ENTROPIC APPROXIMATION FOR MATHEMATICAL PROGRAMS WITH ROBUST EQUILIBRIUM CONSTRAINTS

YONGCHAO LIU* AND HUIFU XU†

Abstract. In this paper, we consider a class of mathematical programs with robust equilibrium constraints represented by a system of semi-infinite complementarity constraints (SICC). We propose a numerical scheme for tackling SICC. Specifically, by relaxing the complementarity constraints and then randomizing the index set of SICC, we employ the well-known entropic risk measure to approximate the semi-infinite constraints with a finite number of stochastic inequality constraints. Under some moderate conditions, we quantify the approximation in terms of the feasible set and the optimal value. The approximation scheme is then applied to a class of two stage stochastic mathematical programs with complementarity constraints in combination with the polynomial decision rules. Finally, we extend the discussion to a mathematical program with distributionally robust equilibrium constraints which is essentially a one stage stochastic program with semi-infinite stochastic constraints indexed by some probability measures from an ambiguity set defined through the KL-divergence.

Key words. Entropic risk measure, robust equilibrium constraints, polynomial decision rule, stability analysis, KL-divergence.

AMS subject classifications. 90C15, 90C30, 90C33.

1. Introduction. Consider the following mathematical program with semi-infinite complementarity constraints

$$\text{(MPSICC)} \qquad \begin{array}{ll} \min\limits_{x\in X} & f(x) \\ \text{s.t.} & 0 \le x \perp g(x,t) \ge 0, \quad \forall t \in T, \end{array} \tag{1.1}$$

where X is a nonempty compact set of $\mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \times T \to \mathbb{R}^n$ are continuous functions, T is a compact index set and $a \perp b$ denotes orthogonality of vectors a and b. In practical applications, the complementarity constraints are often used to describe an equilibrium arising from economic competition or traffic flow, whereas the index parameter t may represent various uncertainties such as market demand, economic or environmental conditions. Consequently MPSICC may be called a mathematical program with *robust* complementarity constraints where the robustness is in the sense that the constraints must hold for every realization of the uncertain parameter. In the case when T is a finite set, (1.1) is known as a mathematical program with complementarity constraints (MPCC) which has been extensively investigated over the past two decades, see monographs [28, 30] for a comprehensive treatment of the topic.

Our interest in this paper is on the case when T is an infinite set. To ease the exposition, we assume that T is a compact set of a finite dimensional space although most of our technical results hold when T is a subset of a Banach space. The MPSICC may be viewed as an extension of the mathematical programs with robust equality and/or inequality constraints which has been investigated extensively over the past two decades [1, 5, 6, 7, 8, 12, 13, 14, 16, 20, 45]. For instance, in multiload truss optimization, mechanical response of a truss is often represented by a robust linear system of equations parameterized by nodal displacement vector, which is a special case of robust complementarity constraints.

Our focus in this paper is on the numerical methods for solving MPSICC. Note that the optimization problems with robust constraints in [12, 2] are convex while problem (1.1) is typically nonconvex because of combinatorial structure of the complementarity constraints. Moreover, the feasible set of the problem does not have an interior which means that a direct application of a randomization approach is often numerically unstable. To address the issue, we employ the well known NLP-regularization scheme (see [36]) to relax the complementarity constraints by replacing them with a parameterized semi-infinite system of inequalities:

$$\text{(RMPSICC)} \qquad \begin{array}{ll} \min\limits_{x\in X_+} & f(x) \\ \text{s.t.} & \left.\begin{array}{l} -g(x,t) \le 0 \\ x \circ g(x,t) \le \tau e \end{array}\right\} \forall t \in T, \end{array} \tag{1.2}$$

*Department of Mathematics, Dalian Maritime University, Dalian 116026, China (googleliu@dlmu.edu.cn). The work of this author is supported by NSFC 11201044 and FRFCU 3132014037.

