**Noname manuscript No.**
(will be inserted by the editor)

# Distributionally Robust Stochastic Variational Inequalities

**Hailin Sun · Alexander Shapiro ·
Xiaojun Chen**

Received: date / Accepted: date

**Abstract** We propose a formulation of the Distributionally Robust Variational Inequality (DRVI) to deal with uncertainties of distributions of the involved random variables in variational inequalities. Examples of the DRVI are provided, including the optimality conditions for distributionally robust optimization and Distributionally Robust Games (DRG). The existence of solutions and monotonicity of the DRVI are discussed. Moreover, we propose a Sample Average Approximation (SAA) approach to the DRVI and study its convergence properties. Numerical examples of DRG are presented to illustrate solutions of the DRVI and convergence properties of the SAA approach.

**Keywords** Distributional robustness, variational inequalities, monotonicity, sample average approximation, stochastic games

Hailin Sun
School of Mathematical Sciences, Jiangsu Key Laboratory for NSLSCS, Nanjing Normal University, Nanjing, China, 210023 E-mail: mathhlsun@163.com. Research of this author was partly supported by National Natural Science Foundation of China 12122108 and 11871276.

Alexander Shapiro
School of Industrial and Systems Engineering, Georgia Institute of Technology, USA E-mail: ashapiro@isye.gatech.edu. Research of this author was partly supported by NSF grant 1633196.

Xiaojun Chen
Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China E-mail: maxjchen@polyu.edu.hk. Research of this author was partly supported by Hong Kong Research Grant Council PolyU15300219.

**Mathematics Subject Classification (2020)** 90C33, 90C15

## 1 Introduction

Let $X \subseteq \mathbb{R}^n$ be a nonempty closed convex set and

$$\mathcal{N}_X(x) = \left\{ y \in \mathbb{R}^n : y^\top (x' - x) \le 0,\ x' \in X \right\}$$

be the normal cone to $X$ at $x \in X$ (note that $\mathcal{N}_X(x) = \emptyset$ if $x \notin X$). Let $\xi \in \mathbb{R}^\ell$ be a random vector with support set $\Xi \subset \mathbb{R}^\ell$ equipped with its Borel sigma algebra $\mathcal{B}$ and probability distribution $P$. Consider the stochastic variational inequality (SVI):

$$0 \in \mathbb{E}_P[\Phi(x, \xi)] + \mathcal{N}_X(x), \tag{1}$$

where $\Phi : X \times \Xi \to \mathbb{R}^n$ is such that the corresponding expectation is well defined. By writing $\mathbb{E}_P$ we emphasize that the expectation is taken with respect to the considered probability measure (distribution) $P$ on $(\Xi, \mathcal{B})$. With some abuse of the notation we use $\xi$ to denote a random vector whose probability distribution is supported on the set $\Xi$, and also a point (an element) of the set $\Xi$, specific meaning will be clear from the context.

The SVI provides a unified form of the first order optimality conditions of stochastic optimization, and models numerous equilibrium problems in economics, finance, management and engineering [27,29,31]. In the recent two decades, the SVI has been studied extensively and many new algorithms for solving the SVI have been developed [6,8,12]. Moreover, the two-stage SVI and multi-stage SVI have been introduced and investigated actively in the last few years [7,9,10,24,25,31]. In the SVI, the probability distribution of $\xi$ is supposed to be known (specified) exactly. However, unlike the well-studied distributionally robust optimization (DRO), theories and algorithms of DRVI are very limited. In practice the "true" distribution $\mathbb{P}$ of random variables is not known and could be estimated at best from historical data. The uncertainty of the "true" distribution in itself motivates the distributionally robust approach. We suggest the following formulation of the DRVI as a counterpart of (1):

$$0 \in \mathbb{E}_P[\Phi(x, \xi)] + \mathcal{N}_X(x), \tag{2}$$

$$P \in \arg\max_{Q \in \mathfrak{M}} \mathbb{E}_Q[\phi(x, \xi)], \tag{3}$$

where $\phi : X \times \Xi \to \mathbb{R}$ and $\mathfrak{M}$ is a specified set of probability measures (distributions) on $(\Xi, \mathcal{B})$. Note that by solving the above DRVI we mean to find a pair $\bar{x} \in X$ and $\bar{P} \in \mathfrak{M}$ satisfying (2)-(3).

The robust Linear Complementarity Problem (LCP) is another kind of robust version of the stochastic LCP. The earliest work is [32] which considers $\rho$-robust solutions of robust LCPs. Xie and Shanbhag analyse tractable strictly robust counterparts of uncertain LCPs [34]. More recently, the $\Gamma$-robustness [4] is applied to the robust LCPs [18]. Note that robust LCPs can be considered as a robust version of expected residual minimization formulation [6]. It is also valuable to point out that the robust LCPs try to find a solution which can minimize the gap function over an uncertain set, while the DRVI (2)-(3) is to find a solution of the SVI under the worst distribution over the ambiguity set $\mathfrak{M}$, see Remark 2 for more details.

We give the definition and examples of such DRVIs in section 2. We also highlight the difference between DRVI and SVI, see Remark 1. Our main contributions in this paper are threefold.

- Based on (2)-(3), we propose a comprehensive formulation of the DRVI to deal with the uncertain distribution in the SVI. We show that the first order optimality conditions of distributionally robust optimization and DRG are special cases of this formulation of the DRVI. Moreover, the DRVI formulation can be used to describe stochastic equilibrium problems in the case when the distribution of random variables is ambiguous.
- We define the monotonicity of the DRVI and show the existence of the solutions of the DRVI under certain conditions.
- We investigate convergence properties of the SAA approach to discretization of DRVIs. Moreover, we use numerical examples of DRG to illustrate the formulation of the DRVI and the convergence of the SAA method.

In section 2, we review four fundamental examples that are special cases of the DRVI. The first two examples are the first order optimality conditions of two types of DRO problems. The third example is an equivalent formulation of DRG with convex objective functions of players and share constraints among players. The fourth example discusses a Walrasian equilibrium problem with uncertain costs. In section 3, we define the monotonicity of the DRVI and prove the existence of solutions to the DRVI. In section 4, we propose an SAA approach for the DRVI with the corresponding convergence analysis. In section

5, we use numerical examples to illustrate the DRVI and the convergence of the SAA.

## 2 Formulation of the DRVI

In this section we give an extended (multivariate) definition of the DRVI and consider four relevant examples.

**Definition 1** (DRVI) Let $\mathfrak{M}_i,\ i=1,...,r$, be sets of probability measures on the sample space $(\Xi,\mathcal{B})$, $X\subseteq\mathbb{R}^n$ be a closed convex set with nonempty interior, $\Phi:X\times\Xi\to\mathbb{R}^n$, $\phi_i:X\times\Xi\to\mathbb{R}$, $i=1,\ldots,r$, be continuous functions in $x\in X$ such that $\Phi(x,\cdot)$ and $\phi_i(x,\cdot)$ are measurable. The DRVI is to find a pair $(x,P)\in X\times\mathfrak{M}$ satisfying

$$0\in\mathbb{E}_P[\Phi(x,\xi)]+\mathcal{N}_X(x), \tag{4}$$

$$P_i\in\arg\max_{Q\in\mathfrak{M}_i}\mathbb{E}_Q[\phi_i(x,\xi)],\quad i=1,\ldots,r, \tag{5}$$

where[1] $\mathfrak{M}:=\{P_1\times\ldots\times P_r:P_i\in\mathfrak{M}_i,\ i=1,...,r\}$,

$$\mathbb{E}_P[\Phi(x,\xi)]:=\left(\mathbb{E}_{P_1}[\Phi_1(x,\xi)]^\top,\cdots,\mathbb{E}_{P_r}[\Phi_r(x,\xi)]^\top\right)^\top,$$

with $\Phi(x,\xi)=(\Phi_1^\top(x,\xi),\cdots,\Phi_r^\top(x,\xi))^\top$, $\Phi_i(x,\xi)\in\mathbb{R}^{n_i}$ and $\sum_{i=1}^r n_i=n$.

*Remark 1* It is worthwhile to explain the difference between SVI and DRVI defined above. In the case when the decision makers know the true distributions $P_i$, i.e., $\mathfrak{M}_i=\{P_i\}$, $i=1,\cdots,r$, then (5) trivially holds and the DRVI reduces to SVI. When the decision makers do not know the true distributions $P_i$, but the ambiguity sets $\mathfrak{M}_i$ are specified, then they would like to solve the SVI (4) under the worst random environment (worst distribution), and (5) is used to define the worst distribution over $\mathfrak{M}_i$, $i=1,\cdots,r$. It is interesting to consider how to design $\phi_i$ in (5) (how to define the worst random environment). We give several examples below to illustrate this issue. In Examples 1-3, the worst distributions come from the DRO, which make the worst objective values. Moreover, if there is no objective function, the worst case distributions are still used to describe the worst case random environment, which depends on the decision makers' purposes, see Example 4.

[1] The notation $P_1\times\ldots\times P_r$ stands for the product of measures $P_1,\ldots,P_r$.

*Example 1* Consider the following distributionally robust stochastic program

$$\min_{x\in X}\sup_{P\in\mathfrak{M}}\mathbb{E}_P[\phi(x,\xi)],\tag{6}$$

where $\phi : X\times\Xi\to\mathbb{R}$. A point $(\bar{x},\bar{P})\in X\times\mathfrak{M}$ is a saddle point of the minimax problem (6) iff

$$\bar{x}\in\arg\min_{x\in X}\mathbb{E}_{\bar{P}}[\phi(x,\xi)]\ \text{ and }\ \bar{P}\in\arg\max_{P\in\mathfrak{M}}\mathbb{E}_P[\phi(\bar{x},\xi)].\tag{7}$$

Assuming that $\phi$ is differentiable in $x$ and the differentiation and expectation operators can be interchanged, we can write the optimality conditions for the first problem in (7) in the form (4) with $\Phi(x,\xi):=\nabla_x\phi(x,\xi)$. This leads to the DRVI of the form (4) - (5) with $r=1$.

*Example 2* Consider the following distributionally robust stochastic program[2]

$$\begin{aligned}&\min_{x\in X}\sup_{P_0\in\mathfrak{M}}\mathbb{E}_{P_0}[\phi_0(x,\xi)]\\&\text{s.t.}\ \ \sup_{P_1\in\mathfrak{M}}\mathbb{E}_{P_1}[\phi_1(x,\xi)]\le 0,\end{aligned}\tag{8}$$

where $\phi_i(x,\xi)$, $i=0,1$, are convex and twice continuously differentiable with respect to $x$.

The corresponding Lagrangian function is

$$L(x,\lambda):=\sup_{P_0\in\mathfrak{M}}\mathbb{E}_{P_0}[\phi_0(x,\xi)]+\lambda\sup_{P_1\in\mathfrak{M}}\mathbb{E}_{P_1}[\phi_1(x,\xi)],$$

where $\lambda\ge 0$. Suppose that the supremum in (8) is finite valued for every $x\in X$ and the Slater constraint qualification holds [21], then distributionally robust stochastic program (8) is equivalent to

$$\min_{x\in X}\max_{\lambda\ge 0}\ \sup_{P_0\in\mathfrak{M}}\mathbb{E}_{P_0}[\phi_0(x,\xi)]+\lambda\sup_{P_1\in\mathfrak{M}}\mathbb{E}_{P_1}[\phi_1(x,\xi)].$$

Since $\phi_i, i=0,1$, are convex, the above problem is equivalent to

$$\min_{x\in X}\max_{\lambda\ge 0,P_0\in\mathfrak{M},P_1\in\mathfrak{M}}\ \mathbb{E}_{P_0}[\phi_0(x,\xi)]+\lambda\mathbb{E}_{P_1}[\phi_1(x,\xi)].$$

Then the corresponding DRVI is

$$\begin{aligned}&0\in\mathbb{E}_{P_0}[\nabla_x\phi_0(x,\xi)]+\lambda\mathbb{E}_{P_1}[\nabla_x\phi_1(x,\xi)]+\mathcal{N}_X(x),\\&0\in-\mathbb{E}_{P_1}[\phi_1(x,\xi)]+\mathcal{N}_{\mathbb{R}_+}(\lambda),\\&P_i\in\arg\max_{Q\in\mathfrak{M}}\mathbb{E}_Q[\phi_i(x,\xi)],\quad i=0,1.\end{aligned}$$

[2] For the sake of simplicity we consider here just one constraint; of course this can be extended to a finite number of such constraints in a straightforward way.

*Example 3* Consider the following distributionally robust formulation of Nash equilibrium with $r$ players: find $(x_1^*, ..., x_r^*) \in \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_r}$ such that

$$x_i^* \in \arg\min_{x_i \in X_i} \max_{P_i \in \mathfrak{M}_i} \mathbb{E}_{P_i}[\phi_i(x_i, x_{-i}^*, \xi)],\ i = 1, ..., r. \tag{9}$$

Here $X_i \subset \mathbb{R}^{n_i}$ is a nonempty convex closed set with nonempty interior, $\mathfrak{M}_i$ is a set of probability measures on $(\Xi_i, \mathcal{B}_i)$, $\Xi_i \subset \mathbb{R}^{\ell_i}$, and $\phi_i : \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_r} \times \Xi_i \to \mathbb{R}$, $i = 1, ..., r$. Similar to (7), problem (9) leads to the following DRVI formulation (under appropriate differentiability assumptions)

$$\begin{aligned} &0 \in \mathbb{E}_{P_i}[\Phi_i(x_1, ..., x_r, \xi)] + \mathcal{N}_{X_i}(x_i),\ i = 1, ..., r, \\ &P_i \in \arg\max_{Q_i \in \mathfrak{M}_i} \mathbb{E}_{Q_i}[\phi_i(x_1, ..., x_r, \xi)],\ i = 1, ..., r, \end{aligned}$$

with $\Phi_i(x_1, ..., x_r, \xi) := \nabla_{x_i} \phi_i(x_1, ..., x_r, \xi)$, $i = 1, ..., r$.

*Example 4* Consider the Walrasian equilibrium problem with uncertain costs. Let $m$ and $n$ be the number of economic activities and goods, respectively. The uncertain cost is described by random cost functions $c_i(x_i, \xi_i)$, where $x_i$ denotes the unknown level of the $i$-th activity and $\xi_i$ is a random variable, $i = 1, \cdots, m$. The initial endowment of the $j$-th good is $b_j$ and the demand function for the $j$-th good is $d_j(p)$, where $p \in \mathbb{R}^n$ is the price vector of all goods, $j = 1, \cdots, n$. The technology input-output matrix of the economy is given by the $m \times n$ matrix $A$. Then a pair of activity-price patterns $(y, p)$ is a general equilibrium if the following conditions are satisfied (in the sense of expectation):

$$0 \le x \perp \mathbb{E}_P[c(x, \xi)] - Ap \ge 0, \tag{10}$$

$$0 \le p \perp b + A^\top x - d(p) \ge 0, \tag{11}$$

where $P := (P_1, \cdots, P_m)$ and

$$\mathbb{E}_P[c(x, \xi)] = (\mathbb{E}_{P_1}[c_1(x_1, \xi_1)], \cdots, \mathbb{E}_{P_m}[c_m(x_m, \xi_m)])^\top.$$

Consider the setting where the distributions $P_1, \cdots, P_m$ are not known, and respective ambiguity sets $\mathfrak{M}_1, ..., \mathfrak{M}_m$ are specified. We hope to reduce the total cost, which leads to the worst case random environment

$$P_i \in \arg\max_{Q_i \in \mathfrak{M}_i} \mathbb{E}_{Q_i}[c_i(x_i, \xi_i)],\ i = 1, \cdots, m.$$

Then the DRVI should be

$$0 \le x \perp \mathbb{E}_P[c(x,\xi)] - Ap \ge 0, \tag{12}$$

$$0 \le p \perp b + A^\top x - d(p) \ge 0, \tag{13}$$

$$P_i \in \arg\max_{Q_i \in \mathfrak{M}_i} \mathbb{E}_{Q_i}[c_i(x_i,\xi_i)],\ i = 1,\cdots,m. \tag{14}$$

Note also that the worst distribution can be defined in different ways, which depends on the purpose of decision makers.

*Remark 2* If $X = \mathbb{R}^n_+$, then (2)-(3) reduces to the formulation of the distributionally robust complementarity problem (DRCP)

$$0 \le x \ \perp \mathbb{E}_P[\Phi(x,\xi)] \ge 0, \qquad P \in \arg\max_{Q \in \mathfrak{M}} \mathbb{E}_Q[\phi(x,\xi)]. \tag{15}$$

Other formulation of the DRCP from [9] can be written as follows

$$0 \le x, \quad \max_{P \in \mathfrak{M}} \mathbb{E}_P[-\Phi_i(x,\xi)] \le 0,\ i = 1,\ldots,n, \tag{16}$$

$$\max_{P \in \mathfrak{M}} \mathbb{E}_P[x^\top \Phi(x,\xi)] = 0, \tag{17}$$

which is to find $x$ to solve $0 \le x \ \perp \mathbb{E}_P[\Phi(x,\xi)] \ge 0$ for all $P \in \mathfrak{M}$. Obviously, if $(x^*, P^*)$ is a solution of (16)-(17), then it is a solution of (15) with $\phi(x,\xi) := x^\top \Phi(x,\xi)$.

In the case when (16)-(17) has no solution, it can be considered as a distributionally robust version of expected residual minimization formulation:

$$\begin{aligned} &\min_x \max_{P \in \mathfrak{M}} \mathbb{E}_P[x^\top \Phi(x,\xi)] \\ &\quad s.t. \quad 0 \le x, \quad \mathbb{E}_P[-\Phi(x,\xi)] \le 0,\ \forall P \in \mathfrak{M}. \end{aligned} \tag{18}$$

It is also interesting to see the difference between distributionally robust LCP (15) (with $\Phi(x,\xi) := M(\xi_1)x + q(\xi_2)$) and the robust LCP (e.g. [32,34, 18])

$$\begin{aligned} &\min_x \max_{(\xi_1,\xi_2) \in \Xi} x^\top (M(\xi_1)x + q(\xi_2)) \\ &\quad s.t. \qquad 0 \le x, \quad -(M(\xi_1)x + q(\xi_2)) \le 0,\ \forall (\xi_1,\xi_2) \in \Xi, \end{aligned} \tag{19}$$

where $\xi := (\xi_1,\xi_2)$ is a random vector with support set $\Xi$. Note that distributionally robust LCP (15) is to find a solution of the stochastic LCP under the worst distribution over $\mathfrak{M}$. In contrast, the robust LCP (19) does not consider the distribution of $\xi$, and its solution is not a solution of an LCP in general.

