

That is, $\varrho_\zeta = \rho_{\mathfrak{A}}$ for $\zeta \in \mathfrak{D}^*$ and $\mathfrak{A} = \mathcal{O}(\zeta)$. Note that there is a certain symmetry between the paired spaces $\mathcal{Z}$ and $\mathcal{Z}^*$. Therefore with some abuse of the notation for $Z \in \mathcal{Z}$ we also consider the functional

$$\varrho_Z(\zeta) := \sup_{Y \in \mathcal{O}(Z)} \langle \zeta, Y \rangle, \quad \zeta \in \mathcal{Z}^*. \tag{2.2}$$

Consider the following conditions.

(A(i)) For every $\zeta \in \mathcal{Z}^*$ the functional $\varrho_\zeta : \mathcal{Z} \to \overline{\mathbb{R}}$ is law invariant.

(A(ii)) For every $Z \in \mathcal{Z}$ the functional $\varrho_Z : \mathcal{Z}^* \to \overline{\mathbb{R}}$ is law invariant.

In Lemma 5 below we show that these conditions always hold.

Lemma 4. (i) *If the uncertainty set* $\mathfrak{A}$ *is law invariant and condition* (A(i)) *holds, then the corresponding functional* $\rho = \rho_{\mathfrak{A}}$ *is law invariant.* (ii) *Conversely, if the functional* $\rho = \rho_{\mathfrak{A}}$ *is law invariant, the set* $\mathfrak{A}$ *is convex and weakly** *closed and condition* (A(ii)) *holds, then* $\mathfrak{A}$ *is law invariant.*

Proof. (i) We have that

$$\rho(Z) = \sup_{\zeta\in\mathfrak{A}} \langle \zeta, Z\rangle = \sup_{\zeta\in\mathfrak{A},\, \eta\in\mathcal{O}(\zeta)} \langle \eta, Z\rangle = \sup_{\zeta\in\mathfrak{A}} \varrho_\zeta(Z),$$

where the second equality follows by the law invariance of $\mathfrak{A}$ and the last equality follows from the definition of ϱ_ζ. Hence law invariance of ρ follows from law invariance of each ϱ_ζ. This completes the proof of (i).

(ii) Consider the conjugate of ρ:

$$\rho^*(\zeta) := \sup_{Z\in\mathcal{Z}} \langle \zeta, Z\rangle - \rho(Z).$$

Let us observe that $\rho^*(\zeta)$ is law invariant. Indeed since ρ is law invariant, for $Y \stackrel{\mathcal{D}}{\sim} Z$ we have that $\rho(Y) = \rho(Z)$, and hence

(2.3)
$$\rho^*(\zeta) = \sup_{Z\in\mathcal{Z},\, Y\in\mathcal{O}(Z)} \langle \zeta, Y\rangle - \rho(Y) = \sup_{Z\in\mathcal{Z},\, Y\in\mathcal{O}(Z)} \langle \zeta, Y\rangle - \rho(Z) = \sup_{Z\in\mathcal{Z}} \varrho_Z(\zeta) - \rho(Z).$$

If $\zeta' \in \mathcal{Z}^*$ is distributionally equivalent to ζ, then by assumption (A(ii)) we have that $\varrho_Z(\zeta') = \varrho_Z(\zeta)$, and hence it follows that $\rho^*(\zeta') = \rho^*(\zeta)$.

Furthermore we have that the conjugate of ρ is the indicator function $\mathbb{I}_{\mathfrak{A}}(\zeta)$ (e.g. [9, Example 2.115]). It is straightforward to see that $\mathbb{I}_{\mathfrak{A}}$ is law invariant iff the set $\mathfrak{A}$ is law invariant. This completes the proof of (ii). ∎

We show now that conditions (A(i)) and (A(ii)) always hold. Together with Lemma 4 this will complete the proof of Theorem 3. It is said that the probability measure P is *nonatomic* if for any measurable set $A \in \mathcal{F}$ with $P(A) > 0$ there exists a measurable set $B \subset A$ such that $P(A) > P(B) > 0$. If P is nonatomic, then the space $(\Omega, \mathcal{F}, P)$ is also called nonatomic.

Lemma 5. *Conditions* (A(i)) *and* (A(ii)) *hold for any probability space.*

Proof. If the measure P is nonatomic, then (cf. [14, Lemma 4.55])

$$\sup_{\eta\in\mathcal{O}(\zeta)} \int_\Omega Z\eta\, dP = \int_0^1 F_Z^{-1}(t) F_\zeta^{-1} dt. \tag{2.4}$$

Since $Z \stackrel{\mathcal{D}}{\sim} Z'$ means that $F_Z = F_{Z'}$, it follows that for any nonatomic probability measure, condition (A(i)) is satisfied, and by the same argument condition (A(ii)) holds as well.

When the reference space has atoms we use the following construction. Consider a nonatomic probability space $(\Xi, \mathcal{G}, Q)$. For example we can use $\Xi = [0, 1]$ equipped with its Borel sigma algebra and uniform probability measure. Let $(\hat\Omega, \hat{\mathcal{F}}, \hat P)$ be the corresponding product space, i.e., $\hat P := Q \times P$ is the product measure on $\hat{\mathcal{F}} := \mathcal{G} \times \mathcal{F}$.

Since Q is nonatomic, the product space is also nonatomic (e.g. [8]). In the product space consider sigma algebra $\mathcal{F}'$ of sets of the form $\Xi \times A$, $A \in \mathcal{F}$. This sigma algebra is a subalgebra of $\hat{\mathcal{F}} = \mathcal{G} \times \mathcal{F}$. With marginal measure $P'(\Xi \times A) = P(A)$ on this subalgebra, we obtain that the reference probability space is isomorphic to $(\hat{\Omega}, \mathcal{F}', P')$. We then identify $(\Omega, \mathcal{F}, P)$ with $(\hat{\Omega}, \mathcal{F}', P')$, and with some abuse of the notation write $(\hat{\Omega}, \mathcal{F}, P)$ for the embedded space.

For $\mathcal{F}$-measurable $\zeta : \hat{\Omega} \to \mathbb{R}$, consider orbit $\mathcal{O}(\zeta)$ consisting of $\mathcal{F}$-measurable $\zeta' : \hat{\Omega} \to \mathbb{R}$ distributionally equivalent to ζ. We also consider the orbit $\hat{\mathcal{O}}(\zeta)$ consisting of $\hat{\mathcal{F}}$-measurable $\zeta' : \hat{\Omega} \to \mathbb{R}$ distributionally equivalent to ζ. That is, $\mathcal{O}(\zeta)$ is the orbit of ζ in the reference space and $\hat{\mathcal{O}}(\zeta)$ is the orbit of ζ in the respective nonatomic space. Note that $\mathcal{O}(\zeta)$ is a subset of $\hat{\mathcal{O}}(\zeta)$.

For $\mathcal{F}$-measurable $Z \in \mathcal{Z}$ and $\hat{\mathcal{F}}$-measurable $\zeta' \in \mathcal{Z}^*$ we have that

$$\int_{\hat{\Omega}} Z\zeta' dP = \mathbb{E}[Z\zeta'] = \mathbb{E}\left[\mathbb{E}_{|\mathcal{F}}[Z\zeta']\right] = \mathbb{E}\left[Z\, \mathbb{E}_{|\mathcal{F}}[\zeta']\right], \tag{2.5}$$

where $\mathbb{E}_{|\mathcal{F}}$ denotes the conditional expectation and the last equality holds since Z is $\mathcal{F}$-measurable. That is

$$\int_{\hat{\Omega}} Z\zeta'\, d\hat{P} = \int_{\hat{\Omega}} Z\eta\, dP, \tag{2.6}$$

where $\eta := \mathbb{E}_{|\mathcal{F}}[\zeta']$ is $\mathcal{F}$-measurable. It follows

$$\sup_{\zeta' \in \hat{\mathcal{O}}(\zeta)} \int_{\hat{\Omega}} Z\zeta' d\hat{P} = \sup_{\eta \in \mathcal{O}(\zeta)} \int_{\hat{\Omega}} Z\eta\, dP. \tag{2.7}$$

Since $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{P})$ is nonatomic, we have that if $Z' : \hat{\Omega} \to \mathbb{R}$ is distributionally equivalent to Z, then

$$\sup_{\zeta' \in \hat{\mathcal{O}}(\zeta)} \int_{\hat{\Omega}} Z\zeta' d\hat{P} = \sup_{\zeta' \in \hat{\mathcal{O}}(\zeta)} \int_{\hat{\Omega}} Z'\zeta' d\hat{P}. \tag{2.8}$$

It follows from (2.7) and (2.8) that condition (A(i)) holds for the reference space $(\hat{\Omega}, \mathcal{F}, P)$. Condition (A(ii)) can be shown in a similar way. ∎

Theorem 3 now follows from Lemmas 4 and 5.