†School of Engineering and Mathematical Sciences, City University London, EC1V 0HB, London, UK (Huifu.Xu.1@city.ac.uk). The work of this author is supported by EPSRC grant EP/J014427/1.

where $X_+ := X \cap \mathbb{R}^n_+$, $\mathbb{R}^n_+$ denotes the set of n-dimensional vectors whose components are nonnegative, $\tau \geq 0$ is a fixed positive parameter, $e \in \mathbb{R}^n$ is a vector with components 1 and "$\circ$" denotes the Hadamard product. The regularized problem MPSICC has at least two advantages from numerical point of view: (a) it involves ordinary inequality constraints, and (b) its feasible solution set is more likely to have an interior as opposed to RMPSICC. Under some moderate conditions, RMPSICC approximates MPSICC as τ is driven to zero. We will come back to the details of this in Section 3. For a fixed parameter τ, we propose a randomization approach based on the entropic risk measure to solve RMPSICC. Specifically, by treating t as a random parameter[1], we use the entropic risk measure of $g(x,t)$ and $x \circ g(x,t)$ to approximate the semi-infinite constraints of RMPSICC.

Much of the rest of the work is on the theoretical analysis of the entropic approximation of RMPSICC as the entropic risk measure parameter varies in terms of the optimal value and the optimal solutions (Section 3). Moreover, we propose a sample average approximation (SAA for short) for the entropic risk measure and investigate the convergence of the optimal value and the optimal solutions as the sample size increases (Section 3). As an application, we apply the proposed numerical schemes to a two stage stochastic mathematical program with complementarity constraints in a combination with the well known polynomial decision rule method [4] (Section 4) and to a stochastic mathematical program with distributionally robust equilibrium constraints where the distributional set is constructed through a nominal sample average approximated distribution within a range specified through the Kullback-Leibler divergence (Section 5). As far as we are concerned, the main contributions of the paper can be summarized as follows:

- Entropic approximation for MPSICC.

$$\text{MPSICC(1.1)} \xrightarrow{\text{Reg}} \text{RMPSICC(1.2)} \xrightarrow[\text{SAA}]{\text{Entropic Appr}} \text{SAA-EA-RMPSICC(3.4)}.$$

  We propose a new optimization model with robust complementarity constraints (1.1) which extends the existing optimization models with robust inequality and/or equality constraints. We develop a numerical scheme for solving the problem (1.1) which utilizes the entropic risk measure and sample average approximation, and carry out qualitative and quantitative stability analysis of the approximation scheme in terms of the optimal value and the optimal solutions. Differing from the mainstream approaches in distributionally robust optimization, the approximation scheme does not require dualization for transforming the semi-infinite constraint into a semi-definite constraint and hence can be applied to a nonconvex function $g(x,t)$. It also differs from Calafiore and Campi's randomization method [12, 13] in that the resulting optimization problem does not increase the number of constraints as the sample size increases, and the use of the entropic risk measure may lead to stability of the optimal values.
- Entropic approximation for two stage SMPECs.

$$\text{SMPEC(4.1)} \xrightarrow[\text{PDL}]{\text{Reg}} \text{PDL-RSMPEC(4.3)} \xrightarrow[\text{SAA}]{\text{Entropic Appr}} \text{SAA-EA-PDL-RSMPEC(4.6)}.$$

  By applying the numerical scheme in a combination with the polynomial decision rule (PDL) [4] and the sample average approximation method to a two stage stochastic mathematical program with complementarity constraints (SMPEC), we provide an alternative method to tackle two stage SMPECs which are often notoriously difficult to solve. A significant advantage of the new framework is that it does not scale in the number of the sample size.
- Entropic approximation for one stage mathematical programs with distributionally robust equilibrium constraints (MPDRE).

$$\text{MPDRE(5.1)} \xrightarrow{\text{Reg}} \text{RMPDRE(5.3)} \xrightarrow[\text{SAA}]{\text{Entropic Appr}} \text{SAA-EA-EMPDRE(5.7)}.$$

  The MPDRE model provides a mathematical tool for the study of robust equilibria. It has potential applications in market design where equilibria withstand uncertain economic circumstances. The new model and the numerical scheme extend the recent research on optimization problems with distributionally robust inequality constraints by Hu and Hong [21].

Throughout the paper, we will use the following notation. We denote by $\|\cdot\|$ the Euclidean norm of a vector, $d(x,D)$ the distance from a point x to a set D and $\text{Diam}(D)$ the diameter of the set D, that is, $d(x,D) := \inf_{x' \in D} \|x - x'\|$ and $\text{Diam}(D) := \sup_{x',x'' \in D} \|x' - x''\|$. For two compact sets D_1 and D_2, we write $\mathbb{D}(D_1, D_2)$ the

[1]Even if t is a random parameter in practical applications, there is usually no information on the distribution of t in this formulation. Indeed, the constraints depend only on the support set of t rather than its distribution.

deviation of D_1 from D_2, that is, $\mathbb{D}(D_1, D_2) := \sup_{x\in D_1} d(x, D_2)$ and $\mathbb{H}(D_1, D_2) := \max(\mathbb{D}(D_1, D_2), \mathbb{D}(D_2, D_1))$ for the Hausdorff distance between D_1 and D_2. Finally, we use standard notation cl$D$ for the closure of set D and $\overline{\lim}_{k\to\infty} D_k$ for the outer limit of a sequence of sets $\{D_k\}$.