## 3 Existence of solutions of the DRVI

In this section we investigate the existence of solutions of the DRVI in three cases: discrete distributions, continuous distributions and monotone setting.

### 3.1 Finite dimensional setting

Suppose that the random vector $\xi$ has a discrete distribution with a finite support $\Xi := \{\xi^1, ..., \xi^m\}$ of cardinality $m$. This setting is especially relevant for discretization of the involved, possibly continuous, distributions which is required for numerical solutions. Then a probability distribution on $\Xi$ can be identified with probability vector $q \in \Delta_m$, where

$$\Delta_m := \left\{q \in \mathbb{R}^m_+ : q^1 + ... + q^m = 1\right\}.$$

That is, each set $\mathfrak{M}_i$, $i = 1, ..., r$, can be viewed as a subset of $\Delta_m$, and can be assumed to be convex and closed. Condition (5) can be then written as $0 \in -\phi^x + \mathcal{N}_{\mathfrak{M}}(p)$, where

$$\phi^x := (\phi_1(x, \xi^1), ..., \phi_1(x, \xi^m), ..., \phi_r(x, \xi^1), ..., \phi_r(x, \xi^m))^\top \in \mathbb{R}^{rm}$$

and $\mathcal{N}_{\mathfrak{M}}(p)$ is the normal cone to the set $\mathfrak{M} := \mathfrak{M}_1 \times ... \times \mathfrak{M}_r \subset \mathbb{R}^{rm}$ at $p := (p_1, \cdots, p_r)$. Thus in that case the corresponding DRVI can be written as the following finite dimensional Variational Inequality (VI):

$$0 \in \sum_{j=1}^m p^j \Phi(x, \xi^j) + \mathcal{N}_X(x), \tag{20}$$

$$0 \in -\phi^x + \mathcal{N}_{\mathfrak{M}}(p), \tag{21}$$

in variables $(x, p) \in X \times \Delta_m^r$ and $p^j \Phi(x, \xi^j) = (p_1^j \Phi_1(x, \xi^j)^\top, ..., p_r^j \Phi_r(x, \xi^j)^\top)^\top$.

In that setting existence of solution follows by the standard results (e.g., [16, Corollary 2.2.5]).

**Proposition 1** *Suppose $\Phi(x, \xi)$ and $\phi_i(x, \xi)$ are continuous in $x$ for every $\xi \in \Xi$, $i = 1, \cdots, r$, and the set $X$ is bounded (and hence the set $X \times \mathfrak{M} \subset \mathbb{R}^n \times \mathbb{R}^{rm}$ is convex and compact). Then finite dimensional VI* (20)-(21) *has a nonempty and compact solution set.*

### 3.2 Continuous distributions setting

Let us consider now settings with continuous distributions of the random vector $\xi$. We assume the existence of a reference probability measure $\mathbb{P}$ on $(\Xi, \mathcal{B})$ and that the ambiguity set consists of probability measures in some sense close to the reference measure[3] $\mathbb{P}$. To proceed consider the space[4] $\mathcal{Z} := L_p(\Xi, \mathcal{B}, \mathbb{P})$, $p \in [1, \infty)$, and its dual space $\mathcal{Z}^* := L_q(\Xi, \mathcal{B}, \mathbb{P})$, $q \in (1, \infty]$, $1/p + 1/q = 1$. We assume $p = q = 2$ in this section. We also use notations $\Phi_i^x(\cdot) := \Phi_i(x, \cdot)$ and $\phi_i^x(\cdot) := \phi_i(x, \cdot)$.

**Assumption 1** *Suppose that, for $i = 1, ..., r$, the set $\mathfrak{M}_i$ in (5) consists of probability measures that are absolutely continuous with respect to $\mathbb{P}$ and consider the set $\mathfrak{A}_i := \{\zeta = dQ/d\mathbb{P} : Q \in \mathfrak{M}_i\}$ of the corresponding density functions. Suppose further that $\mathfrak{A}_i$ is a bounded, convex and weakly* closed subset of $\mathcal{Z}^*$, and that $\phi_i^x \in \mathcal{Z}$ for every $x \in X$ and $i = 1, ..., r$.*

Assumption 1 will hold for several settings of ambiguity sets, e.g., law invariant coherent risk measure [27, section 6.3.2], [28], $\psi$-divergence ball[5] [22], [28, Section 3.2] and so on. Moreover, Assumption 1 can also hold in the discrete distributions setting when $\mathbb{P}$ is discrete (e.g. $\mathbb{P}$ is empirical distribution).

*Remark 3* Historically the first approach to what is now called "distributional robustness", was based on the following argument. The employed probability distribution is not 'exact' and hence one can hedge against a worst case by considering a family of probability distributions. In that approach, construction (estimation) of the nominal distribution from the available data is based on a (parametric) model, and typically the constructed distribution is continuous. More recently the 'data driven' approach became popular. It is assumed that

[3] For convenience, we use the same notation $\mathbb{P}$ for the true distribution and the reference measure. We can distinguish them by context.

[4] Banach spaces $\mathcal{Z}$ and $\mathcal{Z}^*$, equipped with the respective weak and weak* topologies, are paired topological vector spaces with respect to the bilinear form $\langle \zeta, Z \rangle = \int_\Xi \zeta Z d\mathbb{P}$, $Z \in \mathcal{Z}$, $\zeta \in \mathcal{Z}^*$. Note that the weak topology of $\mathcal{Z}$ and weak* topology of $\mathcal{Z}^*$, restricted to respective bounded sets, are metrizable and hence can be described in terms of convergent sequences. The weak convergence $Z_k \overset{w}{\to} \bar{Z}$ means that $\langle \zeta, Z_k \rangle$ converges to $\langle \zeta, \bar{Z} \rangle$ for any $\zeta \in \mathcal{Z}^*$. The weak* convergence $\zeta_k \overset{w^*}{\to} \bar{\zeta}$ means that $\langle \zeta_k, Z \rangle$ converges to $\langle \bar{\zeta}, Z \rangle$ for any $Z \in \mathcal{Z}$.

[5] In some publications "$\phi$-divergence", rather than "$\psi$-divergence", terminology is used. Here we use the definition of $\psi$-divergence from [28, Section 3.2] and its references. The precise definition will be given later (see Example 7 below).

the observed (finite) random sample is coming from the 'true' distribution. If this assumption about the existence of the true distribution is accepted, then the question is why it is better to use the distributionally robust approach, with quite an artificially constructed ambiguity set, rather then the empirical distribution which was used in statistics for a long time. Of course, a polemic about this is far beyond the scope of this paper.

Since $\phi_i^x \in \mathcal{Z}$, it follows that for any $\zeta \in \mathfrak{A}_i$ and $dQ = \zeta d\mathbb{P}$ the integral

$$\mathbb{E}_Q[\phi_i^x] = \int_\Xi \phi_i^x \, \zeta d\mathbb{P}$$

is well defined and finitely valued. In what follows, we consider the ambiguity sets that satisfy Assumption 1 and hence (4)-(5) can be rewritten as

$$0 \in \int_\Xi \Phi_i^x \zeta_i d\mathbb{P} + \mathcal{N}_{X_i}(x_i), \quad i = 1, \ldots, r, \tag{22}$$

$$\zeta_i \in \arg\max_{\eta \in \mathfrak{A}_i} \int_\Xi \phi_i^x \eta d\mathbb{P}, \quad i = 1, \ldots, r. \tag{23}$$

Under Assumption 1, for $i = 1, \cdots, r$, the set $\mathfrak{A}_i$ is convex and closed in the weak* topology of $\mathcal{Z}_i^*$, and hence is weakly* compact. It follows that the set

$$\bar{\mathfrak{A}}_i^x := \arg\max_{\eta \in \mathfrak{A}_i} \int_\Xi \phi_i^x \, \eta d\mathbb{P}$$

is nonempty for any $x \in X$ (note that the set $\bar{\mathfrak{A}}_i^x$ represents the set of densities of the "arg max" probability measures in the right hand side of (5)) and $i$. Consider the mapping $\Phi$ and denote $\Phi^x(\cdot) := \Phi(x, \cdot)$. Suppose that for every $x \in X$, every component of $\Phi^x$ belongs to the space $\mathcal{Z} := \mathcal{Z}^1 \times \cdots \times \mathcal{Z}^r$. Consider the multifunction $\mathfrak{F} : X \rightrightarrows \mathbb{R}^n$ defined as

$$\mathfrak{F}(x) := \left\{ y = \int_\Xi \Phi^x \, \zeta d\mathbb{P} : \zeta \in \bar{\mathfrak{A}}^x \right\},$$

where $\zeta = (\zeta_1, ..., \zeta_r)$ with $\zeta_i \in \bar{\mathfrak{A}}_i^x$, $i = 1, ..., r$ and

$$\Phi^x \, \zeta := (\Phi_1(x, \cdot) \, \zeta_1, ..., \Phi_r(x, \cdot) \, \zeta_r)^\top.$$

In order to show the existence of solutions of the DRVI, we need to verify that the following generalized equations have a solution

$$0 \in \mathfrak{F}(x) + \mathcal{N}_X(x). \tag{24}$$

**Proposition 2** *Suppose that the set $X$ is nonempty convex closed and bounded, and the mappings $\phi^x$ and $\Phi^x$ are weakly continuous with respect to $x \in X$. Then the generalized equation* (24) *has a solution.*

*Proof* It suffices to verify that the multifunction $\mathfrak{F}$ is closed, that is, for any sequences $x_k \in X$ converging to $\bar{x}$ and $y_k \in \mathfrak{F}(x_k)$ converging to $\bar{y}$, we have $\bar{y} \in \mathfrak{F}(\bar{x})$. Indeed, consider the multifunction $\mathfrak{G} : X \rightrightarrows X$ defined as

$$\mathfrak{G}(x) := \arg\min_{v \in X} \{\mathrm{dist}(v, \mathfrak{F}(x))\},$$

where $\mathrm{dist}(x, A)$ denotes the Euclidean distance from $x$ to set $A \subset \mathbb{R}^n$. Note that if the set $A$ is convex, then $\mathrm{dist}(\cdot, A)$ is a convex function. We have that for every $x \in X$, the set $\bar{\mathfrak{A}}^x$ is convex and hence the set $\mathfrak{F}(x)$ is convex, and thus $\mathfrak{G}(x)$ is convex. Also if $\mathfrak{F}$ is closed, then $\mathfrak{G}$ is closed. It follows by Kakutani's fixed-point theorem that the multifunction $\mathfrak{G}$ has a fixed point $\bar{x} \in X$. Let $\bar{y}$ be the closest point of $\mathfrak{F}(\bar{x})$ to $\bar{x}$. Then $\bar{y} - \bar{x} \in \mathcal{N}_X(\bar{x})$.

In order to verify that $\mathfrak{F}$ is closed we can proceed as follows. By the weak$^*$ compactness of $\mathfrak{A}$ and the weak continuity of $\phi^x$, we have that the multifunction $X \ni x \mapsto \bar{\mathfrak{A}}^x$ is weakly$^*$ closed[6] (e.g., [5, Propsition 4.4 and discussion on p. 264 ]). By the weak continuity of $\Phi^x$ it follows that $\mathfrak{F}$ is closed. This follows from the fact that if $Z_k \stackrel{w}{\to} \bar{Z}$ and $\zeta_k \stackrel{w^*}{\to} \bar{\zeta}$, then $\langle \zeta_k, Z_k \rangle \to \langle \bar{\zeta}, \bar{Z} \rangle$ (e.g., [5, Theorem 2.23 (iv)]). □

*Remark 4* Recall that it is assumed that $\phi^x \in \mathcal{Z}$ for every $x \in X$. The mapping $\phi^x$ is weakly continuous if $\phi(x, \xi)$ is continuous in $x$ and there is $\eta \in \mathcal{Z}$ such that $|\phi(x, \xi)| \le \eta(\xi)$ for all $x \in X$ and $\xi \in \Xi$. Indeed, for any $\zeta \in \mathcal{Z}^*$ we have that $|\phi^x(\xi)\zeta(\xi)| \le \eta(\xi)|\zeta(\xi)|$ and $\int \eta(\xi)|\zeta(\xi)| d\mathbb{P} < \infty$. Thus for a sequence $\{x_k\} \subset X$ converging to $\bar{x}$ it follows by the Lebesgue dominated convergence theorem that

$$\lim_{k \to \infty} \int_{\Xi} \phi^{x_k}(z)\, \zeta(z) d\mathbb{P}(z) = \int_{\Xi} \lim_{k \to \infty} \phi^{x_k}(z)\, \zeta(z) d\mathbb{P}(z) = \int_{\Xi} \phi^{\bar{x}}(z)\, \zeta(z) d\mathbb{P}(z).$$

This shows that $\phi^x$ is weakly continuous. Similar conditions can be applied to every component of the mapping $\Phi^x$ to guarantee its weak continuity.

*Remark 5* For the set $\mathfrak{A}_i \subset \mathcal{Z}_i^*$ of density functions, we consider functional $\mathcal{R}_i : \mathcal{Z}_i \to \mathbb{R}$ defined as

$$\mathcal{R}_i(Z) := \sup_{\zeta_i \in \mathfrak{A}_i} \int_{\Xi} Z \zeta^i d\mathbb{P}. \tag{25}$$

[6] That is, if $x_k \in X$ converges to $\bar{x}$ and $\zeta_k \in \bar{\mathfrak{A}}_{x_k}$ is such that $\zeta_k \stackrel{w^*}{\to} \bar{\zeta}$, then $\bar{\zeta} \in \bar{\mathfrak{A}}_{\bar{x}}$.

Since $\mathfrak{A}_i$ is a bounded subset of $\mathcal{Z}_i^*$, the value $\mathcal{R}_i(Z)$ is finite for any $Z \in \mathcal{Z}_i$. This functional can be viewed as the dual representation of the corresponding so-called coherent risk measure. Various examples of coherent risk measures, their dual representations and closed forms for the corresponding sets $\bar{\mathfrak{A}}^x$ are given in (e.g., [27, section 6.3.2]).

The optimality condition (5) can be written in the VI form as follows. For each player $i$, $i = 1, ..., r$, recall that $\mathcal{Z}_i$ and $\mathcal{Z}_i^*$ can be viewed as paired spaces with respect to the bilinear form $\langle \zeta_i, Z \rangle = \int_\Xi \zeta_i Z d\mathbb{P}$. Consider the indicator function $\mathbb{I}_{\mathfrak{A}_i}(\cdot)$ of the set $\mathfrak{A}_i \subset \mathcal{Z}_i^*$, that is $\mathbb{I}_{\mathfrak{A}_i}(\zeta_i) = 0$ for $\zeta_i \in \mathfrak{A}_i$ and $\mathbb{I}_{\mathfrak{A}_i}(\zeta_i) = +\infty$ for $\zeta_i \notin \mathfrak{A}_i$. At a point $\zeta_i \in \mathfrak{A}_i$ the subdifferential $\partial \mathbb{I}_{\mathfrak{A}_i}(\zeta)$ is equal to the normal cone

$$\mathcal{N}_{\mathfrak{A}_i}(\zeta_i) = \{Z \in \mathcal{Z}_i : \langle \eta - \zeta_i, Z \rangle \leq 0, \ \forall \eta \in \mathfrak{A}_i\}.$$

For $\zeta_i \notin \mathfrak{A}_i$ the normal cone $\mathcal{N}_{\mathfrak{A}_i}(\zeta_i) = \emptyset$. For $Z \in \mathcal{Z}_i$ we have that $\bar{\zeta}_i \in \arg\min_{\zeta_i \in \mathfrak{A}_i} \langle \zeta_i, -Z \rangle$ iff $\bar{\zeta}_i \in \arg\min_{\zeta_i \in \mathcal{Z}_i^*} \langle \zeta_i, -Z \rangle + \mathbb{I}_{\mathfrak{A}_i}(\zeta_i)$. Since the subdifferential of $\langle \zeta_i, -Z \rangle + \mathbb{I}_{\mathfrak{A}_i}(\zeta_i)$ at $\bar{\zeta}_i$ is equal to $-Z + \partial \mathbb{I}_{\mathfrak{A}_i}(\bar{\zeta}_i)$, it follows that $\bar{\zeta}_i \in \arg\min_{\zeta_i \in \mathfrak{A}_i} \langle \zeta_i, -Z \rangle$ iff $0 \in -Z + \mathcal{N}_{\mathfrak{A}_i}(\bar{\zeta}_i)$, that is, $0 \in -\phi(x, \cdot) + \mathcal{N}_{\mathfrak{A}_i}(\zeta_i)$. Therefore the optimality condition (5) can be written here as

$$0 \in -\phi^x + \mathcal{N}_{\mathfrak{A}}(\zeta), \tag{26}$$

where $\mathfrak{A} := \mathfrak{A}^1 \times \cdots \times \mathfrak{A}^r$ and $\zeta = (\zeta^1, ..., \zeta^r)$. Note that by pairing $\mathcal{Z}$ and $\mathcal{Z}^*$, the normal cone $\mathcal{N}_{\mathfrak{A}}(\zeta)$ is a subset of the space $\mathcal{Z}$.

This can be compared with the finite dimensional setting discussed in section 3.1. Let $\mathbb{P}$ be the probability measure on the corresponding finite set $\Xi = \{\xi^1, ..., \xi^m\}$ assigning equal probability $1/m$ to each elementary event. Then any probability measure $Q$ on $\Xi$ is absolutely continuous with respect to $\mathbb{P}$ and its density $dQ/d\mathbb{P}$ is given by $mq$ where $q \in \Delta_m$ is the respective probability vector.