REMARK 2.1. Recall that we assume that the uncertainty set $\mathfrak{M}$ consists of probability measures *absolutely continuous* with respect to P. For example let the probability measure P be discrete, i.e., there is a countable set $\Omega' \subset \Omega$ such that $P(\Omega') = 1$ and $P(\{\omega\}) > 0$ for every $\omega \in \Omega'$. Then Q is absolutely continuous with respect to P iff Q is supported on Ω', i.e., $Q(\Omega') = 1$.

REMARK 2.2. If the space $(\Omega, \mathcal{F}, P)$ is nonatomic and the set $\mathfrak{A}$ is law invariant, then the functional $\rho = \rho_{\mathfrak{A}}$ is law invariant and hence $\rho(Z) \geq \mathbb{E}_P(Z)$ for all $Z \in \mathcal{Z}$ (e.g. [24, Corollary 6.52]). It follows that if moreover the set $\mathfrak{A}$ is convex and weakly* closed, then $\mathbf{1}_\Omega \in \mathfrak{A}$. Without the assumption that the space $(\Omega, \mathcal{F}, P)$ is nonatomic, this may not hold. For example suppose that the reference probability measure P is discrete with $\Omega = \{\omega_1, ...\}$ and respective probabilities $p_i > 0$. Suppose further that $\sum_{i \in \mathcal{I}} p_i = \sum_{i \in \mathcal{I}'} p_i$ iff the index sets $\mathcal{I}, \mathcal{I}' \subset \mathbb{N}$ are equal to each other. In such case we say that probabilities p_i are *essentially different* from each other. In case of the discrete probability space two random variables $Z, Z' : \Omega \to \mathbb{R}$ are distributionally equivalent iff $P(Z = a) = P(Z' = a)$ for any $a \in \mathbb{R}$. If the set $\{\omega \in \Omega : Z(\omega) = a\}$ is empty, then $P(Z = a) = 0$. So suppose that sets $\mathcal{I} := \{i \in \mathbb{N} : Z(\omega_i) = a\}$ and $\mathcal{I}' := \{i \in \mathbb{N} : Z'(\omega_i) = a\}$ are nonempty. Then $P(Z = a) = P(Z' = a)$ iff $\sum_{i\in\mathcal{I}} p_i = \sum_{i\in\mathcal{I}'} p_i$. By the above condition this happens iff $\mathcal{I} = \mathcal{I}'$. That is, here Z and Z' are distributionally equivalent iff their level sets do coincide. This means that Z and Z' are distributionally equivalent iff $Z = Z'$. In that case any set $\mathfrak{A}$ and functional ρ are law invariant. Of course for an arbitrary convex closed set $\mathfrak{A}$ (of densities) there is no guarantee that $\mathbf{1}_\Omega \in \mathfrak{A}$. In particular, the set $\mathfrak{A}$ can be a singleton.

3. Construction of the uncertainty sets of probability measures. In this section we discuss some generic approaches to construction of the sets $\mathfrak{M}$ of probability measures used in (1.2), and consider examples. We assume existence of a reference probability measure P on $(\Omega, \mathcal{F})$ and consider probability measures Q in some sense close to P.

3.1. Distance approach. Consider the following construction. Let $\mathfrak{H}$ be a nonempty set of measurable functions $h : \Omega \to \mathbb{R}$. For a probability measure Q on $(\Omega, \mathcal{F})$ consider

$$d(Q, P) := \sup_{h\in\mathfrak{H}} \int_\Omega h dQ - \int_\Omega h dP. \tag{3.1}$$

Of course the integrals and the difference in the right hand sides of (3.1) should be well defined. If the set $\mathfrak{H}$ is *symmetric*, i.e., $h \in \mathfrak{H}$ implies that $-h \in \mathfrak{H}$, then it follows that

$$d(Q, P) = \sup_{h\in\mathfrak{H}} \left| \int_\Omega h dQ - \int_\Omega h dP \right|. \tag{3.2}$$

Formula (3.2) defines a semi-distance between probability measures Q and P (it could happen that the right hand side of (3.2) is zero even if $Q \neq P$), while $d(Q, P)$ defined in (3.1) could be not symmetric.

Assume further that $\mathfrak{H} \subset \mathcal{Z}$ and Q is absolutely continuous with respect to P, with the corresponding density $\zeta = dQ/dP \in \mathcal{Z}^*$. Then

$$d(Q, P) = \sup_{h\in\mathfrak{H}} \int_\Omega h dQ - \int_\Omega h dP = \sup_{h\in\mathfrak{H}} \int_\Omega h(\zeta - 1) dP = \sup_{h\in\mathfrak{H}} \langle h, \zeta - 1\rangle. \tag{3.3}$$

Since $\mathfrak{H} \subset \mathcal{Z}$ and $\zeta \in \mathcal{Z}^*$ it follows that the scalar product $\langle h, \zeta - 1\rangle$ is well defined and finite valued for every $h \in \mathfrak{H}$. Moreover if the set $\mathfrak{H} \subset \mathcal{Z}$ is bounded, then $d(Q, P)$ is finite valued. With the set $\mathfrak{H} \subset \mathcal{Z}$ and $\varepsilon > 0$ we associate the following set of density functions[1] in the dual $\mathcal{Z}^*$ of the space $\mathcal{Z}$,

$$\mathcal{A}_\varepsilon(\mathfrak{H}) := \{\zeta \in \mathfrak{D}^* : d(Q, P) \le \varepsilon\} = \{\zeta \in \mathfrak{D}^* : \langle h, \zeta - 1\rangle \le \varepsilon, \ \forall h \in \mathfrak{H}\}. \tag{3.4}$$

For $\varepsilon = 1$ we drop the subscript ε and simply write $\mathcal{A}(\mathfrak{H})$. Note that

$$\mathcal{A}_\varepsilon(\mathfrak{H}) = \mathcal{A}(\varepsilon^{-1}\mathfrak{H}), \tag{3.5}$$

and that $\mathbf{1}_\Omega \in \mathcal{A}_\varepsilon(\mathfrak{H})$, where $\mathbf{1}_\Omega(\omega) = 1$ for all $\omega \in \Omega$.

[1] Recall that $\mathfrak{D}^* = \mathcal{Z}^* \cap \mathfrak{D}$ is the set of probability density functions in the dual space $\mathcal{Z}^*$.

Definition 6. *Polar (one-sided) of a nonempty set $\mathcal{S} \subset \mathcal{Z}$ is the set*

$$\mathcal{S}^\circ := \{\zeta \in \mathcal{Z}^* : \langle \zeta, Z \rangle \le 1,\ \forall Z \in \mathcal{S}\}.$$

Similarly (one-sided) polar of a set $\mathcal{C} \subset \mathcal{Z}^*$ is

$$\mathcal{C}^\circ := \{Z \in \mathcal{Z} : \langle \zeta, Z \rangle \le 1,\ \forall \zeta \in \mathcal{C}\}.$$

Note that the set $\mathcal{S}^\circ \subset \mathcal{Z}^*$ is convex weakly* closed, and the set $\mathcal{C}^\circ \subset \mathcal{Z}$ is convex weakly closed.

We have the following duality result (e.g. [2, Theorem 5.103]).

Theorem 7. *If $\mathcal{C}$ is a convex weakly* closed subset of $\mathcal{Z}^*$ and $0 \in \mathcal{C}$, then it follows that $(\mathcal{C}^\circ)^\circ = \mathcal{C}$.*

This has the following implications for our analysis. Consider a convex weakly* closed set $\mathfrak{A} \subset \mathfrak{D}^*$ of probability densities and define

$$\mathfrak{H} := \{h \in \mathcal{Z} : \langle h, \zeta - 1 \rangle \le 1,\ \forall \zeta \in \mathfrak{A}\}. \tag{3.6}$$

That is, $\mathfrak{H} = (\mathfrak{A} - \mathbf{1}_\Omega)^\circ$ is the (one-sided) polar of the set $\mathfrak{A} - \mathbf{1}_\Omega$. Suppose that $\mathbf{1}_\Omega \in \mathfrak{A}$. Then by Theorem 7 we have that $\mathfrak{A} - \mathbf{1}_\Omega$ is (one-sided) polar of the set $\mathfrak{H}$, i.e.,

$$\mathfrak{A} = \left\{ \zeta \in \mathcal{Z}^* : \sup_{h \in \mathfrak{H}} \langle h, \zeta - 1 \rangle \le 1 \right\}. \tag{3.7}$$

We obtain the following result.