2. Entropic risk measure approximation. Let $\mathbb{M}$ denote the linear space of bounded measurable functions defined on some measurable space $(\Omega, \mathcal{F})$. Consider a set $\mathcal{A} \subset \mathbb{M}$ such that

$$Z \in \mathcal{A}, U \in \mathbb{M}, U \geq Z \Longrightarrow U \in \mathcal{A}.$$

Define the functional $\rho : \mathbb{M} \to \mathbb{R}$ by $\rho(Z) := \inf\{m \in \mathbb{R} : Z + m \in \mathcal{A}\}$. It is easy to verify that $\rho(Z_1) \leq \rho(Z_2)$, for $Z_1 \geq Z_2$ and $\rho(Z+m) = \rho(Z)-m$. In the literature of finance risks, $\rho(\cdot)$ is known as a monetary risk measure when Z represents a financial position, e.g., capital, see [18]. Let γ be a positive number and $u(z) = 1 - e^{-\gamma z}$ be an exponential utility function. If we define $\mathcal{A} := \{Z \in L^\infty : |\mathbb{E}_P[u(Z)]| \geq u(0)\}$, then the resulting risk measure is

$$e_\gamma(Z) := \rho(Z) = \frac{1}{\gamma} \ln \mathbb{E}_P[e^{-\gamma Z}] \tag{2.1}$$

for $Z \in L^\infty$. The risk measure has a robust representation

$$e_\gamma(Z) = \sup_{Q\in\mathcal{M}_1} \{\mathbb{E}_Q[Z] - \frac{1}{\gamma} H(Q|P)\},$$

where $\mathcal{M}_1$ denotes the set of all probability measures on $\mathbb{M}$, and

$$H(Q|P) := \begin{cases} \mathbb{E}_Q\left[\ln \frac{dQ}{dP}\right], & \text{for } Q \succeq P \\ +\infty, & \text{otherwise} \end{cases} \tag{2.2}$$

denotes the relative entropy of Q with respect to P, $Q \succeq P$ means Q is absolutely continuous with respect to P. Consequently $e_\gamma(Z)$ is called *entropic risk measure*, see [17] for a thorough treatment of the concept. It is well known (see e.g. [18, formulation (4)]) that $e_\gamma(Z)$ is monotonically increasing in γ and

$$\lim_{\gamma\to\infty} e_\gamma(Z) = \operatorname{ess\,sup}(-Z) \tag{2.3}$$

in the case that $\operatorname{ess\,inf} Z > -\infty$. Our focus in this paper is mainly on the limit rather than the financial background of the risk measure.

Let $h : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$ be a continuous function and X be a subset of $\mathbb{R}^n$. Let $\xi : \Omega \to \mathbb{R}^k$ be a random variable defined on the probability space $(\Omega, \mathcal{F}, P)$ with support set Ξ. Our interest is the uniformity of limit (2.3) when $z = h(x, \xi)$ w.r.t. x. We address this in the following proposition.

Proposition 2.1. **(Entropic approximation of random functions)** *Let $H(x)$, $F_x(t)$ and Ω_x denote respectively the essential supremum, the cumulative distribution function and the support set of $-h(x, \xi)$. Let* Diam(Ω_x) *denote the diameter of the support set Ω_x which is the distance between $H(x)$ and essential infimum of $-h(x, \xi)$. Assume: (a) $X \subset \mathbb{R}^n$ be a compact set, (b) for each fixed $x \in X$,*

$$\inf_{\xi\in\Xi} h(x, \xi) > -\infty. \tag{2.4}$$

Then for each fixed $x \in X$,

$$\lim_{\gamma\to\infty} e_\gamma(h(x, \xi)) = H(x). \tag{2.5}$$

Assume in addition that (c)

$$\inf_{x\in X} \inf_{\xi\in\Xi} h(x, \xi) > -\infty, \tag{2.6}$$

and (d) for any fixed small positive number ϵ, there exists $\delta(\epsilon) \in (0, 1)$ such that

$$1 - F_x(H(x) - \epsilon) \geq \delta(\epsilon), \quad \forall x \in X_\epsilon, \tag{2.7}$$

where $X_\epsilon := \{x \in X : \mathrm{Diam}(\Omega_x) > 2\epsilon\}$. *Then*

$$|e_\gamma(h(x,\xi)) - H(x)| < 2\epsilon + \frac{1}{\gamma}\left|\ln \delta(\epsilon)\right|. \tag{2.8}$$