### 3.3 Monotonicity property

By (26) in section 3.2, we can write DRVI (22)-(23) as follows:

$$0 \in \int_\Xi \Phi^x \zeta d\mathbb{P} + \mathcal{N}_X(x), \tag{27}$$

$$\mathbf{0} \in -\phi^x + \mathcal{N}_{\mathfrak{A}}(\zeta), \tag{28}$$

where $\mathbf{0} : \varXi \to \mathbb{R}^r$ is a constant function with value 0, $\phi^x = (\phi_1(x,\cdot), \cdots, \phi_r(x,\cdot))^\top$ is a vector-valued random function, and $\mathcal{N}_{\mathfrak{A}}(\zeta) := \mathcal{N}_{\mathfrak{A}_1}(\zeta_1) \times \cdots \times \mathcal{N}_{\mathfrak{A}_r}(\zeta_r)$.

Note that $\mathbb{R}^n \times \mathcal{Z}$ and $\mathbb{R}^n \times \mathcal{Z}^*$ are paired by the respective bilinear form; that is, for $x, z \in \mathbb{R}^n$, $u \in \mathcal{Z}$ and $\zeta \in \mathcal{Z}^*$,

$$\langle (x,u),(z,\zeta)\rangle := x^\top z + \sum_{i=1}^r \int_\varXi u_i \zeta_i d\mathbb{P}.$$

Consider mapping $\mathcal{G} : \mathbb{R}^n \times \mathcal{Z}^* \to \mathbb{R}^n \times \mathcal{Z}$ defined as

$$\mathcal{G}(x,\zeta) := \begin{pmatrix} \int_\varXi \varPhi^x \zeta d\mathbb{P} \\ -\phi^x \end{pmatrix},$$

and denote the DRVI (4)-(5) by DRVI($\mathcal{G}, (X, \mathfrak{A})$). Monotonicity properties of this mapping are defined in the usual way. In particular, the mapping $\mathcal{G}$ is said to be monotone if for any $(x,\zeta),(z,\eta) \in \mathbb{R}^n \times \mathfrak{A}$, we have

$$\left\langle \mathcal{G}(x,\zeta) - \mathcal{G}(z,\eta), \begin{pmatrix} x-z \\ \zeta-\eta \end{pmatrix} \right\rangle \geq 0, \tag{29}$$

and $\mathcal{G}$ is said to be strongly monotone if there is $\alpha > 0$ such that

$$\left\langle \mathcal{G}(x,\zeta) - \mathcal{G}(z,\eta), \begin{pmatrix} x-z \\ \zeta-\eta \end{pmatrix} \right\rangle \geq \alpha(\|x-z\|^2 + \|\zeta-\eta\|_{L_2}^2), \tag{30}$$

where $\|\zeta - \eta\|_{L_2}$ is defined by function metric in $L_2$ space. Moreover, $\mathcal{G}$ cannot be strongly monotone, since $\mathcal{G}(x,\zeta) - \mathcal{G}(x,\eta) = \big((\int_\varXi \varPhi^x \zeta d\mathbb{P})^\top - (\int_\varXi \varPhi^x \eta d\mathbb{P})^\top, \mathbf{0}\big)^\top$. However, $\mathcal{G}$ can be monotone under some reasonable conditions.

In what follows, we investigate the monotonicity of $\mathcal{G}$. For $j = 1, \cdots, r$, $\tilde{\xi}_j$ is a random vector with support set $\varXi$ and continuous distribution $Q_j$ such that $dQ_j = \zeta_j d\mathbb{P}$, let $S_j(\xi)^\top, S_j(\tilde{\xi}_j)^\top \in \mathbb{R}^r$, $\tilde{\xi} := (\tilde{\xi}_1, \cdots, \tilde{\xi}_r) \in \varXi_r \subseteq \mathbb{R}^{r\ell}$,

$$S(\tilde{\xi}) := (S_1(\tilde{\xi}_1)^\top, \cdots, S_r(\tilde{\xi}_r)^\top)^\top \quad \text{and} \quad S(\xi) := (S_1(\xi)^\top, \cdots, S_r(\xi)^\top)^\top.$$

Note that

$$\int_\varXi S(\xi)\zeta d\mathbb{P} = \left( \left( \int_\varXi S_1(\xi)\zeta_1 d\mathbb{P} \right)^\top, \cdots, \left( \int_\varXi S_r(\xi)\zeta_r d\mathbb{P} \right)^\top \right)^\top,$$

and the positive semidefiniteness of $S(\xi)$ for every $\xi \in \varXi$ cannot guarantee that $\int_\varXi S(\xi)\zeta d\mathbb{P}$ is positive semidefinite, unless $\zeta_i = \zeta_j, i,j \in \{1, \cdots, r\}$. The following lemma gives a sufficient condition of the positive semidefiniteness of $\int_\varXi S(\xi)\zeta d\mathbb{P}$ and is used to show the monotonicity of $\mathcal{G}$. The key idea of the proof for the lemma is: we consider an SAA approach of $\int_{\Xi} S(\xi)\zeta d\mathbb{P}$, and then rewrite the SAA term as a finite sum of many positive semidefinite matrices by a decomposition procedure.

**Lemma 1** *Suppose for any $\tilde{\xi} \in \Xi_r$, $S(\tilde{\xi})$ is a positive semidefinite matrix. Then for any $\zeta \in \mathfrak{A}$, $\int_{\Xi} S(\xi)\zeta d\mathbb{P}$ is positive semidefinite.*

*Proof* We consider a discrete approximation of $\int_{\Xi} S(\xi)\zeta d\mathbb{P}$ firstly. Let $\Xi_j^N := \{\xi_j^1, \cdots, \xi_j^N\}$, $j = 1, \cdots, N$, be a discrete approximation of $\Xi$ with the weight vectors $\{p_1^1, \cdots, p_1^N\}, \cdots, \{p_r^1, \cdots, p_r^N\}$ such that for $i = 1, \cdots, N$, $j = 1, \cdots, r$, $p_j^i \geq 0$, $\sum_{i=1}^N p_j^i = 1$ and w.p.1

$$\lim_{N\to\infty} \sum_{i=1}^N p_j^i S_j(\xi^i) = \int_{\Xi} S_j(\xi)\zeta_j d\mathbb{P}. \tag{31}$$

For any given $\zeta \in \mathfrak{A}$, there are several ways to construct an approximation above. One way is to construct i.i.d. samples $\Xi_j^N := \{\xi_j^1, \cdots, \xi_j^N\}$ from continuous distribution $Q_j$ such that $dQ_j = \zeta_j d\mathbb{P}$ for $j = 1, \cdots, r$. Then $p_j^i = \frac{1}{N}$ if $\xi_j^i \in \Xi_j^N$ and $p_j^i = 0$ otherwise.

Let $P^i := \mathrm{diag}(p_1^i, \cdots, p_r^i)$ for $i = 1, \cdots, N$. Then

$$\sum_{i=1}^N P^i S(\xi^i) := \left(\sum_{i=1}^N p_1^i S_1(\xi^i)^\top, \cdots, \sum_{i=1}^N p_r^i S_r(\xi^i)^\top\right)^\top$$

is a discrete approximation of $\int_{\Xi} S(\xi)\zeta d\mathbb{P}$. In what follows, we prove that $\sum_{i=1}^N P^i S(\xi^i)$ is positive semidefinite.

To this end, we do the following procedure.

Step 0. Let $k = 1$. We reorder the weight vectors $\{p_1^1, \cdots, p_1^N\}, \cdots, \{p_r^1, \cdots, p_r^N\}$ to $\{p_1^{(1)}, \cdots, p_1^{(N)}\}, \cdots, \{p_r^{(1)}, \cdots, p_r^{(N)}\}$ such that $p_j^{(1)} \geq p_j^{(2)} \cdots \geq p_j^{(N)}$ for $j = 1, \cdots, r$. Let $\tilde{\xi}^k = (\tilde{\xi}_1^{(1)}, \cdots, \tilde{\xi}_r^{(1)})$ and $\tilde{p}_k = \min\{p_j^{(1)}, j = 1, \cdots, r\}$. We construct $\tilde{p}_k S(\tilde{\xi}^k)$ and reduce $\tilde{p}_k$ from $p_j^{(1)}$, that is new $p_j^{(1)} := p_j^{(1)} - \tilde{p}_k$, for $j = 1, \cdots, r$. Note that by the condition of the lemma, $\tilde{p}_k S(\tilde{\xi}^k)$ is positive semidefinite. Let $k = k + 1$.

Step 1. For $j = 1, \cdots, r$, since we have reduced $\tilde{p}_{k-1}$ from $p_j^{(1)}$ and $p_j^{(1)}$ may not be the largest one of $\{p_j^{(1)}, \cdots, p_j^{(N)}\}$, we reorder the the weight vectors again. To easy notation, we still denote the newly reordered weight vectors as $\{p_1^{(1)}, \cdots, p_1^{(N)}\}, \cdots, \{p_r^{(1)}, \cdots, p_r^{(N)}\}$. Note that now $\sum_{i=1}^N p_j^{(i)} = 1 - \sum_{i=1}^{k-1} \tilde{p}_i$. If $\sum_{i=1}^N p_j^{(i)} = 0$, for $j = 1, \cdots, r$, stop. Note that $\sum_{i=1}^N p_j^{(i)} = \sum_{i=1}^N p_l^{(i)}$ for all $j, l \in \{1, \cdots, r\}$. Otherwise, go to Step 2.

Step 2. Let $\tilde{\xi}^k = (\tilde{\xi}_1^{(1)}, \cdots, \tilde{\xi}_r^{(1)})$ and $\tilde{p}_k = \min\{p_j^{(1)}, j = 1, \cdots, r\}$. We construct $\tilde{p}_k S(\tilde{\xi}^k)$ and reduce $\tilde{p}_k$ from $p_j^{(1)}$, that is new $p_j^{(1)} := p_j^{(1)} - \tilde{p}_k$, for $j = 1, \cdots, r$. Note that by the condition of the lemma, $\tilde{p}_k S(\tilde{\xi}^k)$ is positive semidefinite. Let $k = k+1$. Go to Step 1.

Note that $\tilde{\xi}^k \in \Xi_r^N$ and $|\Xi_r^N| = N^r$, the procedure above will stop at finite iterations, denote by $K$, $K \le N^r$. Since $S(\tilde{\xi}^k)$ is positive semidefinite for $k = 1, \cdots, K$ and $\sum_{k=1}^K \tilde{p}_k = 1$, then

$$\sum_{i=1}^N P^i S(\xi^i) = \sum_{k=1}^K \tilde{p}_k S(\tilde{\xi}^k)$$

is positive semidefinite, and by (31), $\int_\Xi S(\xi)\zeta d\mathbb{P}$ is positive semidefinite w.p.1. Consequently the deterministic matrix $\int_\Xi S(\xi)\zeta d\mathbb{P}$ is positive semidefinite. $\square$

We use $J_x f(x, \cdot)$ to denote the partial Jacobian of $f$ with respect to $x$. Let $\tilde{\xi} := (\tilde{\xi}_1, \cdots, \tilde{\xi}_r) \in \Xi_r \subset \mathbb{R}^{r\ell}$, and $\int_\Xi J_x \Phi(x,\xi)\zeta d\mathbb{P} := \begin{pmatrix} \int_\Xi J_x \Phi_1^x \zeta_1 d\mathbb{P} \\ \vdots \\ \int_\Xi J_x \Phi_r^x \zeta_r d\mathbb{P} \end{pmatrix}$.

**Proposition 3** *Consider DRVI* (27)-(28)*. Suppose:* (a) *for $i = 1, \cdots, r$ and $\xi \in \Xi$, $\Phi_i(\cdot, \xi)$ and $\phi_i(\cdot, \xi)$ are continuously differentiable;* (b) *for any $\tilde{\xi} \in \Xi^r$ and $x \in X$, $\left((J_x\Phi_1(x, \tilde{\xi}_1))^\top, \cdots, (J_x\Phi_r(x, \tilde{\xi}_r))^\top\right)^\top$ is a positive semidefinite matrix;* (c) *for $\mathbb{P}$-a.e. $\xi \in \Xi$, $\tilde{x}_i^\top \Phi_i(x,\xi) - J_x\phi_i(x,\xi)\tilde{x} \ge 0$, for all $x \in X$ and $\tilde{x} \in \mathbb{R}^n$, $i = 1, \cdots, r$. Then $\mathcal{G}$ is monotone over $X \times \mathfrak{A}$.*

*Proof* It is easy to observe that

$$J\mathcal{G}_{(x,\zeta)} = \begin{pmatrix} \int_\Xi J_x\Phi_1^x\zeta_1 d\mathbb{P} & \Phi_1(x,\cdot) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \int_\Xi J_x\Phi_r^x\zeta_r d\mathbb{P} & \mathbf{0} & \cdots & \Phi_r(x,\cdot) \\ -J_x\phi_1(x,\cdot) & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ -J_x\phi_r(x,\cdot) & \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix}, J\mathcal{G}_{(x,\zeta)} \begin{pmatrix} \tilde{x} \\ \tilde{\zeta} \end{pmatrix} = \begin{pmatrix} \int_\Xi J_x\Phi_1^x\zeta_1 d\mathbb{P}\tilde{x} + \int_\Xi \Phi_1^x\tilde{\zeta}_1 d\mathbb{P} \\ \vdots \\ \int_\Xi J_x\Phi_r^x\zeta_r d\mathbb{P}\tilde{x} + \int_\Xi \Phi_r^x\tilde{\zeta}_r d\mathbb{P} \\ -J_x\phi_1(x,\cdot)\tilde{x} \\ \vdots \\ -J_x\phi_r(x,\cdot)\tilde{x} \end{pmatrix}$$

and

$$\left\langle (\tilde{x}, \tilde{\zeta}), J\mathcal{G}_{(x,\zeta)} \begin{pmatrix} \tilde{x} \\ \tilde{\zeta} \end{pmatrix} \right\rangle = \sum_{i=1}^r \left( \tilde{x}_i^\top \int_\Xi J_x\Phi_i^x\zeta_i d\mathbb{P}\tilde{x} + \tilde{x}_i^\top \int_\Xi \Phi_i^x\tilde{\zeta}_i d\mathbb{P} - \int_\Xi J_x\phi_i^x\tilde{\zeta}_i d\mathbb{P}\tilde{x} \right).$$

By condition (b) and Lemma 1, $\int_\Xi J_x\Phi(x,\xi)\zeta d\mathbb{P}$ is positive semidefinite. By condition (c), for any $\tilde{x}\in X$ and $\tilde{\zeta}\in\mathcal{Z}$,

$$\sum_{i=1}^{r}\left(\tilde{x}_i^\top\int_\Xi \Phi_i^x\tilde{\zeta}_i d\mathbb{P}-\int_\Xi J_x\phi_i^x\tilde{\zeta}_i d\mathbb{P}\tilde{x}\right)\geq 0. \tag{32}$$

Then $\left\langle(\tilde{x},\tilde{\zeta}),J\mathcal{G}_{x,\zeta}\begin{pmatrix}\tilde{x}\\ \tilde{\zeta}\end{pmatrix}\right\rangle\geq 0$ holds for any $\tilde{x}\in\mathbb{R}^n$ and $\tilde{\zeta}\in\mathcal{Z}$, and by [13, Theorem 3.1], $\mathcal{G}(\cdot,\cdot)$ is monotone over $X\times\mathfrak{A}$. □

The above proposition shows the monotonicity properties of the DRVI in the continuous distributions setting (section 3.2). Note that we may weaken the continuous differentiability condition to Lipschitz continuity by considering the Clarke subdifferential and Clarke Jacobian instead of differential and Jacobian. Note also that in the case of finite dimensional setting (section 3.1), we can simply rewrite Proposition 3 as follows.

**Corollary 1** *Consider DRVI* (20)-(21). *Suppose:* (a) *for* $i=1,\cdots,r$ *and* $\xi\in\Xi^m$, $\Phi_i(\cdot,\xi)$ *and* $\phi_i(\cdot,\xi)$ *are continuously differentiable;* (b) *for any* $\tilde{\xi}\in(\Xi^m)^r$ *and* $x\in X$, $\left((J_x\Phi_1(x,\tilde{\xi}_1))^\top,\cdots,(J_x\Phi_r(x,\tilde{\xi}_r))^\top\right)^\top$ *is a positive semidefinite matrix;* (c) *for all* $\xi\in\Xi^m$, $x\in X$, $\tilde{x}\in\mathbb{R}^n$, $i=1,...,r$, $\tilde{x}_i^\top\Phi_i(x,\xi)-J_x\phi_i(x,\xi)\tilde{x}\geq 0$. *Then* $\mathcal{G}$ *corresponding to DRVI* (20)-(21) *is monotone over* $X\times\mathfrak{M}$.

In what follows, we give conditions for the existence of solutions of the DRVI based on the monotonicity properties.

**Definition 2** ([17, Definition 12.1]) The mapping $\mathcal{G}:\mathbb{R}^n\times\mathfrak{A}\to\mathbb{R}^n\times\mathcal{Z}$ is hemicontinuous on $\mathbb{R}^n\times\mathcal{Z}^*$ if $\mathcal{G}$ is continuous on line segments in $\mathbb{R}^n\times\mathcal{Z}^*$, i.e., for every pair of points $(x,\zeta),(z,\eta)\in\mathbb{R}^n\times\mathcal{Z}^*$, the following function is continuous

$$t\mapsto\left\langle\mathcal{G}(tx+(1-t)z,t\zeta+(1-t)\eta),\begin{pmatrix}x-z\\ \zeta-\eta\end{pmatrix}\right\rangle,\quad 0\leq t\leq 1.$$

**Definition 3** ([17, Definition 12.3 (i)]) The mapping $\mathcal{G}:X\times\mathcal{Z}^*\to X\times\mathcal{Z}$ is weakly coercive if there exists $(x_0,\zeta^0)\in\mathbb{R}^n\times\mathcal{Z}^*$ such that

$$\left\langle\mathcal{G}(x,\zeta),\begin{pmatrix}x-x_0\\ \zeta-\zeta^0\end{pmatrix}\right\rangle\to\infty\ \text{ as }\ \|x-x_0\|+\|\zeta-\zeta^0\|\to\infty\ \text{ and }(x,\zeta)\in X\times\mathcal{Z}^*.$$

**Theorem 1** *Suppose the conditions of Proposition 3 hold, $X \subseteq \mathbb{R}^n$ is a closed and convex set, $\mathfrak{A}$ is convex and weakly* compact in $\mathcal{Z}$ and $\mathcal{G}$ is weakly coercive. Then DRVI$(\mathcal{G},(X,\mathfrak{A}))$ has a solution.*

By Proposition 3, it is obvious that $\mathcal{G}$ is hemicontinuous and monotone on $\mathbb{R}^n \times \mathscr{P}$. Then Theorem 1 is from [17, Theorem 12.2 and Corollary 12.2] directly.