Proposition 8. *For any convex weakly* closed set $\mathfrak{A} \subset \mathfrak{D}^*$ containing the constant density function $\mathbf{1}_\Omega$, there exists a convex weakly closed set $\mathfrak{H} \subset \mathcal{Z}$ such that $\mathfrak{A} = \mathcal{A}(\mathfrak{H})$.*

For a given uncertainty set $\mathfrak{A} \subset \mathfrak{D}^*$, the equation $\mathfrak{A} = \mathcal{A}(\mathfrak{H})$ does not define the (convex weakly closed) set $\mathfrak{H}$ uniquely. This is because of the additional constraint for the set $\mathfrak{A} \subset \mathcal{Z}^*$ to be a set of *probability densities*. In particular for any $h \in \mathcal{Z}$, $\lambda \in \mathbb{R}$ and $\zeta \in \mathfrak{D}^*$ we have that $\langle h + \lambda, \zeta - 1 \rangle = \langle h, \zeta - 1 \rangle$.

We discuss now law invariance of the set $\mathcal{A}(\mathfrak{H})$. A function $h \in \mathfrak{H}$ is assumed to be measurable and hence can be viewed as a random variable defined on the reference probability space $(\Omega, \mathcal{F}, P)$. Therefore we can apply derivations of Section 2.

Proposition 9. *Suppose that the set $\mathfrak{H} \subset \mathcal{Z}$ is law invariant. Then the set $\mathfrak{A} := \mathcal{A}_\varepsilon(\mathfrak{H})$ is law invariant. Conversely, if the set $\mathfrak{A} \subset \mathcal{Z}^*$ is law invariant, then the set $\mathfrak{H} := (\mathfrak{A} - \mathbf{1}_\Omega)^\circ$ is law invariant.*

Proof. Consider the functional

$$\psi(\zeta) := \sup_{h \in \mathfrak{H}} \langle h, \zeta - 1 \rangle,\ \zeta \in \mathcal{Z}^*.$$

Note that if $\zeta \in \mathcal{Z}^*$, then $\zeta - \mathbf{1}_\Omega \in \mathcal{Z}^*$, and hence the functional $\psi : \mathcal{Z}^* \to \overline{\mathbb{R}}$ is well defined. Note also $\zeta \overset{\mathcal{D}}{\sim} \zeta'$ iff $\zeta - \mathbf{1}_\Omega \overset{\mathcal{D}}{\sim} \zeta' - \mathbf{1}_\Omega$. Since $\mathfrak{H}$ is law invariant, it follows that ψ is law invariant. This can be shown in the same way as in the proof of Theorem 3. Since $\mathcal{A}_\varepsilon(\mathfrak{H}) = \{\zeta \in \mathfrak{D}^* : \psi(\zeta) \le \varepsilon\}$, it follows that $\mathcal{A}_\varepsilon(\mathfrak{H})$ is law invariant.

For the converse implication recall that $\mathfrak{H} = (\mathfrak{A} - \mathbf{1}_\Omega)^\circ$ can be defined as in (3.6). By law invariance of $\mathfrak{A} - \mathbf{1}_\Omega$, it follows that $\mathfrak{H}$ is law invariant. ∎

Example 3.1 (Expectation). Let $\mathfrak{H} := L_1(\Omega, \mathcal{F}, P)$. Then $d(Q, P) = +\infty$ for any $Q \neq P$, and hence $\mathcal{A}_\varepsilon(\mathfrak{H}) = \{\mathbf{1}_\Omega\}$ and the corresponding functional $\rho(Z) = \mathbb{E}_P[Z]$, $Z \in L_1(\Omega, \mathcal{F}, P)$. Of course, the sets $\mathfrak{H}$, $\mathcal{A}_\varepsilon(\mathfrak{H})$ and the corresponding functional $\rho(Z)$ are law invariant here.

Example 3.2 (Total Variation Distance). Consider the set

$$\mathfrak{H} := \{h : |h(\omega)| \le 1,\ \omega \in \Omega\}. \tag{3.8}$$

The set $\mathfrak{H} \subset L_\infty(\Omega, \mathcal{F}, P)$ is symmetric and is law invariant. The total variation norm of a finite signed measure μ on $(\Omega, \mathcal{F})$ is defined as

$$\|\mu\|_{TV} := \sup_{A \in \mathcal{F}} \mu(A) - \inf_{B \in \mathcal{F}} \mu(B). \tag{3.9}$$

In this example $d(Q, P) = \|Q - P\|_{TV}$ (e.g. [19, p. 44]). If we assume further that measures Q are absolutely continuous with respect to P, then for $dQ = \zeta dP$ we have

$$d(Q, P) = \sup_{h \in \mathfrak{H}} \int_\Omega h(\zeta - 1) dP = \int_\Omega |\zeta - 1| dP = \|\zeta - 1\|_1.$$

The corresponding set

$$\mathcal{A}_\varepsilon(\mathfrak{H}) = \{\zeta \in \mathfrak{D} : \|\zeta - 1\|_1 \le \varepsilon\} \subset L_1(\Omega, \mathcal{F}, P)$$

is law invariant. Law invariance of $\mathcal{A}_\varepsilon(\mathfrak{H})$ can be verified directly by noting that if $\zeta, \zeta' \in L_1(\Omega, \mathcal{F}, P)$ and $\zeta \overset{\mathcal{D}}{\sim} \zeta'$, then $\|\zeta\|_1 = \|\zeta'\|_1$. The corresponding functional $\rho(Z)$ is defined (finite valued) on $L_\infty(\Omega, \mathcal{F}, P)$ and is law invariant (see Example 3.7 below).

Remark 3.1. Consider the set $\mathfrak{H}$ defined in (3.8) and the corresponding distance $d(Q, P)$. Without assuming that Q is absolutely continuous with respect to P, structure of the set of probability measures Q satisfying $d(Q, P) \le \varepsilon$ is more involved. By the Lebesgue Decomposition Theorem we have that any probability measure Q on $(\Omega, \mathcal{F})$ can be represented as a convex combination $Q = \gamma Q_1 + (1 - \gamma) Q_2$, $\gamma \in [0, 1]$, of absolutely continuous with respect to P probability measure Q_1 and probability measure Q_2 supported on a set $S \in \mathcal{F}$ of P-measure zero, i.e., $Q_2(S) = 1$ and $P(S) = 0$. By (3.9) we have that $d(Q_2, P) = 2$.

Example 3.3. Consider the set

$$\mathfrak{H} := \{h : h(\omega) \in [0, 1],\ \omega \in \Omega\},$$

and probability measures $dQ = \zeta dP$ absolutely continuous with respect to P. This set $\mathfrak{H}$ is law invariant, but is not symmetric, and

$$d(Q, P) = \int_\Omega [\zeta - 1]_+ dP.$$

The corresponding set $\mathfrak{A} = \mathcal{A}_\varepsilon(\mathfrak{H})$ and functional $\rho(Z)$ are law invariant (see Example 3.8 below).

Example 3.4 (Wasserstein distance). Let Ω be a closed subset of $\mathbb{R}^d$ equipped with its Borel sigma algebra. Consider the set of Lipschitz continuous functions modulus one,

$$\mathfrak{H} := \{h : h(\omega) - h(\omega') \le \|\omega - \omega'\|, \forall \omega, \omega' \in \Omega\}, \tag{3.10}$$

where $\|\cdot\|$ is the standard Euclidean norm on $\mathbb{R}^d$. The corresponding distance $d(Q,P)$ is called Wasserstein (also called Kantorovich) distance between probability measures Q and P (see, e.g., [15],[19] for a discussion of properties of this metric). It is not difficult to see that if $h \in \mathfrak{H}$ and $h' \overset{\mathcal{D}}{\sim} h$, then h' is not necessarily Lipschitz continuous modulus one. Hence the set $\mathfrak{H}$ is not necessarily law invariant.

Consider for example finite set $\Omega = \{\omega_1, ..., \omega_m\} \subset \mathbb{R}^d$ and the reference probability measure P assigns to each point $\omega_i \in \mathbb{R}^d$ equal probability $p_i = 1/m$, $i = 1, ..., m$. A function $h : \Omega \to \mathbb{R}$ can be identified with vector $(h(\omega_1), ..., h(\omega_m))$. Therefore we can view $\mathfrak{H}$ as a subset of $\mathbb{R}^m$, and thus

$$\mathfrak{H} = \{h \in \mathbb{R}^m : h_i - h_j \le \|\omega_i - \omega_j\|,\ i, j = 1, ..., m\}. \tag{3.11}$$

By adding the constraint $\sum_{i=1}^m h_i = 0$ to the right hand side of (3.11) we do not change the corresponding uncertainty set $\mathfrak{A} = \mathcal{A}(\mathfrak{H})$. With this additional constraint the set $\mathfrak{H} \subset \mathbb{R}^m$ becomes a bounded polytope. The uncertainty set $\mathfrak{A} = \mathcal{A}(\mathfrak{H})$ is also a bounded polytope in $\mathbb{R}^m$.