Proof. The convergence (2.5) is well known, see the comments following [18, Formulation (4)]. Here, we provide a proof for completeness. We proceed the proof in two steps according to the distribution of ξ.

Step 1. ξ follows a discrete distribution, that is, $\mathrm{Prob}\left\{\xi = \xi^j\right\} = p_j,\ j = 1, 2, \cdots$. This includes both infinite and finite distribution (in which case $p_j = 0$ when $j \geq k$ for some k). Let x be fixed and ϵ be a fixed small positive number. Let $\mathcal{J}_\epsilon(x) := \min\{j : -h(x,\xi^j) \geq H(x) - \epsilon\}$. Then

$$e_\gamma(h(x,\xi)) \geq \frac{1}{\gamma}\ln\left(\sum_{j\in\mathcal{J}_\epsilon(x)} e^{\gamma(-h(x,\xi^j))}p_j\right) = H(x) + \frac{1}{\gamma}\ln\left(\sum_{j\in\mathcal{J}_\epsilon(x)} e^{\gamma(-h(x,\xi^j)-H(x))}p_j\right).$$

Since $-h(x,\xi^j) - H(x) \geq -\epsilon$ for $j \in \mathcal{J}_\epsilon(x)$, then

$$0 > \frac{1}{\gamma}\ln\left(\sum_{j\in\mathcal{J}_\epsilon(x)} e^{\gamma(-h(x,\xi^j)-H(x))}p_j\right) > -\epsilon + \frac{1}{\gamma}\ln\left(\sum_{j\in\mathcal{J}_\epsilon(x)} p_j\right). \tag{2.9}$$

Note that for fixed x, $\sum_{j\in\mathcal{J}_\epsilon(x)} p_j > 0$. Therefore, by driving γ to ∞, we obtain

$$0 > \lim_{\gamma\to\infty}\frac{1}{\gamma}\ln\left(\sum_{j\in\mathcal{J}_\epsilon(x)} e^{\gamma(-h(x,\xi^j)-H(x))}p_j\right) \geq -\epsilon,$$

which implies (2.5) because ϵ can be arbitrarily small.

To show (2.8), we note that under condition (2.7),

$$\frac{1}{\gamma}\left|\ln\left(\sum_{j\in\mathcal{J}_\epsilon(x)} p_j\right)\right| \leq \frac{1}{\gamma}\left|\ln\delta(\epsilon)\right|, \quad \forall x \in X_\epsilon,$$

where $X_\epsilon \neq \emptyset$. Through (2.9), the inequality above implies (2.8). When $x \in X \setminus X_\epsilon$, $\mathrm{Diam}(\Omega_x) \leq 2\epsilon$. Since

$$\mathbb{E}[-h(x,\xi)] \leq e_\gamma(h(x,\xi)) \leq H(x), \quad \forall\gamma > 0,$$

then

$$|e_\gamma(h(x,\xi)) - H(x)| \leq 2\epsilon < 2\epsilon + \frac{1}{\gamma}|\ln\delta(\epsilon)|. \tag{2.10}$$

Summarising the two cases, we have shown that (2.8) holds for all $x \in X$ under condition (2.7).

Step 2. ξ follows a continuous distribution [2]. Observe first that (2.10) holds regardless of the distribution of ξ. Therefore, we only need to consider the case when $x \in X_\epsilon$. Recall that $F_x(t)$ is the cumulative distribution function of $-h(x,\xi)$. It is easy to verify that

$$\mathbb{E}[e^{-\gamma h(x,\xi)}] = e^{\gamma H(x)}\left[1 - \gamma\int_{-\infty}^{H(x)} e^{\gamma(t-H(x))}F_x(t)dt\right].$$

Therefore

$$\frac{1}{\gamma}\ln\left(\mathbb{E}[e^{-\gamma h(x,\xi)}]\right) = H(x) + \frac{1}{\gamma}\ln\left(1 - \gamma\int_{-\infty}^{H(x)} F_x(t)e^{\gamma(t-H(x))}dt\right).$$

[2]We gratefully acknowledge that this part of the proof follows from a private communication with Thomas Knispel.