Moreover, for the finite dimensional case with coercivity condition, we have the following result directly from [16, Proposition 2.2.7].

**Corollary 2** *Suppose for $i = 1, \cdots, r$ and $\xi \in \Xi^m$, $\Phi_i(\cdot,\xi)$ and $\phi_i(\cdot,\xi)$ are continuous, $X \subseteq \mathbb{R}^n$ is a closed and convex set, $\mathfrak{M}$ is closed and $\mathcal{G}$ is coercive. Then DRVI$(\mathcal{G},(X,\mathfrak{M}))$ (20)-(21) has a solution.*

Note that, in contrast to Proposition 1, Corollary 2 does not require the compactness of $X$, but replaces it with coercivity. Note also that Corollary 2 does not need the monotonicity condition. However, monotonicity is still important for the algorithm design.

### 3.4 Examples of monotone DRVI

We illustrate the monotonicity property and coerciveness of $\mathcal{G}$ in the DRVI by two examples from the distributionally robust stochastic program and distributionally robust generalized Nash equilibrium.

*Example 5* Consider distributionally robust stochastic program (6) with $X = \mathbb{R}^2_+$, where $\phi$ is convex and twice continuously differentiable w.r.t. $x$, $\Xi := \{\xi^1, \xi^2\}$, $\mathfrak{M} \subset \{(p^1,p^2) : p^i \geq 0, p^1 + p^2 = 1, i = 1,2\}$ is convex and compact.

Then the corresponding DRVI is

$$0 \in p^1 \nabla_x \phi(x,\xi^1) + p^2 \nabla_x \phi(x,\xi^2) + \mathcal{N}_X(x), \tag{33}$$

$$0 \in \begin{pmatrix} -\phi(x,\xi^1) \\ -\phi(x,\xi^2) \end{pmatrix} + \mathcal{N}_{\mathfrak{M}}((p^1,p^2)). \tag{34}$$

And the corresponding function is

$$\mathcal{G}(x,P) = \begin{pmatrix} p^1 \nabla_x \phi(x,\xi^1) + p^2 \nabla_x \phi(x,\xi^2) \\ -\phi(x,\xi^1) \\ -\phi(x,\xi^2) \end{pmatrix}.$$

Moreover,

$$J\mathcal{G}_{(x,P)} = \begin{pmatrix} p^1\nabla_{xx}\phi(x,\xi^1) + p^2\nabla_{xx}\phi(x,\xi^2) & \nabla_x\phi(x,\xi^1) & \nabla_x\phi(x,\xi^2) \\ -\nabla_x\phi(x,\xi^1)^\top & 0 & 0 \\ -\nabla_x\phi(x,\xi^2)^\top & 0 & 0 \end{pmatrix}$$

is positive semidefinite over $X \times \mathfrak{M}$ and then $\mathcal{G}$ is monotone.

Then we prove the coercivity of $\mathcal{G}$. Let $x_0 = (0,0)$ and $P^0 = (1,0)$. Suppose for any $\xi \in \Xi$, $\phi(x,\xi)$ is a strongly convex function of $x$ with parameter $m(\xi^i) > 0$, $i = 1,2$. We have when $x$ sufficiently large, $\phi(x,\xi^i) \geq 0$, $i = 1,2$ and

$$\phi(0,\xi^i) \geq \phi(x,\xi^i) - \nabla_x\phi(x,\xi^i)^\top x + \frac{m(\xi^i)}{2}\|x\|_2^2.$$

Then

$$\begin{aligned}
&\liminf_{x\geq 0,\|x\|\to\infty} \frac{\left\langle \mathcal{G}(x,P), \begin{pmatrix} x - x_0 \\ P - P^0 \end{pmatrix} \right\rangle}{\|x\|} \\
&= \liminf_{x\geq 0,\|x\|\to\infty} \frac{x^\top(p^1\nabla_x\phi(x,\xi^1) + p^2\nabla_x\phi(x,\xi^2)) + (P-P^0)^\top \begin{pmatrix} -\phi(x,\xi^1) \\ -\phi(x,\xi^2) \end{pmatrix}}{\|x\|} \\
&= \liminf_{x\geq 0,\|x\|\to\infty} \frac{p_1(x^\top\nabla_x\phi(x,\xi^1) - \phi(x,\xi^1)) + p_2(x^\top\nabla_x\phi(x,\xi^2) - \phi(x,\xi^2)) + \phi(x,\xi^1)}{\|x\|} \\
&\geq \liminf_{x\geq 0,\|x\|\to\infty} \frac{p_1(x^\top\nabla_x\phi(x,\xi^1) - \phi(x,\xi^1)) + p_2(x^\top\nabla_x\phi(x,\xi^2) - \phi(x,\xi^2))}{\|x\|} \\
&\geq \liminf_{x\geq 0,\|x\|\to\infty} \frac{\sum_{i=1}^2 p^i(\frac{m(\xi^i)}{2}\|x\|^2 - \phi(0,\xi^i))}{\|x\|} > 0.
\end{aligned}$$

Combining the monotonicity and coercivity of $\mathcal{G}$, by Corollary 2, the DRVI has a solution.

We can also consider a distributionally robust generalized Nash equilibrium problem as follows.

*Example 6* Consider the distributionally robust generalized Nash equilibrium problem as follows:

$$\min_{x_i\in X_i} \max_{P_i\in\mathfrak{M}_i} \mathbb{E}_{P_i}[f_i(x,\xi)] + g_i(x), \quad \text{s.t. } b_1x_1 + b_2x_2 \leq c, \quad i = 1,2, \tag{35}$$

where $x = (x_1,x_2)$, $x_i \in X_i = \mathbb{R}_+$, for $P_i$-a.e. $\xi$, $\forall P_i \in \mathfrak{M}_i$ $f_i(\cdot,\xi)$ is convex and twice continuously differentiable with respect to $x$, and $g_i$ is convex and twice continuously differentiable, $\Xi := \{\xi^1,\xi^2\}$, $\mathfrak{M}_i \subset \{(p_i^1,p_i^2) : p_i^j \geq 0, p_i^1 + p_i^2 =$

$1, j = 1, 2\}$ is convex and compact, $i = 1, 2$. Suppose $J\left(\nabla_{x_1} g_1(x), \nabla_{x_2} g_2(x)\right)^\top$ is positive semidefinite.

Then the corresponding DRVI is

$$0 \in \sum_{j=1}^{2} p_i^j \nabla_{x_i} f_i(x, \xi^j) + \nabla_{x_i} g_i(x) + b_i \mu + \mathcal{N}_{X_i}(x_i), \quad i = 1, 2 \tag{36}$$

$$0 \in c - b_1 x_1 - b_2 x_2 + \mathcal{N}_{\mathbb{R}_+}(\mu), \tag{37}$$

$$0 \in \begin{pmatrix} -f_i(x, \xi^1) \\ -f_i(x, \xi^2) \end{pmatrix} + \mathcal{N}_{\mathfrak{M}_i}((p_i^1, p_i^2)), \qquad i = 1, 2. \tag{38}$$

Let

$$\Phi(x, \xi) = \begin{pmatrix} \nabla_{x_1} f_1(x_1, \xi) + \nabla_{x_1} g_1(x) + b_1 \mu \\ \nabla_{x_2} f_2(x_2, \xi) + \nabla_{x_2} g_2(x) + b_2 \mu \end{pmatrix} \quad \text{and} \quad \phi(x, \xi) = \begin{pmatrix} -f_1(x, \xi) \\ -f_2(x, \xi) \end{pmatrix}.$$

Then the DRVI (36)-(38) is corresponding to (20)-(21) with (37). Moreover,

$$\mathcal{G}(x, \mu, P) = \begin{pmatrix} p_1^1 \nabla_{x_1} f_1(x, \xi^1) + p_1^2 \nabla_{x_1} f_1(x, \xi^2) + \nabla_{x_1} g_1(x) + b_1 \mu \\ p_2^1 \nabla_{x_2} f_2(x, \xi^1) + p_2^2 \nabla_{x_2} f_2(x, \xi^2) + \nabla_{x_2} g_2(x) + b_2 \mu \\ c - b_1 x_1 - b_2 x_2 \\ -f_1(x, \xi^1) \\ -f_1(x, \xi^2) \\ -f_2(x, \xi^1) \\ -f_2(x, \xi^2) \end{pmatrix}.$$

For $i, j = 1, 2$, let

$$a_{ij} = p_i^1 \nabla_{x_i x_j} f_i(x_i, \xi^1) + p_i^2 \nabla_{x_i x_j} f_i(x_i, \xi^2) + \nabla_{x_i x_j} g_i(x).$$

Then $J\mathcal{G}_{(x,\mu,P)}$ is

$$\begin{pmatrix} a_{11} & a_{12} & b_1 & \nabla_{x_1} f_1(x, \xi^1) & \nabla_{x_1} f_1(x, \xi^2) & 0 & 0 \\ a_{21} & a_{22} & b_2 & 0 & 0 & \nabla_{x_2} f_2(x, \xi^1) & \nabla_{x_2} f_2(x, \xi^2) \\ -b_1 & -b_2 & 0 & 0 & 0 & 0 & 0 \\ -\nabla_{x_1} f_1(x, \xi^1)^\top & -\nabla_{x_2} f_1(x, \xi^1)^\top & 0 & 0 & 0 & 0 & 0 \\ -\nabla_{x_1} f_1(x, \xi^2)^\top & -\nabla_{x_2} f_1(x, \xi^2)^\top & 0 & 0 & 0 & 0 & 0 \\ -\nabla_{x_1} f_2(x, \xi^1)^\top & -\nabla_{x_2} f_2(x, \xi^1)^\top & 0 & 0 & 0 & 0 & 0 \\ -\nabla_{x_1} f_2(x, \xi^2)^\top & -\nabla_{x_2} f_2(x, \xi^2)^\top & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

It is obvious that in general, $\mathcal{G}$ is nonmonotone. However, $\mathcal{G}$ can be monotone if $f_1$ is only w.r.t. $(x_1, \xi)$ and $f_2$ is only w.r.t. $(x_2, \xi)$, and

$$\begin{pmatrix} \nabla_{x_1x_1} f_1(x_1, \xi^i) + \nabla_{x_1x_1} g_1(x) & \nabla_{x_1x_2} g_1(x) \\ \nabla_{x_2x_1} g_2(x) & \nabla_{x_2x_2} f_2(x_2, \xi^j) + \nabla_{x_2x_2} g_2(x) \end{pmatrix}$$

is positive semidefinite for $i, j = 1, 2$.

Then we show the coerciveness of $\mathcal{G}$. Similar as in Example 5, let $x_{j0} = 0$, $\mu_0 = 0$, $P_j^0 = (1, 0)$ for $j = 1, 2$. Suppose in (35), $c > 0$, $f_j(x, \xi^i)$ and $g_j(x)$ are strongly convex with parameter $m_j(\xi^i) > 0$ and $m_j^g > 0$ respectively, $i = 1, 2$ and $j = 1, 2$. Then

$$\begin{aligned}
&\liminf_{(x,\mu)\geq 0, \|(x,\mu)\|\to\infty} \frac{\left\langle \mathcal{G}(x,\mu,P), \left(x - x_0, \mu - \mu_0, P - P^0\right)^\top \right\rangle}{\|(x,\mu,P)\|} \\
&\geq \liminf_{(x,\mu)\geq 0, \|(x,\mu)\|\to\infty} \frac{\sum_{i=1}^{2}\sum_{j=1}^{2} \left(p_j^i(x_j \nabla_{x_j} f_j(x_j,\xi^i) + x_j \nabla_{x_j} g_j(x) - f_j(x_j,\xi^i)) + p_{j0}^i f_j(x_j,\xi^i)\right) + \mu c}{\|(x,\mu)\|} \\
&\geq \liminf_{(x,\mu)\geq 0, \|(x,\mu)\|\to\infty} \frac{\sum_{i=1}^{2}\sum_{j=1}^{2} p_j^i(x_j \nabla_{x_j} f_j(x_j,\xi^i) - f_j(x_j,\xi^i) + x_j \nabla_{x_j} g_j(x)) + \mu c}{\|(x,\mu)\|} \\
&\geq \liminf_{(x,\mu)\geq 0, \|(x,\mu)\|\to\infty} \frac{\sum_{i=1}^{2}\sum_{j=1}^{2} p_j^i(\frac{m_j(\xi^i)}{2}\|x_j\|^2 - f_j(0,\xi^i) + \frac{m_j^g}{2}\|x_j\|^2 - g_j(0)) + \mu c}{\|(x,\mu)\|} > 0.
\end{aligned}$$

Combining the monotonicity and coerciveness of $\mathcal{G}$, by Corollary 2, the DRVI has a solution.

## 4 Discretization of probability distributions

In this section, we consider the discretization of DRVI with the ambiguity sets formed from continuous distributions in the setting specified in Assumption 1. There are several ways to discretize the ambiguity set [11,28,33]. We propose the SAA approach to the DRVI. For the sake of simplicity we assume here that $r = 1$ and drop the subscript $i$ in $\Phi_i^x$ and $\phi_i^x$, etc. An extension for $r > 1$ will be straightforward. Recall that $\mathbb{P}$ is the reference probability measure (distribution) on $(\Xi, \mathcal{B})$, $\mathcal{Z} = L_p(\Xi, \mathcal{B}, \mathbb{P})$, $\mathcal{Z}^* = L_q(\Xi, \mathcal{B}, \mathbb{P})$, $\mathfrak{A}$ is a convex bounded weakly* closed subset of $\mathcal{Z}^*$ of densities associated with the ambiguity set $\mathfrak{M}$, and

$$\mathcal{R}(Z) = \sup_{\zeta \in \mathfrak{A}} \int_{\Xi} Z(s)\zeta(s) d\mathbb{P}(s), \; Z \in \mathcal{Z}. \tag{39}$$

Let us introduce some definitions.

It is said that random variables $Y, Y' : \Xi \to \mathbb{R}$ are distributionally equivalent (with respect to $\mathbb{P}$), denoted $Y \overset{\mathcal{D}}{\sim} Y'$, if $\mathbb{P}(Y \le y) = \mathbb{P}(Y' \le y)$ for all $y \in \mathbb{R}$. It is said that a functional $\mathcal{R} : \mathcal{Z} \to \mathbb{R}$ is *law invariant* if $\mathcal{R}(Z) = \mathcal{R}(Z')$ for any distributionally equivalent $Z, Z' \in \mathcal{Z}$. The set $\mathfrak{A} \subset \mathcal{Z}^*$ is said to be law invariant if $\zeta \in \mathfrak{A}$ and $\zeta' \overset{\mathcal{D}}{\sim} \zeta$, then $\zeta' \in \mathfrak{A}$. It is known that the functional $\mathcal{R}$ is law invariant iff the corresponding set $\mathfrak{A}$ is law invariant (cf., [28, Theorem 2.3]).

**Assumption 2** *The set $\mathfrak{A}$ is law invariant.*

By Assumption 2 we have that the functional $\mathcal{R}(Z)$ is law invariant, and hence can be viewed as a function of the respective cumulative distribution function (CDF) of $Z$. It is possible to proceed with the required discretization by making discretization of the corresponding CDF of $\phi^x(\xi)$. However, such approach is indirect and inconvenient for applications. Therefore we discuss below several important cases where this can be performed in a rather straightforward way.

Consider an iid sample $\xi^j \in \Xi$, $j = 1, ..., N$, from the reference distribution $\mathbb{P}$. The law invariant risk measure $\mathcal{R}$ is associated with the corresponding empirical functional[7] $\hat{\mathcal{R}}_N : \mathbb{R}^N \to \mathbb{R}$. The functional $\mathcal{R}_N$ has the dual representation

$$\mathcal{R}_N(Z) := \sup_{\zeta \in \mathfrak{A}^N} N^{-1} \sum_{j=1}^{N} \zeta_j Z(\xi^j), \tag{40}$$

where $\mathfrak{A}^N$ is the respective convex closed set of densities[8] $\zeta = (\zeta_1, ..., \zeta_N)$. In the next section we give examples of how the empirical functional can be constructed.

For the generated sample,

$$\hat{\mathcal{R}}_N(\phi^x) = \sup_{\zeta \in \mathfrak{A}^N} N^{-1} \sum_{j=1}^{N} \zeta_j \phi^x(\xi^j) \tag{41}$$

can be considered as an empirical estimate of $\mathcal{R}(\phi^x)$. We have that under mild regularity conditions, $\hat{\mathcal{R}}_N(\phi^x)$ epiconverges w.p.1 to $\mathcal{R}(\phi^x)$ on $X$ (cf., [26]).

[7] Any $Z : \{\xi^1, ..., \xi^N\} \to \mathbb{R}$ can be identified with $N$-dimensional vector $(Z(\xi_1), ..., Z(\xi_N))$, and hence the empirical risk measure can be viewed as defined on $\mathbb{R}^N$.

[8] Note that $\zeta$ is a density on $\{\xi^1, ..., \xi^N\}$ if $\zeta \ge 0$ and $N^{-1} \sum_{i=1}^{N} \zeta_i = 1$, i.e., $N^{-1}\zeta \in \Delta_N$.