We have here that two variables $h, h' : \Omega \to \mathbb{R}$ are distributionally equivalent iff there exists a permutation $\pi : \Omega \to \Omega$ such that $h' = h \circ \pi$, where the notation $h \circ \pi$ stands for the composition $h(\pi(\cdot))$. For a permutation $\pi : \Omega \to \Omega$ and uncertainty set $\mathfrak{A} = \mathcal{A}(\mathfrak{H})$ we have that $\mathfrak{A} \circ \pi = \mathcal{A}(\mathfrak{H} \circ \pi^{-1})$, and

$$\mathfrak{H} \circ \pi^{-1} = \left\{h \in \mathbb{R}^m : h_i - h_j \le \|\omega_{\pi(i)} - \omega_{\pi(j)}\|,\ i, j = 1, ..., m\right\}.$$

Unless the respective distances $\|\omega_i - \omega_j\|$ are equal to each other, the set $\mathfrak{H} \circ \pi^{-1}$ is different from the set $\mathfrak{H}$ and the uncertainty set $\mathfrak{A}$ is not necessarily equal to the set $\mathfrak{A} \circ \pi$. That is, by changing order of the points $\omega_1, ..., \omega_m$ we may change the corresponding uncertainty set and the associated functional $\rho(Z) = \sup_{q\in\mathfrak{A}} \sum_{i=1}^m q_i Z(\omega_i)$. Of course, making such permutation does not change the corresponding expectation $\mathbb{E}_P[Z] = \frac{1}{m}\sum_{i=1}^m Z(\omega_i)$.

That is, for the uncertainty set defined by the Wasserstein distance we are not guaranteed that the uncertainty set $\mathfrak{A} = \mathcal{A}(\mathfrak{H})$ and the corresponding functional $\rho = \rho_{\mathfrak{A}}$ are law invariant.

3.2. Approach of ϕ-divergence. In this section we consider the ϕ-divergence approach to construction of the uncertainty sets. The concept of ϕ-divergence is originated in Csiszár [10] and Morimoto [18], and was extensively discussed in Ben-Tal and Teboulle [6]. We also can refer to Bayraksan and Love [4] for a recent survey of this approach. Consider a convex lower semicontinuous function $\phi : \mathbb{R} \to \mathbb{R}_+ \cup \{+\infty\}$ such that $\phi(1) = 0$. For $x < 0$ we set $\phi(x) = +\infty$. Let (cf. [1])

$$\mathfrak{A} := \left\{\zeta \in \mathfrak{D} : \int_\Omega \phi(\zeta(\omega))dP(\omega) \le c\right\} \tag{3.12}$$

for some $c > 0$. If $\zeta \overset{\mathcal{D}}{\sim} \zeta'$, then $\int_\Omega \phi(\zeta(\omega))dP(\omega) = \int_\Omega \phi(\zeta'(\omega))dP(\omega)$, and hence it follows that the set $\mathfrak{A}$ is law invariant.

We view $\mathfrak{A}$ as a subset of an appropriate dual space $\mathcal{Z}^*$. Consider functional

$$\nu(\zeta) := \int_\Omega \phi(\zeta(\omega))dP(\omega),\ \zeta \in \mathcal{Z}^*. \tag{3.13}$$

By Fenchel-Moreau Theorem we have that

$$\phi(x) = \sup_{y\in\mathbb{R}} \{yx - \phi^*(y)\}, \tag{3.14}$$

where $\phi^*(y) := \sup_{x\geq 0}\{yx - \phi(x)\}$ is the conjugate of ϕ. Note that since $\phi(x) = +\infty$ for $x < 0$, it suffices to take maximum in calculation of the conjugate with respect to $x \geq 0$. Note also that $\phi^*(y)$ can be $+\infty$ for some $y \in \mathbb{R}$, and since $\phi(x) \geq 0$ and $\phi(1) = 0$ it follows that $\phi^*(0) = 0$ and $\phi^*(y) \geq y$ for all $y \in \mathbb{R}$.

By using representation (3.14) and interchanging the sup and integral operators[2], we can write functional $\nu(\cdot)$ in the form[3]

$$\nu(\zeta) = \sup_{Y\in\mathcal{Z}} \left\{ \langle Y, \zeta\rangle - \int_\Omega \phi^*(Y(\omega))dP(\omega) \right\}. \tag{3.15}$$

That is, the functional $\nu(\cdot)$ is given by maximum of convex and weakly* continuous (affine) functions, and hence $\nu(\cdot)$ is convex and weakly* lower semicontinuous. It follows that the set $\mathfrak{A} \subset \mathcal{Z}^*$ is convex and weakly* closed.

The corresponding functional $\rho = \rho_{\mathfrak{A}}$ is given by the optimal value of the problem:

$$\begin{array}{cl} \sup\limits_{\zeta\in\mathcal{Z}^*_+} & \int_\Omega Z(\omega)\zeta(\omega)dP(\omega) \\ \text{s.t.} & \int_\Omega \phi(\zeta(\omega))dP(\omega) \leq c,\ \int_\Omega \zeta(\omega)dP(\omega) = 1, \end{array} \tag{3.16}$$

where $\mathcal{Z}^*_+ := \{\zeta \in \mathcal{Z}^* : \zeta \succeq 0\}$. Lagrangian of problem (3.16) is

$$\mathcal{L}_Z(\zeta,\lambda,\mu) = \int_\Omega \left[\zeta(\omega)Z(\omega) - \lambda\phi(\zeta(\omega)) - \mu\zeta(\omega)\right] dP(\omega) + \lambda c + \mu.$$

The Lagrangian dual of problem (3.16) is the problem

$$\inf_{\lambda\geq 0,\mu} \sup_{\zeta\succeq 0} \mathcal{L}_Z(\zeta,\lambda,\mu). \tag{3.17}$$

Since Slater condition holds for problem (3.16) (for example take $\zeta(\cdot) \equiv 1$) and the functional $\nu(\cdot)$ is lower semicontinuous, there is no duality gap between (3.16) and its dual problem (3.17) and the dual problem has a nonempty set of optimal solutions (e.g. [9, Theorem 2.165]).

Since the space $L_q(\Omega,\mathcal{F},P)$ is decomposable, the maximum in (3.17) can be taken inside the integral (cf. [20, Theorem 14.60]), that is

$$\sup_{\zeta\succeq 0} \int_\Omega \left[\zeta(\omega)Z(\omega) - \mu\zeta(\omega) - \lambda\phi(\zeta(\omega))\right] dP(\omega) = \int_\Omega \sup_{z\geq 0}\{z(Z(\omega)-\mu) - \lambda\phi(z)\} dP(\omega).$$

We obtain (cf. [1, Theorem 5.1],[6])

$$\rho(Z) = \inf_{\lambda\geq 0,\mu} \left\{\lambda c + \mu + \mathbb{E}_P\left[(\lambda\phi)^*(Z-\mu)\right]\right\}, \tag{3.18}$$

where $(\lambda\phi)^*$ is the conjugate of $\lambda\phi$. It follows directly from the representation (3.18) that $\rho(\cdot)$ is *law invariant*.

Note that it suffices in (3.17) and (3.18) to take the 'inf' with respect to $\lambda > 0$ rather than $\lambda \geq 0$, and that $(\lambda\phi)^*(y) = \lambda\phi^*(y/\lambda)$ for $\lambda > 0$. Hence $\rho(Z)$ can be be written in the following equivalent form

$$\rho(Z) = \inf_{\lambda>0,\mu} \left\{\lambda c + \mu + \lambda\mathbb{E}_P\left[\phi^*\big((Z-\mu)/\lambda\big)\right]\right\}. \tag{3.19}$$

[2] This is justified since the space $L_q(\Omega,\mathcal{F},P)$ is decomposable (cf., [20, Theorem 14.60]).

[3] Of course, it suffices to take maximum in (3.15) for such $Y \in \mathcal{Z}$ that $\int \phi^*(Y)dP < +\infty$. Note that since $\phi^*(y) \geq y$ and $\int Y dP$ is finite for every $Y \in \mathcal{Z}$, the integral $\int \phi^*(Y)dP$ is well defined.

Consider the uncertainty set $\mathfrak{A} = \mathfrak{A}_c$, defined in (3.12), and the corresponding functional $\rho = \rho_c$ as a function of the constant c. Suppose that $\phi(x) > 0$ for all $x \neq 1$. Then for any density $\zeta \in \mathfrak{D}$ different from the constant density $\mathbf{1}_\Omega$, $\phi(\zeta(\cdot))$ is positive on a set of positive measure and hence $\int_\Omega \phi(\zeta(\omega))dP(\omega) > 0$. Thus in that case $\mathfrak{A}_0 = \cap_{c>0}\mathfrak{A}_c = \{\mathbf{1}_\Omega\}$ and $\rho_0(\cdot) = \mathbb{E}_P[\,\cdot\,]$.