This suggests the following discretization of problem (22) - (23) (with $r = 1$):

$$0 \in \sum_{j=1}^{N} \zeta_j \Phi(x, \xi^j) + \mathcal{N}_X(x), \tag{42}$$

$$\zeta \in \arg\max_{\eta \in \mathfrak{A}^N} \sum_{j=1}^{N} \eta_j \phi(x, \xi^j). \tag{43}$$

## 4.1 Construction of the empirical estimates

Here we discuss construction of the empirical estimates of the risk measure $\mathcal{R}$ defined in (39). For a random variable $Z$ we denote by $H_Z(z) := \mathbb{P}(Z \le z)$ its CDF and by $H_Z^{-1}(t) := \inf\{\tau : H_Z(\tau) \ge t\}$ the corresponding quantile function (also called Value-at-Risk). Note that the subdifferential of $\mathcal{R}(Z)$ is given by

$$\partial \mathcal{R}(Z) = \arg\max_{\zeta \in \mathfrak{A}} \int_{\Xi} Z(s)\zeta(s) d\mathbb{P}(s) \tag{44}$$

(e.g., [27, eq. (6.49), p. 284]). Let us first consider spectral risk measure $\mathcal{R}$.

**Spectral risk measure.** Spectral risk measure is[9]

$$\mathcal{R}(H_Z^{-1}) := \int_0^1 \sigma(t) H_Z^{-1}(t) dt, \tag{45}$$

where $\sigma : [0,1) \to \mathbb{R}_+$ is monotonically nondecreasing, and left side continuous function such that $\int_0^1 \sigma(t)dt = 1$. The Average Value-at-Risk $\mathsf{AV@R}_\alpha$ is a spectral risk measure with spectral function $\sigma(t) = 0$ for $t \in [0, 1-\alpha)$, and $\sigma(t) = 1/\alpha$ for $t \in [1-\alpha, 1]$, see Example 8 in the Appendix

Let $H_{\phi^x}(z) := \mathbb{P}\{\phi^x(\xi) \le z\}$ be the CDF of $\phi^x(\xi)$ and $H_{\phi^x,N}$ be the CDF of $\phi^x(\xi^j)$, $j = 1, .., N$. That is, function $H_{\phi^x,N}(\cdot)$ is stepwise constant with jumps $1/N$ at points $\phi^x_{(1)}, ..., \phi^x_{(N)}$, where $\phi^x_{(1)}, ..., \phi^x_{(N)}$ are values $\phi^x(\xi^1), ..., \phi^x(\xi^N)$ arranged in the increasing order, i.e.,

$$H_{\phi^x,N}(\cdot) = N^{-1} \sum_{j=1}^{N} \mathbf{1}_{(-\infty, \phi^x_{(j)}]}(\cdot). \tag{46}$$

Then

$$\mathcal{R}(H^{-1}_{\phi^x,N}) = \int_0^1 \sigma(t) H^{-1}_{\phi^x,N}(t) dt = \sum_{j=1}^{N} q_j \phi^x_{(j)}, \tag{47}$$

[9] By the law invariance of $\mathcal{R}(Z)$ it can be considered as a function of $H_Z$.

where

$$q_j := \int_{(j-1)/N}^{j/N} \sigma(t)dt,\ j = 1, ...., N. \tag{48}$$

Note that $q_j \geq 0$, $\sum_{j=1}^N q_j = \int_0^1 \sigma(t)dt = 1$.

*Remark 6* The corresponding set $\mathfrak{A}^N$ is the convex hull of vectors $(q_{\pi(1)}, ..., q_{\pi(N)})$, $\pi \in \Pi$, where $\Pi$ is the set of permutations of the set $\{1, ..., N\}$. By Hardy - Littlewood inequality, we have here

$$\sup_{\zeta \in \mathfrak{A}^N} \sum_{j=1}^N \zeta_j \phi^x(\xi^j) = \sum_{j=1}^N q_j \phi^x_{(j)}, \tag{49}$$

and the corresponding maximizer $\bar{\zeta} \in \arg\max_{\zeta \in \mathfrak{A}^N} \sum_{j=1}^N \zeta_j \phi^x(\xi^j)$ is given by $\bar{\zeta} = (q_{\pi(1)}, ..., q_{\pi(N)})$ with the permutation $\pi \in \Pi$ corresponding to the order $\phi^x_{(1)} \leq \cdots \leq \phi^x_{(N)}$. Note that this permutation and hence the maximizer $\bar{\zeta}$ depend on $x$.

We can also write this spectral risk measure in the form

$$\mathcal{R}(\phi^x) = \int_0^1 \mathsf{AV@R}_{1-\alpha}(\phi^x) d\mu(\alpha), \tag{50}$$

where $\mu$ is the probability measure on the interval $[0, 1)$ associated with the spectral function $\sigma(\cdot)$, given by

$$\mu(\alpha) = (1 - \alpha)\sigma(\alpha) + \int_0^\alpha \sigma(t)dt.$$

This is the so-called Kusuoka representation of the spectral risk measure (e.g., [27, p. 307]). That is

$$\mathcal{R}(\phi^x) = \int_0^1 \mathbb{E}_{\mathbb{P}}\big\{\tau(\alpha) + (1 - \alpha)^{-1}[\phi^x - \tau(\alpha)]_+\big\} d\mu(\alpha), \tag{51}$$

where $\tau(\alpha) := H^{-1}_{\phi^x}(\alpha)$. The empirical estimate $\mathcal{R}(H^{-1}_{\phi^x, N})$ can be written then as

$$\mathcal{R}(H^{-1}_{\phi^x,N}) = \frac{1}{N} \int_0^1 \sum_{j=1}^N \Big\{\hat{\tau}_N(\alpha) + (1 - \alpha)^{-1}[\phi^x(\xi^j) - \hat{\tau}_N(\alpha)]_+\Big\} d\mu(\alpha), \tag{52}$$

where $\hat{\tau}_N(\alpha)$ is the empirical estimate of $H^{-1}_{\phi^x}(\alpha)$.

The subdifferential of $\mathcal{R}(\phi^x)$ can be taken inside the integral in (50), i.e.,

$$\partial\mathcal{R}(\phi^x) = \int_0^1 \partial\mathsf{AV@R}_{1-\alpha}(\phi^x) d\mu(\alpha). \tag{53}$$

We have that $\partial\mathcal{R}(\phi^x) = \{\bar{\zeta}^x\}$ is a singleton iff $\partial\mathsf{AV@R}_{1-\alpha}(\phi^x)$ is a singleton for $\mu$-almost every $\alpha \in [0,1)$, i.e., iff $\mathbb{P}\{\phi^x = \kappa_\alpha\} = 0$ for $\mu$-almost every $\alpha \in [0,1)$, where $\kappa_\alpha = H^{-1}_{\phi^x}(\alpha)$. Then we have by Example 8 that the subdifferential $\partial\mathsf{AV@R}_{1-\alpha}(\phi^x) = \{\bar{\zeta}_\alpha\}$ is given by

$$\bar{\zeta}_\alpha(s) = \begin{cases} (1-\alpha)^{-1} \; if \; \phi^x(s) > \kappa_\alpha, \; s \in \Xi, \\ \quad\quad 0 \quad\quad\; if \; \phi^x(s) < \kappa_\alpha, \; s \in \Xi. \end{cases} \tag{54}$$

The subdifferential $\hat{\zeta}^x_\alpha = (\zeta^x_{\alpha 1}, ..., \zeta^x_{\alpha N})$ of the corresponding empirical estimate is obtained by replacing $\kappa_\alpha = H^{-1}_{\phi^x}(\alpha)$ with their empirical estimates.

For a continuous and bounded function $g : \Xi \to \mathbb{R}$ we have that

$$\int_\Xi g(s)dP^x_N(s) = \int_{\alpha\in[0,1)} \frac{1}{N}\sum_{j=1}^N g(\xi^j)\zeta^x_{\alpha j}d\mu(\alpha) = \int_{\alpha\in[0,1)} \frac{1}{(1-\alpha)N}\sum_{\phi^x(\xi^j)>\kappa^\alpha_N} g(\xi^j)d\mu(\alpha)$$

converges w.p.1 to

$$\int_\Xi g(s)\bar{\zeta}^x(s)d\mathbb{P}(s) = \int_{\alpha\in[0,1)}\int_\Xi g(s)\bar{\zeta}^x_\alpha(s)d\mathbb{P}(s)d\mu(\alpha) = \int_{\alpha\in[0,1)} \frac{1}{1-\alpha}\int_{\phi^x(\xi)>\kappa_\alpha} g(s)d\mathbb{P}(s)d\mu(\alpha).$$

Then we have that $P^x_N$ converges weakly to $P^x$, where $P^x$ has density $\bar{\zeta}^x$ (see (A4)). Moreover, by Proposition 6 in the Appendix, we have if $\{x_N\}$ is a sequence in $X$ converging to $x$, then $\int_\Xi g(s)dP^{x_N}_N(s)$ converges to $\int_{\phi^x(\xi)>\kappa} g(s)d\mathbb{P}(s)$ w.p.1, and hence $P^{x_N}_N$ converges weakly to $P^x$ as $N \to \infty$.

**Law invariant coherent risk measure.** By dual representation, any law invariant coherent risk measure can be represented as follows

$$\max_{\zeta\in\mathfrak{A}}\int_\Xi Z(s)\zeta(s)d\mathbb{P}(s) = \mathcal{R}(Z) = \mathcal{R}(H^{-1}_Z) = \sup_{\sigma\in\mathfrak{S}}\int_0^1 \sigma(t)H^{-1}_{\phi^x}(t)dt \tag{55}$$

with

$$\mathfrak{S} := \{\sigma = H^{-1}_\zeta : \zeta \in \mathfrak{A}\} \tag{56}$$

being a set of spectral functions.

Let $H_{\phi^x,N}$ denote the CDF of the empirical distribution corresponding to the i.i.d. samples $\{\phi^x(\xi^1), \cdots, \phi^x(\xi^N)\}$. Note that $H_{\phi^x,N}$ is a function of the random sample, and hence is random. We have

$$\partial\mathcal{R}(H^{-1}_{\phi^x}) = \arg\max_{\sigma\in\mathfrak{S}}\int_0^1 \sigma(t)H^{-1}_{\phi^x}(t)dt \text{ and } \partial\mathcal{R}\left(H^{-1}_{\phi^x,N}\right) = \arg\max_{\sigma\in\mathfrak{S}}\int_0^1 \sigma(t)H^{-1}_{\phi^x,N}(t)dt. \tag{57}$$

**Lemma 2** *Consider a point $\bar{x} \in X$ and a sequence $\{x_N\} \subset X$ converging to $\bar{x}$. Suppose that Assumptions 1 and 2 hold, $\phi(\cdot, \xi)$ is continuous and $\partial \mathcal{R}(H_{\phi^{\bar{x}}}^{-1}) = \{\bar{\sigma}\}$ is a singleton. Then any sequence $\sigma_N \in \partial \mathcal{R}\left(H_{\phi^{x_N},N}^{-1}\right)$ weakly* converges to $\bar{\sigma}$ w.p.1.*

*Proof* We first note that $H_{\phi^x}^{-1}$ and $H_{\phi^{x_N},N}^{-1}$ belong to the space $\mathcal{L}_p$. We can apply a general theory of sensitivity analysis applied to the optimization problem (57) with viewing $H_{\phi^x}^{-1}$ as parameter in the space $\mathcal{L}_p$. We have that $H_{\phi^{x_N},N}^{-1}$ converges w.p.1 to $H_{\phi^x}^{-1}$ in the norm topology of $\mathcal{L}_p$ as $N \to \infty$. This can be proved by an extension of [26, Theorem 2.1] (see Theorem 3 in the Appendix). Since the set $\mathfrak{S}$ is weakly* compact and the maximizer $\bar{\sigma}$ of the right hand side of (57) is unique, it follows by [5, Lemma 4.3 and example 4.5] that if

$$\sigma_N \in \arg\max_{\sigma \in \mathfrak{S}} \int_0^1 \sigma(t) H_{\phi^{x_N},N}^{-1}(t) dt, \tag{58}$$

then $\{\sigma_N\}$ is weak* convergent w.p.1 to $\bar{\sigma}$. $\square$

For law invariant coherent risk measure, by the Kusuoka representation, (55) can also be presented as

$$\mathcal{R}(H_{\phi^x}^{-1}) = \sup_{\sigma \in \mathfrak{S}} \int_0^1 \sigma(t) H_{\phi^x}^{-1}(t) dt = \sup_{\mu \in \mathfrak{V}} \int_0^1 \mathrm{AVaR}_{1-\alpha}(\phi^x) d\mu(\alpha) \tag{59}$$

and its SAA can be written as

$$\mathcal{R}(H_{\phi^x,N}^{-1}) = \sup_{\mu \in \mathfrak{V}} \frac{1}{N} \int_0^1 \sum_{j=1}^N \Big\{ \hat{\tau}_N(\alpha) + (1-\alpha)^{-1} [\phi^x(\xi^j) - \hat{\tau}_N(\alpha)]_+ \Big\} d\mu(\alpha), \tag{60}$$

where $\mathfrak{V} := \{\mu : \mu(\alpha) = (1-\alpha)\sigma(\alpha) + \int_0^\alpha \sigma(t)dt, \sigma \in \mathfrak{S}\}$. Then we have that $\partial \mathcal{R}(\phi^x) = \{\bar{\zeta}^x\}$ is a singleton (which implies that $\partial \mathcal{R}(H_{\phi^{\bar{x}}}^{-1})$ is a singleton) if and only if $\partial \mathsf{AV@R}_{1-\alpha}(\phi^x)$ is a singleton for $\mu$-almost every $\alpha \in [0, 1)$, i.e., if and only if $\mathbb{P}\{\phi^x = \kappa_\alpha\} = 0$ for $\mu$-almost every $\alpha \in [0, 1)$, where $\kappa_\alpha = H_{\phi^x}^{-1}(\alpha)$, $\mu \in \mathfrak{V}$. Note that if $\mathbb{P}\{\phi^x = \kappa_\alpha\} = 0$ for every $\alpha \in [0, 1)$, then the condition that $\partial \mathcal{R}(H_{\phi^{\bar{x}}}^{-1}) = \{\bar{\sigma}\}$ is a singleton holds.

Then we consider the convergence analysis between

$$\max_{\eta \in \mathfrak{A}^N} \sum_{j=1}^N \eta_j \phi(x_N, \xi^j) = \sup_{\mu \in \mathfrak{V}} \frac{1}{N} \int_0^1 \sum_{j=1}^N \Big\{ \hat{\tau}_N(\alpha) + (1-\alpha)^{-1} [\phi^{x_N}(\xi^j) - \hat{\tau}_N(\alpha)]_+ \Big\} d\mu(\alpha) \tag{61}$$

and

$$\max_{\eta\in\mathfrak{A}}\int_{\Xi}\phi^{\bar{x}}\eta d\mathbb{P}=\sup_{\mu\in\mathfrak{V}}\int_0^1 \mathrm{AVaR}_{1-\alpha}(\phi^{\bar{x}})d\mu(\alpha), \tag{62}$$

where $x_N\to\bar{x}$. Let $(\zeta_N^{x_N},\mu_N)$ and $(\zeta^*,\bar{\mu})$ denote the optimal solutions of (61) and (62), respectively. Note that $\zeta_N^{x_N}$ corresponds to a discrete distribution $P_N^{x_N}$ and $\zeta^*$ corresponds to a continuous distribution $P^*=\zeta^*\mathbb{P}$.

**Proposition 4** *Consider a point $\bar{x}\in X$ and a sequence $\{x_N\}\subset X$ converging to $\bar{x}$ and the ambiguity set corresponding to a law invariant coherent risk measure. Suppose (i) Assumptions 1 and 2 hold, and $\phi(\cdot,\xi)$ is Lipschitz continuous; (ii) $\partial\mathcal{R}(H^{-1}_{\phi^{\bar{x}}})=\{\bar{\sigma}\}$ is a singleton; (iii) the CDF of $\phi^{\bar{x}}$ is strictly monotone, and (iv) there exists positive measure $\hat{\mu}$ such that for all $N$ sufficiently large, $\int_{[0,1]}h(t)\hat{\mu}(t)\geq\int_{[0,1]}h(t)\mu_N(t)$ for all bounded function $h(t)$. Then $P_N^{x_N}$ converges weakly to $P^{\bar{x}}$.*

The proof of Proposition 4 is in the Appendix.

*Example 7* Consider the $\psi$-divergence approach to construction of the uncertain sets. The concept of $\psi$-divergence is originated in Csiszár [14] and Morimoto [19], and was extensively discussed in Ben-Tal and Teboulle [2]. We also can refer to Bayraksan and Love [1] for a recent survey of this approach. That is, consider a convex lower semicontinuous function $\psi:\mathbb{R}\to\mathbb{R}_+\cup\{+\infty\}$ such that $\psi(1)=0$. For $x<0$ we set $\psi(x)=+\infty$. For $c>0$ consider

$$\mathfrak{A}:=\left\{\zeta\in\mathfrak{D}:\textstyle\int_{\Xi}\psi(\zeta(s))d\mathbb{P}(s)\leq c\right\}, \tag{63}$$

where $\mathfrak{D}:=\{\zeta\in\mathcal{Z}^*:\int\zeta d\mathbb{P}=1,\ \zeta\succeq 0\}$ denotes the set of densities. If $\zeta\overset{\mathcal{D}}{\sim}\zeta'$, then $\int_{\Xi}\psi(\zeta(s))d\mathbb{P}(s)=\int_{\Xi}\psi(\zeta'(s))d\mathbb{P}(s)$. Hence the set $\mathfrak{A}$ and the corresponding functional $\mathcal{R}$ are law invariant. Since $\psi$-divergence is a law invariant coherent risk measure, it has Kusuoka representation (Note that the representation is only for constructing the SAA and proving the weak convergence).

**Proposition 5** *[15, Proposition 5.6] A $\psi$-divergence risk measure can be written in the form*

$$\mathcal{R}(Z)=\sup_{\sigma\in\mathfrak{S}}\int_0^1\sigma(t)H_Z^{-1}(t)dt, \tag{64}$$

*where*

$$\mathfrak{S}:=\left\{\sigma:[0,1]\to[0,\infty]:\sigma\text{ is non-decreasing},\textstyle\int_0^1\sigma(t)dt=1,\int_0^1\psi(\sigma(t))dt\leq c\right\}.$$

Moreover, let $\mu(\alpha)=(1-\alpha)\sigma(\alpha)+\int_0^\alpha \sigma(t)dt$ (that is $\sigma(t)=\int_0^t \frac{1}{1-\alpha}d\mu(\alpha)$). By Kusuoka representation, we have

$$\mathcal{R}(Z)=\sup_{\mu\in\mathfrak{V}}\int_0^1 \mathsf{AV@R}_{1-\alpha}(Z)d\mu(\alpha)=\sup_{\mu\in\mathfrak{V}}\int_0^1 \mathbb{E}_{\mathbb{P}}\{\tau(\alpha)+(1-\alpha)^{-1}[Z-\tau(\alpha)]_+\}d\mu(\alpha), \tag{65}$$

where

$$\mathfrak{V}:=\left\{\mu:[0,1)\to[0,\infty]:\int_0^1 d\mu(\alpha)=1,\int_0^1\psi\left(\int_0^t\frac{1}{1-\alpha}d\mu(\alpha)\right)dt\le c\right\}.$$

Although the structure of $\mathfrak{V}$ looks complicated, the discretization way is exactly SAA and same as in the paper. Then with the conditions of Proposition 4, we can show that $P_N^{x_N}$ converges weakly to $P^{\bar{x}}$.

By discussion above, we have shown that for law invariant coherent risk measure and under mild conditions, $P_N^{x_N}$ converges weakly to $P^{\bar{x}}$.

However, to prove the convergence between (22)-(23) and (42)-(40), we need stronger convergence results between $P_N^{x_N}$ and $P^{\bar{x}}$. To this end, we need the following assumption.

**Assumption 3** *Let $\mathfrak{M}$ and $\mathfrak{M}^N$ be nonempty and closed. Assume:* (a) *there exists a weakly compact set $\hat{\mathfrak{M}}$ such that $\mathfrak{M},\mathfrak{M}^N\subset\hat{\mathfrak{M}}$ holds for $N$ sufficiently large;* (b) $\sup_{P\in\hat{\mathfrak{M}}}\mathbb{E}_P[\|\xi\|]$ *is bounded.*

Note that the ambiguity set $\mathfrak{M}$ is defined in Definition 1 and $\mathfrak{M}^N$ is the corresponding set of discrete probability measures of $\mathfrak{A}^N$ in (40). One sufficient condition for Assumption 3 is the compactness of support set $\Xi$. However, by Prokhorov's theorem, a closed ambiguity set of probability measures is *compact* (under the weak topology) if it is tight, and the support set is not necessarily compact. For more discussion on weak compactness, see [30] and section 5 in [3]. Now we present the main convergence result.

**Theorem 2** *Let $(\hat{x}_N,\hat{\zeta}_N)\in X\times\mathfrak{A}^N$ be a solution of the SAA variational inequalities* (42) - (40)*. Suppose:* (a) *Assumptions 1, 2 and 3 hold;* (b) $\phi(x,\cdot)$ *is Lipschitz continuous on $\Xi$, $\Phi(\cdot,\xi)$ and $\Phi(x,\cdot)$ are Lipschitz continuous with Lipschitz modulus $\kappa(\xi)$ and $\bar{\kappa}$ over $X$ and $\Xi$ respectively, and* $\sup_{P\in\hat{\mathfrak{M}}}\mathbb{E}_P[\kappa(\xi)]<\infty$*;* (c) *$\hat{x}_N$ converges w.p.1 to a point $\bar{x}$;* (d) *$\partial\mathcal{R}(\phi^{\bar{x}})=\{\bar{\zeta}\}$ is a singleton;* (e) *$P^{\bar{x}}$ is probability measure on $(\Xi,\mathcal{B})$ with density $\bar{\zeta}$, and $P_N^{\hat{x}_N}$ is the empirical measure associated with $\hat{\zeta}_N$, and $P_N^{\hat{x}_N}$ weakly converges to $P^{\bar{x}}$. Then $(\bar{x},\bar{\zeta})$ is a solution of the DRVI* (22) - (23)*.*

The proof of Theorem 2 is in the Appendix. Note that the sufficient conditions for assumption (e) are given in Proposition 4.

## 5 Numerical examples

In this section, we use a continuous version of Example 6 of the distributionally robust generalized Nash equilibrium problem to illustrate the SAA approach and its convergence, where $f_i$ and $g_i$ are quadratic convex functions, $\mathfrak{M}_i$ is constructed by modified $\chi^2$-distance, $i = 1, 2$. Particularly, let

$$f_i(x,\xi) := \frac{1}{2} x_i^\top \tilde{M}_i(\xi) x_i + \tilde{c}_i(\xi)^\top x_i, \qquad g_i(x) := \frac{1}{2} x_i^\top M_i x_i + c_i^\top x_i + x_i^\top R_i x_{-i},$$

$X_i = \mathbb{R}^2_+$, $b_1 = b_2 = (1,1)^\top$ and $c = 10$. Let $\mathbb{P}$ follow the uniform distribution over $[-1,1]$, and $\xi$ is a random variable with support set $[-1,1]$. Then the density function of $\mathbb{P}$ is a constant function with value $\frac{1}{2}$ over $[-1,1]$ and the ambiguity set

$$\mathfrak{M}_i := \left\{ P \in \mathscr{P} : \int_{\xi \in [-1,1]} 2(p(\xi) - 1)^2 d\xi \leq 0.05 \right\},$$

where $\mathscr{P}$ denotes all probability measures over $[-1,1]$, $p(\xi)$ is the density function of $P$, $i = 1, 2$. Note that this is a particular case of $\psi$-divergence. It is obvious that $\mathfrak{M}_i$ is a weakly compact subset in $\mathcal{L}_2$ over $[-1,1]$. Let $E$ be the $2 \times 2$ matrix with all elements 1, $R_i = E$, $\tilde{M}_i(\xi) = 5I + \xi I$ and $M_i = I$. Then for any $\xi^1, \xi^2 \in [0,1]$,

$$\begin{pmatrix} \nabla_{x_1 x_1} f_1(x,\xi^i) + \nabla_{x_1 x_1} g_1(x) & \nabla_{x_1 x_2} f_1(x,\xi^i) + \nabla_{x_1 x_2} g_1(x) \\ \nabla_{x_2 x_1} f_2(x,\xi^j) + \nabla_{x_2 x_1} g_2(x) & \nabla_{x_2 x_2} f_2(x,\xi^j) + \nabla_{x_2 x_2} g_2(x) \end{pmatrix} = \begin{pmatrix} \tilde{M}_1(\xi^1) + M_1 & R_1 \\ R_2 & \tilde{M}_2(\xi^2) + M_2 \end{pmatrix}$$

is positive definite and then for any $\xi^1, \xi^2 \in [0,1]$,

$$\begin{pmatrix} \tilde{M}_1(\xi^1) + M_1 & R_1 & b_1 \\ R_2 & \tilde{M}_2(\xi^2) + M_2 & b_2 \\ -b_1^\top & -b_2^\top & 0 \end{pmatrix}$$

is positive semidefinite. Moreover, for any $\xi^1, \xi^2 \in [0,1]$, $\tilde{x}_1^\top \left( \tilde{M}_1(\xi^1) x_1 + \tilde{c}_1(\xi^1) \right) - J_x(f_1(x,\xi^1))\tilde{x} = 0$, and $\tilde{x}_2^\top \left( \tilde{M}_2(\xi^2) x_2 + \tilde{c}_2(\xi^2) \right) - J_x(f_2(x,\xi^2))\tilde{x} = 0$. Let $z := (x, \mu)$. Then by Proposition 3, the corresponding

$$\mathcal{G}(z,\zeta) := \begin{pmatrix} \int_\Xi \Phi^z \zeta d\mathbb{P} \\ -\phi^z \end{pmatrix}$$

is monotone, where

$$\Phi^z = \begin{pmatrix} \tilde{M}_1(\xi^1) + M_1 & R_1 & b_1 \\ R_2 & \tilde{M}_2(\xi^2) + M_2 & b_2 \\ -b_1^\top & -b_2^\top & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \mu \end{pmatrix} + \begin{pmatrix} \tilde{c}_1(\xi^1)^\top + c_1^\top \\ \tilde{c}_2(\xi^2)^\top + c_2^\top \\ 0 \end{pmatrix} \text{ and } \phi^z = \begin{pmatrix} f_1(x,\cdot) \\ f_2(x,\cdot) \end{pmatrix}.$$

Note also that $f_i$ and $g_i$ are strongly monotone, $i = 1, 2$. Similar as the analysis in Example 6, it is easy to see $\mathcal{G}(x, \mu, \zeta)$ is weakly coercive. Then by Theorem 1, the distributionally robust generalized Nash equilibrium problem has a solution.

Since random variable $\phi^x(\xi)$ in (59) is $f_1(x, \xi)$ and $f_2(x, \xi)$ and $\mathbb{P}$ follows the uniform distribution over $[-1, 1]$, $f_1(x, \xi)$ and $f_2(x, \xi)$ follow continuous distribution and their $\alpha$-quantiles, denote by $\kappa_\alpha^1$ and $\kappa_\alpha^2$, are unique for all $\alpha \in [0, 1]$ when $x \neq 0$. In this case, $\mathbb{P}(f_1(x, \xi) = \kappa_\alpha^1) = 0$ and $\mathbb{P}(f_2(x, \xi) = \kappa_\alpha^2) = 0$ for all $\alpha \in [0, 1]$, and then $\partial_{f_1}\mathcal{R}(f_1(x, \xi))$ and $\partial_{f_2}\mathcal{R}(f_2(x, \xi))$ are singleton.

Let $\{\xi^1, \cdots, \xi^N\}$ be the i.i.d. sample of $\xi$ generated from $\mathbb{P}$. Then we solve

$$0 \in \sum_{j=1}^N p_i^j \nabla_{x_i} f_i(x, \xi^j) + \nabla_{x_i} g_i(x) + b_i\mu + \mathcal{N}_{X_i}(x_i), \qquad i = 1, 2 \quad (66)$$

$$0 \in c - b_1x_1 - b_2x_2 + \mathcal{N}_{\mathbb{R}_+}(\mu), \quad (67)$$

where $P_i = (p_i^1, \cdots, p_i^N)$ is from

$$P_i \in \arg\max_{Q_i \in \mathfrak{M}_i^N} \frac{1}{N} \sum_{j=1}^N q_i^j \phi_i(x_1, x_2, \xi^j) \quad (68)$$

with $\phi_i(x_1, x_2, \xi) = f_i(x_1, x_2, \xi) + g_i(x_1, x_2)$, $Q_i = (q_i^1, \cdots, q_i^N)$ and

$$\mathfrak{M}_i^N := \left\{ P \in \mathbb{R}_+^N : \sum_{j=1}^N (p^j - \frac{1}{N})^2 \leq \frac{0.05}{N},\ \sum_{j=1}^N p_j = 1 \right\}.$$

We consider sample size $N = (50, 100, 300, 600, 1200)$. For each sample size, we generate 20 group of samples and solve the corresponding DRVI (66) - (68) by Algorithm 1[10] with $\tau = 0.2$ and randomly generated $z^0 \in [0, 1]^5$ using the uniform distribution in Matlab.

Since the two players are symmetric, then $x_1 = x_2$ and we only show $x_1$ with $x_1 = (x_{11}, x_{12})^\top$ in Figures 1-2. From the two figures, we can observe the tendency of convergence as the sample size increases, which is consistent with our convergence results.

[10] In Step 4 of Algorithm 1, we do not specify how to solve the monotone VI: $0 \in F^k(z) + \mathcal{N}_{X_1 \times X_2 \times \mathbb{R}_+}(z)$. We can solve it by any suitable method, such as extragradient method.

**Algorithm 1** Projection method for solving DRVI.

1: Choose a parameter $\tau \in (0,1)$ and an initial point $z^0 = ((x_1^0)^\top, (x_2^0)^\top, \mu^0)^\top$. Set $k \leftarrow 0$.

2: Solve
$$P_i^k = \arg \max_{Q_i \in \mathfrak{M}_i^N} \mathbb{E}_{Q_i}[\phi_i(x^k, \xi)], \quad \text{for} \quad i = 1, 2.$$

3: Set
$$F^k(z) = \begin{pmatrix} \sum_{j=1}^N (p_1^j)^k \nabla_{x_1} f_1(x, \xi^j) + \nabla_{x_1} g_1(x) + b_1\mu \\ \sum_{j=1}^N (p_2^j)^k \nabla_{x_2} f_2(x, \xi^j) + \nabla_{x_2} g_2(x) + b_2\mu \\ c - b_1 x_1 - b_2 x_2 \end{pmatrix}$$
where $((p_i^1)^k, \cdots, (p_i^N)^k) = P_i^k$, $i = 1, 2$.

4: If $\|\min(z^k, F^k(z^k))\|_2 \le 10^{-8}$ (where 'min' is in the sense of element-wise), stop; otherwise find $z^{k+1}$ such that
$$\|z^{k+1} - (z^{k+1} - \tau F^k(z^{k+1}))_+\| \le 10^{-8}.$$

5: $k \leftarrow k+1$, go to Step 2.

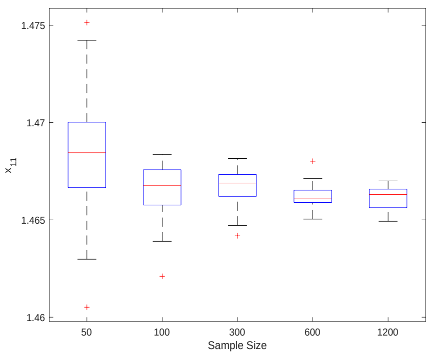


**Fig. 1** convergence of $x_{11}$

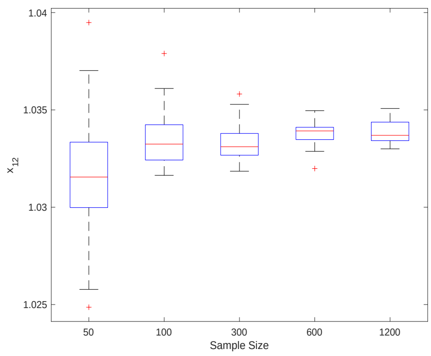


**Fig. 2** convergence of $x_{12}$

## 6 Concluding remarks

To deal with uncertainties of probability distributions $\mathbb{P}$ in the SVI (1), we propose a formulation of the DRVI in Definition 1. This formulation provides a unified framework for the research of many important problems including the optimality conditions for distributionally robust optimization and DRG. We show the existence of solutions of the DRVI under the conditions that the set $X$ of decision variables is convex and bounded or the operator in the DRVI is monotone and coercive. Moreover, under the condition that the set of densities associated with the ambiguity set $\mathfrak{M}$ is law invariant, we propose an SAA approach to the DRVI by using the corresponding law invariant risk measure $\mathcal{R}$ and establish its convergence properties as the sample size $N$ goes to infinity. The formulation of the DRVI, solutions of the DRVI, the monotone condition, the SAA approach and the convergence properties of the SAA are illustrated by several examples. Within this new DRVI framework, some new algorithms can be developed for finding robust solutions of optimization and equilibrium problems under uncertain environment.

**Acknowledgements** We would like to thank the editor and referees for their helpful comments and suggestions.

## References

1. Bayraksan, G., Love, D. K.: Data-driven stochastic programming using phi-divergences, in Tutorials in Operations Research, INFORMS, Catonsville, MD (2015)
2. Ben-Tal, A., Teboulle, M.: Penalty functions and duality in stochastic programming via phi-divergence functionals, Math. Oper. Res. 12, 224-240 (1987)
3. Billingsley, P.: Convergence of Probability Measures, John Wiley & Sons, New York (1999).
4. Bertsimas, D., Sim., M.: The Price of Robustness. Oper. Res. 52, 35-53 (2004)
5. Bonnans J. F., Shapiro, A.: Perturbation Analysis of Optimization Problems, Wiley Series in Probability and Statistics, New York (2000)
6. Chen, X., Fukushima, M.: Expected residual minimization method for stochastic linear complementarity problems. Math. Oper. Res. 30, 1022–1038 (2005)
7. Chen, X., Pong, T. K., Wets, R.: Two-stage stochastic variational inequalities: an ERM-solution procedure. Math. Program. 165, 71-112 (2015)
8. Chen, X., Wets, R., Zhang, Y.: Stochastic variational inequalities: residual minimization smoothing sample average approximations. SIAM J. Optim. 22, 649–673 (2012)
9. Chen, X., Sun, H., Xu, H.: Discrete approximation of two-stage stochastic and distributionally robust linear complementarity problems. Math. Program. 177, 255-289 (2019)
10. Chen, X., Shapiro, A., Sun, H.: Convergence analysis of sample average approximation of two-stage stochastic generalized equations. SIAM J. Optim. 29, 135-161 (2019)
11. Chen, Y., Sun, H., Xu, H.: Decomposition and discrete approximation methods for solving two-stage distributionally robust optimization problems, Comp. Optim. Appl. 28, 205-238 (2021)
12. Chen, Y., Lan, G., Ouyang, Y.: Accelerated schemes for a class of variational inequalities. Math. Program. 165, 113–149 (2017)
13. Chieu, N. H., Trang, N. T. Q.: Coderivative and monotonicity of continuous mappings. Taiwanese J. Math. 16, 353–365 (2012)
14. Csiszár, I.: Eine informationstheoretische ungleichung und ihre anwendung auf den beweis der ergodizitat von markoffschen ketten, Magyar. Tud. Akad. Mat. Kutato Int. Kozls, 8 (1963).
15. Dommel, P., Pichler, A.: Convex risk measures based on divergence, Optimization-online (2020)

16. Facchinei, F., Pang, J. S.: Finite-Dimensional Variational Inequalities and Complementarity Problems, Springer-Verlag, New York (2003)
17. Hadjisavvas, H., Komlósi, S., Schaible, S.: Handbook of Generalized Convexity and Generalized Monotonicity, Springer, New York (2005)
18. Krebs, V., Schmidt, M.: $\Gamma$- Robust linear complementarity problems, Optim. Methods Softw. online first (2020)
19. Morimoto, T.: Markov processes and the h-theorem, J. Phys. Soc. Japan. 18, 328 - 333 (1963)
20. Milz, J., Ulbrich, M.: An approximation scheme for distributionally robust nonlinear optimization. SAIM J. Optim. 30, 1996–2025 (2020)
21. Morton, S.: Lagrange multipliers revisited, Cowles Commission Discussion Paper No. 403 (1950)
22. Pardo, L.: Statistical Inference Based on Divergence Measures. Chapman and Hall/CRC, Boca Raton, FL (2005)
23. Römisch, W.: Stability of Stochastic Programming Problems, in *Stochastic Programming*, Ruszczyński A., Shapiro, A. eds., Elsevier, Amsterdam (2003)
24. Rockafellar, R. T., Wets, R. J.-B.: Stochastic variational inequalities: single-stage to multistage. Math. Program. 165, 331–360 (2017)
25. Rockafellar R. T., Sun, J.: Solving monotone stochastic variational inequalities and complementarity problems by progressive hedging. Math. Program. 174, 453-471 (2018)
26. Shapiro, A.: Consistency of sample estimates of risk averse stochastic programs. J. Appl. Probab. 50, 533-541 (2013)
27. Shapiro, A., Dentcheva, D., Ruszczyński, A.: Lectures on Stochastic Programming: Modeling and Theory, second edition. SIAM, Philadelphia (2014)
28. Shapiro, A.: Distributionally robust stochastic programming. SIAM J. Optim. 27, 2258–2275 (2017)
29. Shanbhag U. V.: Stochastic variational inequality problems: applications, analysis, and algorithms. INFORMS Tutor. Oper. Res. 71–107 (2013)
30. Sun, H., Xu, H.: Convergence analysis for distributionally robust optimization and equilibrium problems. Math. Oper. Res. 41, 377-401 (2016)
31. Sun, H., Chen, X.: Two-stage stochastic variational inequalities: theory, algorithms and application. J. Oper. Res. Soc. China, (2019) https://doi.org/10.1007/s40305-019-00267-8
32. Wu, D., Han, J. Y., Zhu, J. H.: Robust solutions to uncertain linear complementarity problems. Acta Math. Sin. (Engl. Ser.) 27, 339-352 (2011)
33. Xu, H., Liu, Y., Sun, H.: Distributionally robust optimization with matrix moment constraints: Lagrange duality and cutting plane methods. Math. Program. 169, 489–529 (2018)
34. Xie, Y., Shanbhag, U. V.: On robust solutions to uncertain linear complementarity problems and their variants, SIAM J. Optim., 26(4), 2120–2159 (2016)

## 7 Appendix

In this appendix, we give some proofs and necessary results used in this paper.