EXAMPLE 3.5. For $\alpha \in (0, 1]$ let $\phi(\cdot) := \mathbb{I}_A(\cdot)$ be the indicator function of the interval $A = [0, \alpha^{-1}]$, i.e., $\phi(x) = 0$ for $x \in [0, \alpha^{-1}]$, and $\phi(x) = +\infty$ otherwise. Then for any $c \geq 0$ the corresponding uncertainty set

$$\mathfrak{A} = \{\zeta \in \mathfrak{D} : \zeta(\omega) \in [0, \alpha^{-1}],\ a.e.\ \omega \in \Omega\}. \tag{3.20}$$

(For $\alpha > 1$ the set in the right hand side of (3.20) is empty.) Note that for any $\lambda > 0$, $\lambda\phi = \phi$. The conjugate of ϕ is $\phi^*(y) = \max\{0, \alpha^{-1}y\} = [\alpha^{-1}y]_+$. In that case (cf. [1],[4])

$$\rho(Z) = \inf_{\mu,\lambda\geq 0} \left\{\lambda c + \mu + \alpha^{-1}\mathbb{E}_P[Z - \mu]_+\right\} = \inf_{\mu} \left\{\mu + \alpha^{-1}\mathbb{E}_P[Z - \mu]_+\right\}. \tag{3.21}$$

That is, here $\rho(Z) = \mathsf{AV@R}_\alpha(Z)$ is the so-called Average Value-at-Risk functional (also called Conditional Value-at-Risk, Expected Shortfall and Expected Tail Loss).

EXAMPLE 3.6. Consider $\phi(x) := x \ln x - x + 1$, $x \geq 0$. Here $\int \phi(\zeta)dP$ defines the Kullback-Leibler divergence, denoted $D_{KL}(\zeta\|P)$. For $\lambda > 0$ the conjugate of $\lambda\psi$ is $(\lambda\phi)^*(y) = \lambda(e^{y/\lambda} - 1)$. In this case it is natural to take $\mathcal{Z} = L_\infty(\Omega, \mathcal{F}, P)$ and to pair it with $L_1(\Omega, \mathcal{F}, P)$.

By (3.18) we have

$$\rho(Z) = \inf_{\lambda\geq 0,\mu} \left\{\lambda c + \mu + \lambda e^{-\mu/\lambda}\mathbb{E}_P\left[e^{Z/\lambda}\right] - \lambda\right\}. \tag{3.22}$$

Minimization with respect to μ in the right hand side of (3.22) gives $\bar{\mu} = \lambda \ln \mathbb{E}_P[e^{Z/\lambda}]$. By substituting this into (3.22) we obtain (cf. [1],[13],[16])

$$\rho(Z) = \inf_{\lambda>0} \left\{\lambda c + \lambda \ln \mathbb{E}_P[e^{Z/\lambda}]\right\}. \tag{3.23}$$

For $c = 0$ the functional $\rho = \rho_0$ is given by the minimum of entropic risk measures $\lambda \ln \mathbb{E}_P[e^{Z/\lambda}]$. Here $\phi(x) > 0$ for any $x \neq 0$, and hence for $c = 0$ the corresponding functional $\rho_0(\cdot) = \mathbb{E}_P[\,\cdot\,]$.

EXAMPLE 3.7. Consider $\phi(x) := |x - 1|$, $x \geq 0$, and $\phi(x) := +\infty$ for $x < 0$. This gives the same uncertainty set $\mathfrak{A}$ as in Example 3.2. It is natural to take here $\mathcal{Z} := L_\infty(\Omega, \mathcal{F}, P)$ and to pair it with $L_1(\Omega, \mathcal{F}, P)$. We have that

$$(\lambda\phi)^*(y) = \begin{cases} -\lambda + [y + \lambda]_+ & if \quad y \leq \lambda, \\ +\infty & if \quad y > \lambda. \end{cases}$$

Hence

$$\begin{aligned} \rho(Z) &= \inf_{\substack{\lambda\geq 0,\mu \\ \operatorname{ess\,sup}(Z-\mu)\leq\lambda}} \{\lambda c + \mu - \lambda + \mathbb{E}_P[Z - \mu + \lambda]_+\} \\ &= \inf_{\substack{\lambda\geq 0,\mu \\ \operatorname{ess\,sup}(Z)\leq\mu+2\lambda}} \{\lambda c + \mu + \mathbb{E}_P[Z - \mu]_+\}. \end{aligned} \tag{3.24}$$

The minimum in μ in the right hand side of (3.24) is attained at $\bar\mu = \operatorname{ess\,sup}(Z) - 2\lambda$. Suppose that $c \in (0, 2)$. Then

$$\begin{aligned}\rho(Z) &= \operatorname{ess\,sup}(Z) + \inf_{\lambda>0}\{\lambda(c-2) + \mathbb{E}_P[Z - \operatorname{ess\,sup}(Z) + 2\lambda]_+\}\\ &= \operatorname{ess\,sup}(Z) + \inf_{t<0}\{t(1-c/2) + \mathbb{E}_P[Z - \operatorname{ess\,sup}(Z) - t]_+\}\\ &= \operatorname{ess\,sup}(Z) + (1-c/2)\inf_{t\in\mathbb{R}}\left\{t + (1-c/2)^{-1}\mathbb{E}_P[Z - \operatorname{ess\,sup}(Z) - t]_+\right\}.\end{aligned}$$

Note that since $Z - \operatorname{ess\,sup}(Z) \preceq 0$ the minimum in the last equation is attained at some $t \le 0$, and this minimum is equal to

$$\mathsf{AV@R}_{1-c/2}[Z - \operatorname{ess\,sup}(Z)] = \mathsf{AV@R}_{1-c/2}[Z] - \operatorname{ess\,sup}(Z).$$

Hence we obtain (cf. [17])

$$\rho(Z) = (c/2)\operatorname{ess\,sup}(Z) + (1-c/2)\mathsf{AV@R}_{1-c/2}[Z]. \tag{3.25}$$

EXAMPLE 3.8. Consider $\phi(x) := [x-1]_+$, $x \ge 0$, and $\phi(x) := +\infty$ for $x < 0$. This gives the same uncertainty set $\mathfrak{A}$ as in Example 3.3. It is natural to take here $\mathcal{Z} := L_\infty(\Omega, \mathcal{F}, P)$ and to pair it with $L_1(\Omega, \mathcal{F}, P)$. We have that

$$(\lambda\phi)^*(y) = \begin{cases} [y]_+ & if \quad y \le \lambda,\\ +\infty & if \quad y > \lambda.\end{cases}$$

Hence

$$\rho(Z) = \inf_{\substack{\lambda\ge 0,\mu\\ \operatorname{ess\,sup}(Z-\mu)\le\lambda}}\{\lambda c + \mu + \mathbb{E}_P[Z-\mu]_+\}. \tag{3.26}$$

Similar to the previous example, the minimum in the right hand side of (3.26) is attained at $\bar\mu = \operatorname{ess\,sup}(Z) - \lambda$. Suppose that $c \in (0, 1)$. Then

$$\begin{aligned}\rho(Z) &= \operatorname{ess\,sup}(Z) + \inf_{\lambda>0}\{\lambda(c-1) + \mathbb{E}_P[Z - \operatorname{ess\,sup}(Z) + \lambda]_+\}\\ &= \operatorname{ess\,sup}(Z) + \inf_{t<0}\{t(1-c) + \mathbb{E}_P[Z - \operatorname{ess\,sup}(Z) - t]_+\}\\ &= \operatorname{ess\,sup}(Z) + (1-c)\inf_{t\in\mathbb{R}}\left\{t + (1-c)^{-1}\mathbb{E}_P[Z - \operatorname{ess\,sup}(Z) - t]_+\right\}.\end{aligned}$$

Hence

$$\rho(Z) = c\operatorname{ess\,sup}(Z) + (1-c)\mathsf{AV@R}_{1-c}[Z]. \tag{3.27}$$

4. Implications of law invariance.

In this section we discuss some implications of the property of law invariance. Unless stated otherwise we assume that the uncertainty set $\mathfrak{A}$ and the respective functional $\rho = \rho_{\mathfrak{A}}$ are law invariant. As it was discussed in section 2 there is a close relation between law invariance of $\mathfrak{A}$ and ρ.

Consider the set

$$\mathfrak{C}(\mathcal{Z}) := \{F : F(z) = P(Z \le z),\ Z \in \mathcal{Z}\} \tag{4.1}$$

of cdfs associated with the space $\mathcal{Z}$. Since the functional ρ is law invariant, it can be considered as a function of the cdf $F = F_Z$, and we sometimes write $\rho(F)$, $F \in \mathfrak{C}(\mathcal{Z})$, for a law invariant functional.