*Example 8* Consider the Average Value-at-Risk,

$$\mathsf{AV@R}_{1-\alpha}(Z) := \frac{1}{1-\alpha}\int_{\alpha}^{1} H_Z^{-1}(t)dt = \inf_{\tau\in\mathbb{R}}\left\{\tau + (1-\alpha)^{-1}\mathbb{E}_{\mathbb{P}}[Z-\tau]_+\right\},\ \alpha\in(0,1). \tag{A1}$$

Here $\mathcal{Z} = L_1(\Xi, \mathcal{B}, \mathbb{P})$ and a minimizer in the right hand side of (A1) is $\bar{\tau} = H_Z^{-1}(\alpha)$. The empirical estimate of $\mathsf{AV@R}_{1-\alpha}(\phi^x)$ is then

$$\widehat{\mathsf{AV@R}}_{(1-\alpha)N}(\phi^x) = \inf_{\tau\in\mathbb{R}}\left\{\tau + \frac{1}{(1-\alpha)N}\sum_{j=1}^{N}\left[\phi^x(\xi^j)-\tau\right]_+\right\}. \tag{A2}$$

We have that $\partial\mathsf{AV@R}_{1-\alpha}(Z)$ is a singleton iff $\mathbb{P}\{Z=\kappa\}=0$, where $\kappa_\alpha := H_Z^{-1}(\alpha)$. Suppose that $\partial\mathsf{AV@R}_{1-\alpha}(Z) = \{\bar{\zeta}\}$ is a singleton. Then

$$\bar{\zeta}(s) = \begin{cases} (1-\alpha)^{-1} & if\ Z(s) > \kappa,\ s\in\Xi, \\ 0 & if\ Z(s) < \kappa,\ s\in\Xi, \end{cases} \tag{A3}$$

(cf. [27, eq. (6.80), p. 292]). For $x\in X$ and $Z := \phi^x$ let $\{\bar{\zeta}^x\}$ be the corresponding subdifferential. The subdifferential $\hat{\zeta}^x = (\zeta_1^x, ..., \zeta_N^x)$ of the corresponding empirical estimate is obtained by replacing $\kappa_\alpha$ with their empirical estimates. That is $\zeta_j^x = (1-\alpha)^{-1}$ if $\phi^x(\xi^j) > \kappa_{\alpha,N}$ and $\zeta_j^x = 0$ if $\phi^x(\xi^j) < \kappa_{\alpha,N}$, where $\kappa_{\alpha,N}$ is the empirical estimate of $\kappa_\alpha$. Note that because of the assumption $\mathbb{P}\{Z=\kappa\}=0$, the empirical estimate $\kappa_{\alpha,N}$ converges w.p.1 to $\kappa_\alpha$.

Consider the probability distribution $P_N^x$ on $\{\xi^1, ..., \xi^N\}$ associated with density $\hat{\zeta}^x$, i.e., with $\xi^j$ being assigned probability $1/((1-\alpha)N)$ if $\phi^x(\xi^j) > \kappa_{\alpha,N}^x$, and 0 otherwise. We view $P_N^x$ as the empirical counterpart of $P^x$, where $P^x$ is the probability measure absolutely continuous with respect to $\mathbb{P}$ and having density $\bar{\zeta}^x$, i.e.,

$$dP^x = \bar{\zeta}^x d\mathbb{P}. \tag{A4}$$

Consider a continuous bounded function $g : \Xi \to \mathbb{R}$. Since $g(\cdot)$ is bounded and continuous, $\kappa_{\alpha,N}^x \to \kappa_\alpha^x$ w.p.1 and $\mathbb{P}\{\phi^x(\xi) = \kappa_\alpha^x\} = 0$, we have that

$$\int_{\Xi} g(s)dP_N^x(s) = \frac{1}{(1-\alpha)N}\sum_{\phi^x(\xi^j)>\kappa_{\alpha,N}^x} g(\xi^j)$$

converges w.p.1 to

$$\int_{\Xi} g(s)\bar{\zeta}^x(s)d\mathbb{P}(s) = \frac{1}{1-\alpha}\int_{\phi^x(\xi)>\kappa_\alpha^x} g(s)d\mathbb{P}(s).$$

That is $P_N^x$ converges weakly[11] to $P^x$. Moreover, by Proposition 6 in the Appendix, we have if $\{x_N\}$ is a sequence in $X$ converging to $x$, then $\int_\Xi g(s)dP_N^{x_N}(s)$ converges to $\int_{\phi^x(\xi)>\kappa} g(s)d\mathbb{P}(s)$ w.p.1, and hence $P_N^{x_N}$ converges weakly to $P^x$.

**Proposition 6** *Suppose (i) $\phi(\cdot,\xi)$ is Lipschitz continuous in $x \in X$ with a uniform Lipschitz modules $k_\phi$, (ii) the CDF of $\phi^{\bar{x}}$ is strictly monotone, and (iii) $\{x_N\}$ is a sequence in $X$ converging to $\bar{x}$, (iv) $|\kappa_\alpha^{x'}|$ is bounded by a constant number for all $x' \in \mathcal{B}(x) \cap X$. Then for any bounded and continuous function $g$, $\int_\Xi g(s)dP_N^{x_N}(s)$ converges to $\int_{\phi^{\bar{x}}(\xi)>\kappa_\alpha^{\bar{x}}} g(s)d\mathbb{P}(s)$.*

*Proof* For any continuous and bounded function $g(s)$,

$$\begin{aligned}
&\left|\frac{1}{\alpha N}\sum_{\phi^{x_N}(\xi^j)>\kappa_{\alpha,N}^{x_N}} g(\xi^j) - \frac{1}{\alpha}\int_{\phi^{\bar{x}}(\xi)>\kappa_\alpha^{\bar{x}}} g(s)d\mathbb{P}(s)\right| \\
\leq &\left|\frac{1}{\alpha N}\sum_{\phi^{x_N}(\xi^j)>\kappa_{\alpha,N}^{x_N}} g(\xi^j) - \frac{1}{\alpha N}\sum_{\phi^{\bar{x}}(\xi^j)>\kappa_\alpha^{\bar{x}}} g(\xi^j)\right| + \left|\frac{1}{\alpha N}\sum_{\phi^{\bar{x}}(\xi^j)>\kappa_\alpha^{\bar{x}}} g(\xi^j) - \frac{1}{\alpha}\int_{\phi^{\bar{x}}(\xi)>\kappa_\alpha^{\bar{x}}} g(s)d\mathbb{P}(s)\right|.
\end{aligned} \tag{A5}$$

We first prove that $\kappa_{\alpha,N}^{x_N}$ converges to $\kappa_\alpha^{\bar{x}}$ w.p.1. To see this, by condition (ii), we have $\mathbb{P}\{\phi^{\bar{x}}(\xi) = \kappa_\alpha^{\bar{x}}\} = 0$, then

$$\kappa_\alpha^{\bar{x}} = \arg\min_\tau \ \tau + \frac{1}{1-\alpha}\mathbb{E}_\mathbb{P}[(\phi^{\bar{x}} - \tau)_+]$$

and

$$\kappa_{\alpha,N}^{x_N} \in \arg\min_\tau \ \tau + \frac{1}{(1-\alpha)N}\sum_{j=1}^N (\phi^{x_N}(\xi^j) - \tau)_+.$$

It is easy to observe that $(\phi^{\bar{x}}(\xi) - \tau)_+$ is continuous w.r.t. $x$ and dominated by an integrable function, then by uniform law of large numbers [27, Theorem 7.48]

$$\left|\mathbb{E}_\mathbb{P}[(\phi^{\bar{x}} - \tau)_+] - \frac{1}{N}\sum_{j=1}^N (\phi^{x_N}(\xi^j) - \tau)_+\right| \to 0$$

as $N \to \infty$ w.p.1. Then by conditions (i)-(iv) and [5, Proposition 4.4], we have that $\kappa_{\alpha,N}^{x_N}$ converges to $\kappa_\alpha^{\bar{x}}$ w.p.1.

[11] Recall that a sequence $P_N$ of probability measures converges weakly to a probability measure $P$ if $\int g dP_N \to \int g dP$ for any bounded continuous function $g : \Xi \to \mathbb{R}$, see e.g., Billingsley [3] for a discussion of weak convergence of probability measures.

Then we prove the convergence of first part in the right side of (A5). Let

$$A_N^1 = \{\xi \in \Xi : \phi^{x_N}(\xi) > \kappa_{\alpha,N}^{x_N}, \phi^{\bar{x}}(\xi) \le \kappa_\alpha^{\bar{x}}\}, \quad A_N^2 = \{\xi \in \Xi : \phi^{x_N}(\xi) \le \kappa_{\alpha,N}^{x_N}, \phi^{\bar{x}}(\xi) > \kappa_\alpha^{\bar{x}}\}$$

$$A_N^3 = \{\xi \in \Xi : \phi^{\bar{x}}(\xi) \ge \kappa_{\alpha,N}^{x_N} - k_\phi \|\bar{x} - x_N\|, \phi^{\bar{x}}(\xi) \le \kappa_\alpha^{\bar{x}}\},$$

and

$$A_N^4 = \{\xi \in \Xi : \phi^{\bar{x}}(\xi) \le \kappa_{\alpha,N}^{x_N} + k_\phi \|\bar{x} - x_N\|, \phi^{\bar{x}}(\xi) \ge \kappa_\alpha^{\bar{x}}\}.$$

Then

$$\begin{aligned} \left| \frac{1}{\alpha N} \sum_{\phi^{x_N}(\xi^j) > \kappa_{\alpha,N}^{x_N}} g(\xi^j) - \frac{1}{\alpha N} \sum_{\phi^{\bar{x}}(\xi^j) > \kappa_\alpha^{\bar{x}}} g(\xi^j) \right| &\le \tfrac{1}{\alpha} \mathbb{P}_N(A_N^1 \cup A_N^2) |\max_s g(s)| \\ &\le \tfrac{1}{\alpha} \mathbb{P}_N(A_N^3 \cup A_N^4) |\max_s g(s)|, \end{aligned}$$

where $\mathbb{P}_N$ is an empirical estimation of $\mathbb{P}$. Note that $\kappa_{\alpha,N}^{x_N} \to \kappa_\alpha^{\bar{x}}$ and $x_N \to \bar{x}$, $A_N^1 \subset A_N^3$ and $A_N^2 \subset A_N^4$, and $A_N^3$ and $A_N^4$ converge to singleton sets. Then by condition (i) and (ii), $\mathbb{P}_N(A_N^1 \cup A_N^2) \le \mathbb{P}_N(A_N^3 \cup A_N^4) \to 0$ as $N \to \infty$ w.p.1, which implies

$$\left| \frac{1}{\alpha N} \sum_{\phi^{x_N}(\xi^j) > \kappa_{\alpha,N}^{x_N}} g(\xi^j) - \frac{1}{\alpha N} \sum_{\phi^{x_N}(\xi^j) > \kappa_\alpha^{\bar{x}}} g(\xi^j) \right| \to 0 \tag{A6}$$

as $N \to \infty$ w.p.1.

Then we consider the second part in the right side of (A5). Since $g$ is continuous and bounded, by classical law of large numbers, as $N \to \infty$ w.p.1,

$$\left| \frac{1}{\alpha N} \sum_{\phi^{\bar{x}}(\xi^j) > \kappa_\alpha^{\bar{x}}} g(\xi^j) - \frac{1}{\alpha} \int_{\phi^{\bar{x}}(\xi) > \kappa_\alpha^{\bar{x}}} g(s) d\mathbb{P}(s) \right| \to 0.$$

Combining discussion above, we have the conclusion. □

Then we derive a kind of uniform Glivenko-Cantelli theorem which we need in the proof of Lemma 2. Let $f(x, \xi)$ be a random function and $\{x_N\} \to x$ as $N \to \infty$. Moreover, suppose $f(x, \xi)$ is Lipschitz continuous w.r.t. $x$ and $\xi$, and the Lipschitz modules $\kappa(\xi)$ of $f(\cdot, \xi)$ is integrable. We use $H_{x_N}(t)$ and $H_x(t)$ to denote the CDF of $f(x_N, \xi)$ and $f(x, \xi)$ w.r.t. $\mathbb{P}$ and $H_{x_N}^N(t)$ and $H_x^N(t)$ are used to denote the CDF of their empirical distributions i.i.d samples $\{\xi^1, \cdots, \xi^N\}$.

**Lemma 3** *Suppose $f(x,\xi)$ is integrable and continuous w.r.t. $x$, and $\mathbb{P}$ is a continuous distribution. Then for each $\epsilon > 0$, there exists a finite partition of the real line of the form $-\infty = t_0 < t_1 < \cdots < t_k = \infty$ such that for $0 \le j \le k-1$, $H(x_N, t_{j+1}) - H(x_N, t_j) \le \epsilon$ for all $N$ sufficiently large.*

*Proof* Since $\mathbb{P}$ is a continuous distribution, $H_x(t)$ is a CDF of continuous distribution and then, for any $\epsilon > 0$ there exists $-\infty = t_0 < t_1 < \cdots < t_k = \infty$ such that for $0 \le j \le k-1$, $H_x(t_{j+1}) - H_x(t_j) \le \frac{\epsilon}{2}$. Moreover, since $f(x,\xi)$ is integrable and continuous w.r.t. $x$, by Lebesgue's dominated convergence theorem, for any continuous and bounded function $h$,

$$\lim_{N\to\infty} \mathbb{E}[h(f(x,\xi)) - h(f(x_N,\xi))] = \mathbb{E}\left[\lim_{N\to\infty}(h(f(x,\xi)) - h(f(x_N,\xi)))\right] = 0.$$

Then $f(x,\cdot)$ converges to $f(x_N,\cdot)$ weakly, which is equivalent to $\lim_{n\to\infty} |H_{x_N}(t) - H_x(t)| = 0$ for any $t \in \mathbb{R}$. Then there exists sufficiently large $n$ such that $\sup_{j\in\{0,\cdots,k\}} |H_{x_N}(t_j) - H_x(t_j)| \le \frac{\epsilon}{4}$. Then we have

$$\begin{aligned}|H_{x_N}(t_{j+1}) - H_{x_N}(t_j)| &\le |H_{x_N}(t_{j+1}) - H_x(t_{j+1})| + |H_x(t_{j+1}) - H_x(t_j)| + |H_x(t_j) - H_{x_N}(t_j)| \\ &\le \tfrac{\epsilon}{4} + \tfrac{\epsilon}{2} + \tfrac{\epsilon}{4} = \epsilon.\end{aligned}$$

□

**Theorem 3** *Suppose $f(x,\xi)$ is Lipschitz continuous w.r.t. $x$ and $\xi$, and the Lipschitz modules $\kappa(\xi)$ of $f(\cdot,\xi)$ is integrable, $f(x,\cdot) \in \mathcal{L}_P(\Xi,\mathcal{F},\mathbb{P})$ and $\mathbb{P}$ is a continuous distribution. Then w.p.1*

$$\lim_{N\to\infty} \sup_{t\in\mathbb{R}} |H^N_{x_N}(t) - H_x(t)| = 0, \tag{A7}$$

*and $(H^N_{x_N})^{-1}$ converges w.p.1 to $H_x^{-1}$ in the norm topology of $\mathcal{L}_p$ as $N \to \infty$.*

*Proof* Note that

$$|H^N_{x_N}(t) - H_x(t)| \le |H^N_{x_N}(t) - H_{x_N}(t)| + |H_{x_N}(t) - H_x(t)|.$$

It is sufficient to show that for any $\epsilon > 0$,

$$\limsup_{N\to\infty} \sup_t |H^N_{x_N}(t) - H_{x_N}(t)| \le \epsilon \tag{A8}$$

and

$$\limsup_{N\to\infty} \sup_t |H_{x_N}(t) - H_x(t)| \le \epsilon. \tag{A9}$$

We consider (A8) firstly. By Lemma 3, there exists $-\infty = t_0 < t_1 < \cdots < t_k = \infty$ such that for $0 \leq j \leq k-1$, $H_{x_N}(t_{j+1}) - H_{x_N}(t_j) \leq \frac{\epsilon}{2}$ for all $n$ sufficiently large. For any $t$, there exists $j$ such that $t_j \leq t \leq t_{j+1}$. For such $j$,

$$H_{x_N}^N(t_j) \leq H_{x_N}^N(t) \leq H_{x_N}^N(t_{j+1}) \;\text{ and }\; H_{x_N}(t_j) \leq H_{x_N}(t) \leq H_{x_N}(t_{j+1}),$$

which implies

$$H_{x_N}^N(t_j) - H_{x_N}(t_{j+1}) \leq H_{x_N}^N(t) - H_{x_N}(t) \leq H_{x_N}^N(t_{j+1}) - H_{x_N}(t_j).$$

Then we have

$$H_{x_N}^N(t_j) - H_{x_N}(t_j) + H_{x_N}(t_j) - H_{x_N}(t_{j+1}) \leq H_{x_N}^N(t) - H_{x_N}(t)$$

and

$$H_{x_N}^N(t_{j+1}) - H_{x_N}(t_{j+1}) + H_{x_N}(t_{j+1}) - H_{x_N}(t_j) \geq H_{x_N}^N(t) - H_{x_N}(t).$$

Note that by Lemma 3 and by uniform law of large numbers [27, Theorem 7.48], $H_{x_N}(t_{j+1}) - H_{x_N}(t_j) \leq \frac{\epsilon}{2}$ and $|H_{x_N}^N(t_{j+1}) - H_{x_N}(t_j)| \leq \frac{\epsilon}{4}$ for all $N$ sufficiently large and $j = 0, \cdots, k$, then we have (A8). Now we consider (A9). Similar as the procedure above, For any $t$, there exists $j$ such that $t_j \leq t \leq t_{j+1}$. For such $j$,

$$H_x(t_j) \leq H_x(t) \leq H_x(t_{j+1}) \;\text{ and }\; H_{x_N}(t_j) \leq H_{x_N}(t) \leq H(x_N, t_{j+1}).$$

Then by continuous distribution of $\mathbb{P}$, Lipschitz continuity of $f(x,\xi)$ w.r.t. $x$ and Lemma 3, for any $t \in \mathbb{R}$,

$$\begin{aligned} |H_{x_N}(t) - H_x(t)| &\leq |H_x(t) - H_{x_N}(t_j)| + |H_{x_N}(t_j) - H_x(t_j)| + |H_x(t_j) - H_x(t)| \\ &\leq |H_{x_N}(t_{j+1}) - H_{x_N}(t_j)| + |H_{x_N}(t_j) - H_x(t_j)| + |H_x(t_j) - H_x(t_{j+1})| \\ &\leq \epsilon. \end{aligned}$$

Combining (A8) and (A9), we have (A7).

Moreover, (A7) implies that $(H_{x_N}^N)^{-1}$ pointwise converges to $H_x^{-1}$ on the set $[0,1]$. Then, if the sequence $\{|(H_{x_N}^N)^{-1}(s) - H_x^{-1}(s)|^p\}$ is uniformly integrable, $(H_{x_N}^N)^{-1}$ converges w.p.1 to $H_x^{-1}$ in the norm topology of $\mathcal{L}_p$ as $N \to \infty$, that is w.p.1

$$\lim_{N\to\infty} \int_0^1 |(H_{x_N}^N)^{-1}(s) - H_x^{-1}(s)|^p ds = \int_0^1 \lim_{N\to\infty} |(H_{x_N}^N)^{-1}(s) - H_x^{-1}(s)|^p ds = 0,$$

where the first equality comes from the Lebesgue's dominated convergence theorem.

Let us show that the uniform integrability indeed holds. By triangle inequality,

$$|(H_{x_N}^N)^{-1}(s) - H_x^{-1}(s)|^p \leq |(H_{x_N}^N)^{-1}(s)|^p + |H_x^{-1}(s)|^p.$$

Then we only need to show the uniform integrability of $|(H_{x_N}^N)^{-1}(s)|^p$. Note that

$$\int_0^1 |(H_{x_N}^N)^{-1}(s)|^p ds = \int_{\Xi} |f(x_N,\xi)|^p dH_{x_N}^N = \frac{1}{N}\sum_{i=1}^N |f(x_N,\xi^i)|^p.$$

Since the Lipschitz continuity of $f(x,\xi)$ with Lipschitz modules $\kappa(\xi)$,

$$\begin{aligned}
|\tfrac{1}{N}\textstyle\sum_{i=1}^N |f(x_N,\xi^i)|^p - \mathbb{E}_{\mathbb{P}}[|f(x,\xi)|^p]| &\leq |\tfrac{1}{N}\textstyle\sum_{i=1}^N |f(x_N,\xi^i)|^p - \tfrac{1}{N}\sum_{i=1}^N |f(x,\xi^i)|^p| \\
&\quad + |\tfrac{1}{N}\textstyle\sum_{i=1}^N |f(x,\xi^i)|^p - \mathbb{E}_{\mathbb{P}}[|f(x,\xi)|^p]| \\
&\leq |\tfrac{1}{N}\textstyle\sum_{i=1}^N \kappa(\xi)(x-x_N)|^p \\
&\quad + |\tfrac{1}{N}\textstyle\sum_{i=1}^N |f(x,\xi^i)|^p - \mathbb{E}_{\mathbb{P}}[|f(x,\xi)|^p]|.
\end{aligned}$$

Moreover, by the Law of Large Numbers and $x_N \to x$, $\frac{1}{N}\sum_{i=1}^N \kappa(\xi) \to \mathbb{E}_{\mathbb{P}}[\kappa(\xi)]$, $|\frac{1}{N}\sum_{i=1}^N \kappa(\xi)(x-x_N)|^p \to 0$ and $|\frac{1}{N}\sum_{i=1}^N |f(x,\xi^i)|^p - \mathbb{E}_{\mathbb{P}}[|f(x,\xi)|^p]| \to 0$ as $N \to \infty$ w.p.1. It follows that $|(H_{x_N}^N)^{-1}(s)|^p$ converges w.p.1 to a finite limit, which implies that w.p.1 $|(H_{x_N}^N)^{-1}(s)|^p$ is uniformly integrable. $\square$

**Proof of Proposition 4.** For any continuous and bounded function $g : \Xi \to \mathbb{R}$, we have that

$$\int_{\Xi} g(s)\bar{\zeta}^{\bar{x}}(s)d\mathbb{P}(s) = \int_{[0,1)}\int_{\Xi} g(s)\bar{\zeta}_{\alpha}^{\bar{x}}(s)d\mathbb{P}(s)d\bar{\mu}(\alpha) = \int_{[0,1)} \frac{1}{1-\alpha}\int_{\phi^{\bar{x}}(\xi)>\kappa_\alpha} g(s)d\mathbb{P}(s)d\bar{\mu}(\alpha),$$

where $\bar{\mu}$ is corresponding to $\bar{\sigma}$. Moreover,

$$\int_{\Xi} g(s)dP_N^{x_N}(s) = \int_{[0,1)} \frac{1}{N}\sum_{j=1}^N g(\xi^j)(\zeta_j^{x_N})_\alpha d\mu_N(\alpha) = \int_{[0,1)} \frac{1}{(1-\alpha)N}\sum_{\phi^{x_N}(\xi^j)>\kappa_{N,x_N}^\alpha} g(\xi^j)d\mu_N(\alpha).$$

Then

$$\begin{aligned}
&|\textstyle\int_{\Xi} g(s)dP_N^{x_N}(s) - \int_{\Xi} g(s)\bar{\zeta}^{\bar{x}}(s)d\mathbb{P}(s)| \\
&\leq \left|\int_{[0,1)} \frac{1}{(1-\alpha)N}\sum_{\phi^{x_N}(\xi^j)>\kappa_{N,x_N}^\alpha} g(\xi^j)d\mu_N(\alpha) - \int_{[0,1)} \frac{1}{1-\alpha}\int_{\phi^{\bar{x}}(\xi)>\kappa_\alpha} g(s)d\mathbb{P}(s)d\mu_N(\alpha)\right| \\
&+\left|\int_{[0,1)} \frac{1}{1-\alpha}\int_{\phi^{\bar{x}}(\xi)>\kappa_\alpha} g(s)d\mathbb{P}(s)d\mu_N(\alpha) - \int_{[0,1)} \frac{1}{1-\alpha}\int_{\phi^{\bar{x}}(\xi)>\kappa_\alpha} g(s)d\mathbb{P}(s)d\bar{\mu}(\alpha)\right|.
\end{aligned} \tag{A10}$$

We first prove

$$\left|\int_{[0,1)} \frac{1}{1-\alpha}\int_{\phi^{\bar{x}}(\xi)>\kappa_\alpha} g(s)d\mathbb{P}(s)d\mu_N(\alpha) - \int_{[0,1)} \frac{1}{1-\alpha}\int_{\phi^{\bar{x}}(\xi)>\kappa_\alpha} g(s)d\mathbb{P}(s)d\bar{\mu}(\alpha)\right| \to 0 \tag{A11}$$

as $N\to\infty$. From condition (iii), $g(\xi)$ is continuous and bounded and $\phi^{\bar{x}}(\xi)$ is continuous w.r.t. $\xi$. Then for any $\alpha'\to\alpha$, $\alpha',\alpha\in[0,1)$, we have

$$\left|\int_{\phi^{\bar{x}}(\xi)>\kappa_{\alpha'}} g(s)d\mathbb{P}(s) - \int_{\phi^{\bar{x}}(\xi)>\kappa_\alpha} g(s)d\mathbb{P}(s)\right| \le \mathbb{P}((A_{\alpha'}-A_\alpha)\cup(A_\alpha-A_{\alpha'}))\max_s g(s),$$

where $A_\alpha=\{\xi:\phi^{\bar{x}}(\xi)>\kappa_\alpha\}$ and $A_{\alpha'}=\{\xi:\phi^{\bar{x}}(\xi)>\kappa_{\alpha'}\}$. Note that $a'\to a$ and the CDF of $\phi^{\bar{x}}$ is strictly monotone, $A_{\alpha'}\to A_\alpha$ and $\mathbb{P}((A_{\alpha'}-A_\alpha)\cup(A_\alpha-A_{\alpha'}))\to 0$. Then we have that $\frac{1}{1-\alpha}\int_{\phi^{\bar{x}}(\xi)>\kappa_\alpha} g(s)d\mathbb{P}(s)$ is continuous and bounded w.r.t. $\alpha$, and (A11) is from the fact that $\mu_N$ weak* converges to $\mu$. Indeed, by Lemma 2, $\sigma_N$ weak* converges to $\bar{\sigma}$, then for any continuous and bounded function $g(t)$, $\int_{[0,1)} g(t)\sigma_N(t)dt \to \int_{[0,1)} g(t)\bar{\sigma}(t)dt$ as $N\to\infty$. Note that $\mu(\alpha)=(1-\alpha)\sigma(\alpha)+\int_0^\alpha \sigma(t)dt$. Then

$$\begin{aligned}|\textstyle\int_{[0,1)} g(\alpha)\mu_N(\alpha)d\alpha - \int_{[0,1)} g(\alpha)\bar{\mu}(\alpha)d\alpha| &= (1-\alpha)|\textstyle\int_{[0,1)} g(\alpha)\sigma_N(\alpha)d\alpha - \int_{[0,1)} g(\alpha)\bar{\sigma}(\alpha)d\alpha|\\ &\quad+\textstyle\int_{[0,1)}\int_0^\alpha g(\alpha)(\sigma_N(t)-\bar{\sigma}(t))dtd\alpha \to 0\end{aligned}$$

as $N\to\infty$, which implies that $\mu_N$ weak* converges to $\bar{\mu}$.

Then we prove

$$\left|\int_{[0,1)} \frac{1}{(1-\alpha)N}\sum_{\phi^{x_N}(\xi^j)>\kappa^\alpha_{N,x_N}} g(\xi^j)d\mu_N(\alpha) - \int_{[0,1)} \frac{1}{1-\alpha}\int_{\phi^{\bar{x}}(\xi)>\kappa_\alpha} g(s)d\mathbb{P}(s)d\mu_N(\alpha)\right| \to 0.$$

Note that $\phi^{\bar{x}}(\xi)$ is Lipschitz continuous w.r.t. $x$, the given $g$ is continuous and bounded function w.r.t. $\xi$ and $\{\xi^j\}_{j=1}^N$ is i.i.d. samples from $\mathbb{P}$, both $\frac{1}{(1-\alpha)N}\sum_{\phi^{x_N}(\xi^j)>\kappa^\alpha_{N,x_N}} g(\xi^j)$ and $\frac{1}{1-\alpha}\int_{\phi^{\bar{x}}(\xi)>\kappa_\alpha} g(s)d\mathbb{P}(s)$ are bounded by $\max_{s\in[0,1]} g(s)$ and by Proposition 6

$$\lim_{N\to\infty}\left|\frac{1}{(1-\alpha)N}\sum_{\phi^{x_N}(\xi^j)>\kappa^\alpha_{N,x_N}} g(\xi^j) - \frac{1}{1-\alpha}\int_{\phi^{\bar{x}}(\xi)>\kappa_\alpha} g(s)d\mathbb{P}(s)\right| = 0.$$

We then have

$$
\begin{aligned}
&\lim_{N\to\infty}\left|\int_{[0,1)}\frac{1}{(1-\alpha)N}\sum_{\phi^{x_N}(\xi^j)>\kappa^{\alpha}_{N,x_N}} g(\xi^j)d\mu_N(\alpha)-\int_{[0,1)}\frac{1}{1-\alpha}\int_{\phi^{\bar{x}}(\xi)>\kappa_\alpha} g(s)d\mathbb{P}(s)d\mu_N(\alpha)\right|\\
&\le \lim_{N\to\infty}\int_{[0,1)}\left|\frac{1}{\alpha N}\sum_{\phi^{x_N}(\xi^j)>\kappa^{\alpha}_{N,x_N}} g(\xi^j)-\int_{[0,1)}\frac{1}{\alpha}\int_{\phi^{\bar{x}}(\xi)>\kappa_\alpha} g(s)d\mathbb{P}(s)\right|d\mu_N(\alpha)\\
&\le \lim_{N\to\infty}\int_{[0,1)}\left|\frac{1}{\alpha N}\sum_{\phi^{x_N}(\xi^j)>\kappa^{\alpha}_{N,x_N}} g(\xi^j)-\int_{[0,1)}\frac{1}{\alpha}\int_{\phi^{\bar{x}}(\xi)>\kappa_\alpha} g(s)d\mathbb{P}(s)\right|d\hat{\mu}(\alpha)\\
&=0,
\end{aligned}
$$

where the second inequality is from condition (iii) and the third equality is from Lebesgue's dominated convergence theorem.

Combining the above analysis, we have (A10), that is $P_N^{x_N}$ converges weakly to $P^{\bar{x}}$. $\square$

**Proof of Theorem 2.** By conditions (c) - (e),

$$P^{\bar{x}} \in \arg\max_{Q\in\mathfrak{M}} \mathbb{E}_Q[\phi(\bar{x},\xi)].$$

Then we only need to prove that $\bar{x}$ is a solution of (22), which is equivalent to

$$0 \in \mathbb{E}_{P^{\bar{x}}}[\Phi(\bar{x},\xi)] + \mathcal{N}_X(\bar{x}). \tag{A12}$$

Since $\hat{x}_N \to \bar{x}$,

$$\limsup_{N\to\infty} \mathcal{N}_X(\hat{x}_N) \subset \mathcal{N}_X(\bar{x}). \tag{A13}$$

Moreover,

$$
\begin{aligned}
\|\mathbb{E}_{P_N^{\hat{x}_N}}[\Phi(\hat{x}_N,\xi)] - \mathbb{E}_{P^{\bar{x}}}[\Phi(\bar{x},\xi)]\| \le{}& \|\mathbb{E}_{P_N^{\hat{x}_N}}[\Phi(\hat{x}_N,\xi)] - \mathbb{E}_{P^{\bar{x}}}[\Phi(\hat{x}_N,\xi)]\|\\
&+ \|\mathbb{E}_{P^{\bar{x}}}[\Phi(\hat{x}_N,\xi)] - \mathbb{E}_{P^{\bar{x}}}[\Phi(\bar{x},\xi)]\|.
\end{aligned}
$$

Note that since $P_N^{\hat{x}_N} \to P^{\bar{x}}$ weakly and by Assumption 3 (b), $P_N^{\hat{x}_N} \to P^{\bar{x}}$ under Wasserstein metric [23]. Then by condition (b), we have for any $N$, $\Phi(\hat{x}_N,\cdot)$ is Lipschitz continuous and

$$\lim_{N\to\infty}\sup_{z\in\bar{Z}} \|\mathbb{E}_{P_N^{\hat{x}_N}}[\Phi(z,\xi)] - \mathbb{E}_{P^{\bar{x}}}[\Phi(z,\xi)]\| = 0, \tag{A14}$$

where $\bar{Z} := \{\hat{x}_N, N = 1, 2, \cdots\}$. Moreover,

$$\begin{aligned}\lim_{N\to\infty} \|\mathbb{E}_{P^{\bar{x}}}[\Phi(\bar{x},\xi)] - \mathbb{E}_{P^{\bar{x}}}[\Phi(\hat{x}_N,\xi)]\| &\le \lim_{N\to\infty} \mathbb{E}_{P^{\bar{x}}}[\kappa(\xi)]\|\bar{x} - \hat{x}_N\| \\ &\le \lim_{N\to\infty} \sup_{P\in\hat{\mathfrak{M}}} \mathbb{E}_P[\kappa(\xi)]\|\bar{x} - \hat{x}_N\| \\ &= 0.\end{aligned} \tag{A15}$$

Combining (A14)-(A15), we have

$$\lim_{N\to\infty} \|\mathbb{E}_{P_N^{\hat{x}_N}}[\Phi(\hat{x}_N,\xi)] - \mathbb{E}_{P^{\bar{x}}}[\Phi(\bar{x},\xi)]\| = 0, \tag{A16}$$

which implies (A12). □