4.1. Sample Average Approximation method. Given a sample $Z_1, ..., Z_N$ of the random variable Z, we can approximate the corresponding cdf $F(z) = P(Z \le z)$ by the empirical cdf

$$\hat{F}_N(z) := \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}_{(-\infty, z]}(Z_j). \tag{4.2}$$

Consequently we can approximate $\rho(F)$ by $\rho(\hat{F}_N)$. In case of ϕ-divergence, when the uncertainty set $\mathfrak{A}$ is of the form (3.12), we can use (3.18) to write

$$\rho(\hat{F}_N) = \inf_{\lambda \ge 0, \mu} \left\{ \lambda c + \mu + \frac{1}{N} \sum_{j=1}^{N} (\lambda\phi)^*(Z_j - \mu) \right\}. \tag{4.3}$$

In general we can proceed as follows. We can write the functional $\rho(F)$ as (e.g. [14, section 4.5])

$$\rho(F) = \sup_{\sigma \in \Upsilon} \int_0^1 \sigma(t) F^{-1}(t) dt, \tag{4.4}$$

where Υ is a set of monotonically nondecreasing functions $\sigma : [0, 1) \to \mathbb{R}_+$ such that $\int_0^1 \sigma(t)dt = 1$ (referred to as spectral functions). Consequently

$$\rho(\hat{F}_N) = \sup_{\sigma \in \Upsilon} \int_0^1 \sigma(t) \hat{F}_N^{-1}(t) dt = \sup_{\sigma \in \Upsilon} \left\{ \sum_{j=1}^{N} Z_{(j)} \int_{\gamma_{j-1}}^{\gamma_j} \sigma(t) dt \right\}, \tag{4.5}$$

where $Z_{(1)} \le ... \le Z_{(N)}$ are the sample values arranged in increasing order and $\gamma_0 = 0$, $\gamma_j = j/N$, $j = 1, ..., N$. Note that $\sum_{j=1}^{N} \int_{\gamma_{j-1}}^{\gamma_j} \sigma(t)dt = \int_0^1 \sigma(t)dt = 1$ for any $\sigma \in \Upsilon$.

Consider now the distributionally robust stochastic programming problem (1.1). Suppose that for every $x \in \mathcal{X}$ the random variable $G(x, \xi(\omega))$ belongs to $\mathcal{Z}$. Let $\xi_1, ..., \xi_N$ be a sample of the random vector $\xi = \xi(\omega)$. The Sample Average Approximation (SAA) of problem (1.1) is obtained by replacing the cdf of the random variable $G(x, \xi)$ by the corresponding empirical cdf based on the sample $G(x, \xi_j)$, $j = 1, ..., N$. It is possible to show that, under mild regularity conditions, the optimal value and optimal solutions of the SAA problem converge w.p.1 to their true counterparts as the sample size N tends to infinity (cf. [23]).

In particular, in the setting of ϕ-divergence the distributionally robust stochastic program (1.1) can be written in the form

$$\operatorname*{Min}_{x \in \mathcal{X}, \lambda \ge 0, \mu} \mathbb{E}_P[\Psi(x, \lambda, \mu, \xi)], \tag{4.6}$$

and the corresponding SAA problem as

$$\operatorname*{Min}_{x \in \mathcal{X}, \lambda \ge 0, \mu} \frac{1}{N} \sum_{j=1}^{N} \Psi(x, \lambda, \mu, \xi_j), \tag{4.7}$$

where

$$\Psi(x, \lambda, \mu, \xi) := \lambda c + \mu + (\lambda\phi)^*(G(x, \xi) - \mu).$$

Note that

$$(\lambda\phi)^*(G(x,\xi)-\mu)=\sup_{z\geq 0}\big\{z(G(x,\xi)-\mu)-\lambda\phi(z)\big\}. \tag{4.8}$$

Suppose that the set $\mathcal{X}$ is convex and for every $\xi\in\Xi$ the function $G(\cdot,\xi)$ is convex. Then the right hand side of (4.8) is the maximum of a family of convex in (λ,μ,x) functions. Consequently the function $\Psi(\cdot,\cdot,\cdot,\xi)$ is convex for all $\xi\in\Xi$, and hence problems (4.6) and (4.7) are convex.

Let ϑ and $\hat\vartheta_N$ be the optimal values of problems (4.6) and (4.7), respectively.

THEOREM 10. *Suppose that:* (i) *the sample* $\xi_1,...,\xi_N$ *is* iid *(independent identically distributed) from the reference distribution* P*,* (ii) *the set* $\mathcal{X}$ *and function* $G(\cdot,\xi)$*, for all* $\xi\in\Xi$*, are convex* (iii) *problem* (4.6) *has a nonempty and bounded set* $\mathcal{S}\subset\mathbb{R}^n\times\mathbb{R}_+\times\mathbb{R}$ *of optimal solutions,* (iv) *there is a (bounded) neighborhood* $\mathcal{V}\subset\mathbb{R}^n\times\mathbb{R}_+\times\mathbb{R}$ *of the set* $\mathcal{S}$ *and a measurable function* $C:\Xi\to\mathbb{R}_+$ *such that* $\mathbb{E}_P[C(\xi)^2]$ *is finite and*

$$|\Psi(x,\lambda,\mu,\xi)-\Psi(x',\lambda',\mu',\xi)|\leq C(\xi)(\|x-x'\|+|\lambda-\lambda'|+|\mu-\mu'|)$$

for all $(x,\lambda,\mu),(x',\lambda',\mu')\in\mathcal{V}$ *and* $\xi\in\Xi$*,* (v) *for some point* $(x,\lambda,\mu)\in\mathcal{V}$ *the expectation* $\mathbb{E}_P\big[\Psi(x,\lambda,\mu,\xi)^2\big]$ *is finite.*

Then

$$\hat\vartheta_N=\inf_{(x,\lambda,\mu)\in\mathcal{S}}\frac{1}{N}\sum_{j=1}^N\Psi(x,\lambda,\mu,\xi_j)+o_p(N^{-1/2}). \tag{4.9}$$

Moreover, if problem (4.6) *has unique optimal solution, i.e., the set* $\mathcal{S}=\{(\bar x,\bar\lambda,\bar\mu)\}$ *is a singleton, then* $N^{1/2}(\hat\vartheta_N-\vartheta)$ *converges in distribution to normal* $\mathcal{N}(0,\sigma^2)$ *with*

$$\sigma^2=\mathrm{Var}_P[\Psi(\bar x,\bar\lambda,\bar\mu,\xi)].$$

Proof. Since the set $\mathcal{S}$, of optimal solutions, is nonempty and bounded and the problem is convex, an optimal solution of the SAA problem (4.7) converges w.p.1 to the set $\mathcal{S}$ (e.g., [24, Theorem 5.4]). Let $\mathcal{V}\subset\mathbb{R}^n\times\mathbb{R}_+\times\mathbb{R}$ be a compact neighborhood of the set $\mathcal{S}$. Then it suffices to perform the optimization in the neighborhood $\mathcal{V}$. That is, restricting minimization in problem (4.6) to the set $\mathcal{V}$ clearly does not change its optimal value ϑ; and for N large enough w.p.1 $\hat\vartheta_N=\hat\vartheta'_N$, where $\hat\vartheta'_N$ is the optimal solution of the restricted problem

$$\operatorname*{Min}_{(x,\lambda,\mu)\in\mathcal{V}}\frac{1}{N}\sum_{j=1}^N\Psi(x,\lambda,\mu,\xi_j). \tag{4.10}$$

The results then follow from a general theory of asymptotics of SAA problems applied to the restricted problem (cf. [22], [24, section 5.1.2]). ∎

For iid sample the rate of convergence of the SAA estimates typically is of order $O_p(N^{-1/2})$, provided that $\mathcal{Z}=L_p(\Omega,\mathcal{F},P)$, $p\in[1,\infty)$ (e.g. [24, section 6.6]). It is interesting to note that in Examples 3.7 and 3.8 the space $\mathcal{Z}=L_\infty(\Omega,\mathcal{F},P)$, and the corresponding functional ρ is a convex combination of the Average Value-at-Risk and the essential sup operators. In that case statistical properties of the SAA estimates are different, we elaborate on this in Example 4.1 below. In Examples 3.7 and 3.8 the conjugate function ϕ^* is discontinuous and condition (iv) of Theorem 10 does not hold.

EXAMPLE 4.1. Let us consider the essential sup operator $\rho(\cdot) := \operatorname{ess\,sup}(\cdot)$ and $Z := U_1 + ... + U_m$, where $U_1, ..., U_m$ are random variables independent of each other and each having uniform distribution on the interval [0,1]. We have that $\rho(Z) = m$. On the other hand, for large m by the Central Limit Theorem, Z has approximately normal distribution with mean $\mu = m/2$ and variance $\sigma^2 = m/12$. The probability that $Z > 0.9m$, say, is given by the probability that $Z > \mu + 1.38\sqrt{m}\,\sigma$ and is very small. More accurately, by Hoeffding inequality

$$P\big\{Z \geq (0.5+\tau)m\big\} \leq e^{-2\tau^2 m}, \;\; 0 < \tau < 0.5.$$

For example for $m = 100$ and $\tau = 0.4$ it follows that $P(Z \geq 0.9m) \leq e^{-32} \approx \frac{1}{6\times 10^{13}}$. That is, one would need the sample size N of order 10^{14} to ensure that probability of the event "$\rho(\hat{F}_N) \geq 0.9m$", i.e., that the sample estimate is within 10% accuracy of the true value, to be close to one. This is in a sharp contrast with $\rho := \mathsf{AV@R}_\alpha$ and say $\alpha = 0.05$. In that case $\rho(\hat{F}_N)$ will converge to $\rho(F)$ at a rate of $O_p(N^{-1/2})$.

4.2. Ambiguous chance constraints. Consider the following so-called ambiguous chance constraint

$$Q\{C(x,\omega) \leq 0\} \geq 1-\varepsilon, \; \forall Q \in \mathfrak{M}, \tag{4.11}$$

where $C : \mathcal{X} \times \Omega \to \mathbb{R}$ and $\varepsilon \in (0,1)$. It is assumed that for every $x \in \mathcal{X}$ the function $C(x,\cdot)$ is measurable. For a measurable set $A \in \mathcal{F}$ we have

$$\sup_{Q\in\mathfrak{M}} Q(A) = \sup_{Q\in\mathfrak{M}} \mathbb{E}_Q[\mathbf{1}_A] = \sup_{\zeta\in\mathfrak{A}} \int_A \zeta(\omega)dP(\omega) = \rho(\mathbf{1}_A), \tag{4.12}$$

where the last equality follows by the definition of the functional ρ. Therefore we can write (4.11) in the form

$$\rho(\mathbf{1}_{A_x}) \leq \varepsilon, \tag{4.13}$$

where

$$A_x := \{\omega \in \Omega : C(x,\omega) > 0\}.$$

Note that for two measurable sets $A, A' \in \mathcal{F}$ the functions $\mathbf{1}_A$ and $\mathbf{1}_{A'}$ are distributionally equivalent iff $P(A) = P(A')$.

We make the following assumption.

Assumption (B) The following implication holds for any $A, B \in \mathcal{F}$:

$$P(B) \leq P(A) \Rightarrow \sup_{Q\in\mathfrak{M}} Q(B) \leq \sup_{Q\in\mathfrak{M}} Q(A). \tag{4.14}$$

This assumption implies that every $Q \in \mathfrak{M}$ is absolutely continuous with respect to P. Indeed consider $B \in \mathcal{F}$ such that $P(B) = 0$ and let $A := \emptyset$ be the empty set. Then $P(A) = 0$, and hence by assumption (B)

$$\sup_{Q\in\mathfrak{M}} Q(B) \leq \sup_{Q\in\mathfrak{M}} Q(A) = 0.$$

It follows that $Q(B) = 0$ for every $Q \in \mathfrak{M}$, and hence Q is absolutely continuous with respect to P.

Remark 4.1. In case the functional ρ is law invariant and the reference probability measure P is nonatomic, assumption (B) holds automatically. Indeed if $A, B \in \mathcal{F}$ and $P(B) \le P(A)$, then since P is nonatomic there is $B' \in \mathcal{F}$ such that $P(B) = P(B')$ and $B' \subset A$. Since $\mathbf{1}_{B'} \preceq \mathbf{1}_A$ it follows by monotonicity of ρ that $\rho(\mathbf{1}_{B'}) \le \rho(\mathbf{1}_A)$, and by law invariance of ρ we have that $\rho(\mathbf{1}_{B'}) = \rho(\mathbf{1}_B)$. Without assuming that P is nonatomic, assumption (B) may not hold even if ρ is law invariant. For example, suppose that the set $\Omega = \{\omega_1, ..., \omega_m\}$ is finite with respective probabilities $p_i > 0$ being essentially different from each other (see Remark 2.2). Then any uncertainty set $\mathfrak{A}$ and functional $\rho = \rho_{\mathfrak{A}}$ are law invariant. In particular we can take $\mathfrak{A} = \{Q\}$ to be a singleton. Then assumption (B) holds iff $\{p_i \ge p_j\} \Rightarrow \{q_i \ge q_j\}$, $i, j \in \{1, ..., m\}$. On the other hand, if probabilities p_i are equal to each other, i.e. $p_i = 1/m$, $i = 1, ..., m$, and ρ is law invariant, then assumption (B) holds.

Consider function $\mathfrak{p} : [0, 1] \to [0, 1]$ defined as

$$\mathfrak{p}(t) := \sup\big\{Q(A) : P(A) \le t,\ A \in \mathcal{F},\ Q \in \mathfrak{M}\big\}. \tag{4.15}$$

By definition of the functional ρ we can write

$$\mathfrak{p}(t) = \sup\big\{\rho(\mathbf{1}_A) : P(A) \le t,\ A \in \mathcal{F}\big\}. \tag{4.16}$$

Also for $\varepsilon \in [0, 1]$ consider

$$\mathfrak{p}^{-1}(\varepsilon) := \inf\{t \in [0, 1] : \mathfrak{p}(t) \ge \varepsilon\}. \tag{4.17}$$

Clearly $\mathfrak{p}(\cdot)$ is nondecreasing on [0,1] and because of assumption (B) we have that for $A \in \mathcal{F}$ and $t^* := P(A)$ it follows that $\mathfrak{p}(t^*) = \rho(\mathbf{1}_A)$. Therefore we can write constraint (4.13) in the following equivalent form

$$\mathfrak{p}(t) \le \varepsilon \text{ subject to } t \ge P(A_x). \tag{4.18}$$

Moreover condition $\mathfrak{p}(t) \le \varepsilon$ can be written as $t \le \mathfrak{p}^{-1}(\varepsilon)$, and hence constraint (4.18) as $P(A_x) \le \mathfrak{p}^{-1}(\varepsilon)$. We obtain the following result.

Proposition 11. *Suppose that assumption* (B) *is fulfilled. Then the ambiguous chance constraint* (4.11) *can be written as*

$$P\{C(x, \omega) \le 0\} \ge 1 - \varepsilon^*, \tag{4.19}$$

where $\varepsilon^* := \mathfrak{p}^{-1}(\varepsilon)$.

This indicates that if assumption (B) is fulfilled, then the computational complexity of the corresponding ambiguous chance constrained problem is basically the same as the computational complexity of the respective reference chance constrained problem provided value $\mathfrak{p}^{-1}(\varepsilon)$ can be readily computed.

4.2.1. Law invariant case. In this section we consider the case of *law invariant* functional $\rho = \rho_{\mathfrak{A}}$. We also assume that the reference probability space is *nonatomic*. Then as it was pointed in Remark 4.1, assumption (B) follows, and hence the ambiguous chance constraint (4.11) can be written as (4.19). Since P is nonatomic, $P(A)$ can be any number in the interval [0,1] for some $A \in \mathcal{F}$. Thus function $\mathfrak{p}(\cdot)$ can be defined as $\mathfrak{p}(t) = \rho(\mathbf{1}_A)$ for $t = P(A)$. Alternatively $\mathfrak{p}(t)$ can be defined as follows. Let $Z_t \sim Ber(t)$ be Bernoulli random variable, i.e., $P(Z_t = 1) = t$ and $P(Z_t = 0) = 1 - t$, $t \in [0, 1]$. By law invariance of ρ we have that $\rho(Z_t)$ is a function of t, and $\mathfrak{p}(t) = \rho(Z_t)$.

In case of nonatomic reference space function $\mathfrak{p}(\cdot)$ has the following properties (cf. [24, Proposition 6.53]): (i) $\mathfrak{p}(0) = 0$ and $\mathfrak{p}(1) = 1$, (ii) $\mathfrak{p}(\cdot)$ is monotonically nondecreasing on the interval $[0, 1]$, (iii) $\mathfrak{p}(\cdot)$ is monotonically increasing on the interval $[0, \tau]$, where

$$\tau := \inf\{t \in [0, 1] : \mathfrak{p}(t) = 1\} = \mathfrak{p}^{-1}(1), \tag{4.20}$$

(iv) if $\mathfrak{M} = \{P\}$, then $\mathfrak{p}(t) = t$ for all $t \in [0, 1]$, and if $\mathfrak{M} \neq \{P\}$, then $\mathfrak{p}(t) > t$ for all $t \in (0, 1)$, (v) $\mathfrak{p}(\cdot)$ is continuous on the interval $(0, 1]$.

For $\gamma := \lim_{t\downarrow 0} \mathfrak{p}(t)$ and $\varepsilon \in (\gamma, 1)$, value $\mathfrak{p}^{-1}(\varepsilon)$ can be computed by solving equation $\mathfrak{p}(t) = \varepsilon$. It can happen that $\gamma > 0$, in which case $\mathfrak{p}^{-1}(\varepsilon) = 0$ for $\varepsilon \in [0, \gamma]$ (see Example 4.4 below). In some cases function $\mathfrak{p}(\cdot)$ and modified significance level ε^* can be computed in a closed form. Consider the setting of ϕ-divergence discussed in section 3.2. By (3.18) in that case we have

$$\mathfrak{p}(t) = \inf_{\lambda \geq 0, \mu} \left\{ \lambda c + \mu + \mathbb{E}[(\lambda\phi)^*(Z_t - \mu)] \right\},\ t \in [0, 1], \tag{4.21}$$

where $Z_t \sim Ber(t)$. Since Z_t can only take value 1 with probability t and value 0 with probability $1 - t$, it follows that

$$\mathfrak{p}(t) = \inf_{\lambda \geq 0, \mu} \left\{ \lambda c + \mu + t[(\lambda\phi)^*(1 - \mu)] + (1 - t)[(\lambda\phi)^*(-\mu)] \right\},\ t \in [0, 1]. \tag{4.22}$$

We have here that $\mathfrak{p}(\cdot)$ is given by minimum of a family of affine functions, and hence $\mathfrak{p}(\cdot)$ is a concave function. It could be noted that in general the function $\mathfrak{p}(\cdot)$ does not have to be concave. Indeed let ρ_1 and ρ_2 be law invariant functionals of the form (1.4), with the corresponding functions $\mathfrak{p}_1$ and $\mathfrak{p}_2$. Then $\rho(\cdot) := \max\{\rho_1(\cdot), \rho_2(\cdot)\}$ is also a law invariant functional of the form (1.4) with the corresponding function $\mathfrak{p}(\cdot) = \max\{\mathfrak{p}_1(\cdot), \mathfrak{p}_2(\cdot)\}$. Maximum of two concave functions can be not concave. This indicated that not every convex, weakly* closed and law invariant set $\mathfrak{A}$ can be represented in the form (3.12) (see Example 4.3 below).

EXAMPLE 4.2. Consider the setting of Example 3.5 with the set $\mathfrak{A}$ of the form (3.20) and $\rho = \mathsf{AV@R}_\alpha$. Here the function $\mathfrak{p} = \mathfrak{p}_\alpha$ is

$$\mathfrak{p}_\alpha(t) = \begin{cases} \alpha^{-1}t & \text{if} \quad t \in [0, \alpha], \\ 1 & \text{if} \quad t \in (\alpha, 1], \end{cases} \tag{4.23}$$

and hence $\varepsilon^* = \alpha\varepsilon$. In this example the constant τ, defined in (4.20), is equal to α.

EXAMPLE 4.3. Consider risk measure $\rho : \mathcal{Z} \to \mathbb{R}$ of the form

$$\rho(Z) := \int_0^1 \mathsf{AV@R}_\alpha(Z) d\mu(\alpha),$$

where μ is a probability measure on the interval $(0, 1]$. The function $\mathfrak{p}(t)$ of this risk measure is given by $\mathfrak{p}(t) = \int_0^1 \mathfrak{p}_\alpha(t) d\mu(\alpha)$, where $\mathfrak{p}_\alpha$ is given in (4.23). Since each function $\mathfrak{p}_\alpha$ is concave, it follows that $\mathfrak{p}$ is also a concave function.

In particular, let $\rho(\cdot) := \beta\mathsf{AV@R}_\alpha(\cdot) + (1 - \beta)\mathsf{AV@R}_1(\cdot)$ (note that $\mathsf{AV@R}_1(\cdot) = \mathbb{E}_P(\cdot)$), for some $\alpha, \beta \in (0, 1)$. Then (cf., [24, p.322])

$$\mathfrak{p}(t) = \begin{cases} (1 - \beta + \alpha^{-1}\beta)t & \text{if} \quad t \in [0, \alpha], \\ \beta + (1 - \beta)t & \text{if} \quad t \in (\alpha, 1], \end{cases}$$

is a concave piecewise linear function. Maximum of such two functions can be non-concave.

EXAMPLE 4.4. Consider the uncertainty set $\mathfrak{A}$ of Example 3.7. In this example, by using (3.25), it can be computed that $\mathfrak{p}(0) = 0$ and $\mathfrak{p}(t) = \min\{t + c/2, 1\}$ for $t \in (0, 1]$. In this example the function $\mathfrak{p}(t)$ is discontinuous at $t = 0$. Also for $\varepsilon \in [0, c/2]$ we have that $\varepsilon^* = \mathfrak{p}^{-1}(\varepsilon) = 0$. That is, for $\varepsilon \in [0, c/2]$ the ambiguous chance constraint (4.11) is equivalent to the constraint that $C(x, \omega) \le 0$ should be satisfied for P-almost every $\omega \in \Omega$.

REFERENCES

[1] A. AHMADI-JAVID, *Entropic value-at-risk: A new coherent risk measure*, Journal Optimization Theory Applications, 155 (2012), pp. 1105–1123.

[2] C. D. ALIPRANTIS AND K. C. BORDER, *Infinite Dimensional Analysis, A Hitchhiker's Guide*, Springer, Berlin, 3rd ed., 2006.

[3] P. ARTZNER, F. DELBAEN, J.-M. EBER, AND D. HEATH, *Coherent measures of risk*, Mathematical Finance, 9 (1999), pp. 203–228.

[4] G. BAYRAKSAN AND D. K. LOVE, *Data-driven stochastic programming using phi-divergences*, Tutorials in Operations Research, INFORMS, (2015).

[5] A. BEN-TAL, A. EL GHAOUI, AND A. NEMIROVSKI, *Robust Optimizations*, Princeton University Press, 2009.

[6] A. BEN-TAL AND M. TEBOULLE, *Penalty functions and duality in stochastic programming via phi-divergence functionals*, Mathematics of Operations Research, 12 (1987), pp. 224–240.

[7] D. BERTSIMAS AND J. SETHURAMAN, *Moment problems and semidefinite optimization*, in Handbook of Semidefinite Programming, H. Wolkowicz, R. Saigal, and L. Vandenberghe, eds., Kluwer Academic Publishers, 2000, pp. 469 – 510.

[8] K. P. S. BHASKARA RAO AND M. BHASKARA RAO, *A remark on nonatomic measures*, The Annals of Mathematical Statistics, 43 (1972), pp. 369–370.

[9] J. F. BONNANS AND A. SHAPIRO, *Perturbation Analysis of Optimization Problems*, Springer Series in Operations Research, Springer, 2000.

[10] I. CSISZÁR, *Eine informationstheoretische ungleichung und ihre anwendung auf den beweis der ergodizitat von markoffschen ketten*, Magyar. Tud. Akad. Mat. Kutato Int. Kozls, 8 (1063).

[11] E. DELAGE AND Y. YE, *Distributionally robust optimization under moment uncertainty with application to data-driven problemsm*, Operations Research, 58 (2010), pp. 595 – 6127.

[12] P. M. ESFAHANI AND D. KUHN, *Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations*, Optimization online, (2015).

[13] H. FÖLLMER AND T. KNISPEL, *Entropic risk measures: Coherence vs. convexity, model ambiguity, and robust large deviations*, Stochastics Dynam., 11 (2011), pp. 1–19.

[14] H. FÖLLMER AND A. SCHIED, *Stochastic Finance: An Introduction in Discrete Time*, Walter de Gruyter, Berlin, 2nd ed., 2004.

[15] A. L. GIBBS AND F. E. SU, *On choosing and bounding probability metrics*, International Statistical Review, 70 (2002), pp. 419–435.

[16] Z. HU AND J. L. HONG, *Kullback-leibler divergence constrained distributionally robust optimization*, Optimization online, (2012).

[17] R. JIANG AND Y. GUAN, *Data-driven chance constrained stochastic program*, Mathematical Programming, 158 (2016), pp. 291–327.

[18] T. MORIMOTO, *Markov processes and the h-theorem*, J. Phys. Soc. Jap., 18 (1963), pp. 328–333.

[19] G. C. PFLUG AND A. PICHLER, *Multistage Stochastic Optimization*, Springer, New York, 2014.

[20] R. T. ROCKAFELLAR AND R. J.-B. WETS, *Variational Analysis*, Springer, New York, 1998.

[21] H. SCARF, *A min-max solution of an inventory problem*, in Studies in the Mathematical Theory of Inventory and Production, Stanford University Press, 1958, pp. 201–209.

[22] A. SHAPIRO, *Asymptotic analysis of stochastic programs*, Annals of Operations Research, 30 (1991), pp. 169–186.

[23] A. SHAPIRO, *Consistency of sample estimates of risk averse stochastic programs*, Journal of Applied Probability, 50 (2013), pp. 533–541.

[24] A. SHAPIRO, D. DENTCHEVA, AND A. RUSZCZYŃSKI, *Lectures on Stochastic Programming: Modeling and Theory*, SIAM, Philadelphia, 2nd ed., 2014.

[25] J. ŽÁČKOVÁ, *On minimax solution of stochastic linear programming problems*, Cas. Pest. Mat., 91 (1966), pp. 423–430.