Let ϵ be a small positive number. Then

$$\gamma \int_{-\infty}^{H(x)} F_x(t) e^{\gamma(t-H(x))} dt = \gamma \int_{-\infty}^{H(x)-\epsilon} F_x(t) e^{\gamma(t-H(x))} dt + \gamma \int_{H(x)-\epsilon}^{H(x)} F_x(t) e^{\gamma(t-H(x))} dt$$
$$\leq F_x(H(x)-\epsilon) e^{-\gamma\epsilon} + 1 - e^{-\gamma\epsilon},$$

where the inequality follows from the monotonicity of the cumulative distribution function and the exponential function. Therefore

$$\left| \frac{1}{\gamma} \ln\left(\mathbb{E}[e^{-\gamma h(x,\xi)}] \right) - H(x) \right| < \left| \frac{1}{\gamma} \ln(1 - (1 - e^{-\gamma\epsilon}) - F_x(H(x)-\epsilon) e^{-\gamma\epsilon}) \right|$$
$$= \epsilon + \frac{1}{\gamma} |\ln(1 - F_x(H(x)-\epsilon))|.$$

For fixed x, since $1 - F_x(H(x)-\epsilon) > 0$, we arrive at (2.5) by driving γ to infinity and then ϵ to zero. The error bound (2.8) follows from the inequality above under condition (2.7). ■

Note that condition (2.7) is similar to the so called consistent tail behaviour condition for CVaR approximation of the essential superemum of a random function in [2]. Indeed, the latter implies the former. We explain how the condition may be satisfied through a simple example varied from [2, Example 1].

EXAMPLE 2.1. *Consider $h(x,\xi) = -x\xi$, where $x \in [0,1] \subset \mathbb{R}$ and ξ follows a uniform distribution over interval $[-1,1]$. Let $\epsilon < 1$ be a small positive number. Then $H(x) = x$, $\Omega_x = [-x, x]$ and $X_\epsilon = \{x \in [0,1] : 2\epsilon < 2x\} = (\epsilon, 1]$. It is easy to derive that*

$$1 - F_x(H(x)-\epsilon) = \frac{\epsilon}{2x} \geq \frac{\epsilon}{2}, \qquad \forall x \in X_\epsilon.$$

This shows that condition (2.7) holds with $\delta(\epsilon) = \frac{\epsilon}{2}$. Let us now consider the case when $x \in [0,1] \setminus X_\epsilon = [0,\epsilon]$. Then $\mathrm{Diam}(\Omega_x) = 2x \leq 2\epsilon$ and hence $|e_\gamma(h(x,\xi)) - H(x)| \leq 2\epsilon$.

Proposition 2.1 says that under some moderate conditions, the entropic risk measure of $h(x,\xi)$ converges to the essential supremum of $-h(x,\xi)$ uniformly w.r.t. x as $\gamma \to \infty$. Under the similar condition, Anderson et al. [2] presented some conditions which ensure the CVaR of a random function converges uniformly to its essential supremum. The main differences are two fold: (a) CVaR utilizes the tail distribution whereas entropic approximation uses the whole distribution with more weights on the tail when γ is large. (b) CVaR is nonsmooth as it only uses the tail distribution while the entropic risk measure is smooth.

3. Stability analysis. In this section, we use the entropic risk measure as the workhorse to construct an approximation of the robust semi-infinite constraints of problem (1.2). We examine the accuracy and efficiency of the approximation from a stability point of view.

3.1. Stability w.r.t. parameters τ and γ. Let us start by writing problem (1.2) equivalently as

$$\text{(RMPSICC)} \qquad \begin{array}{ll} \min\limits_{x \in X_+} & f(x) \\ \text{s.t.} & \sup\limits_{t \in T} x \circ g(x,t) \leq \tau e, \\ & \sup\limits_{t \in T} -g_j(x,t) \leq 0, \qquad \text{for } j \in \{1, \cdots, n\}, \end{array} \tag{3.1}$$

where $X_+ := X \cap \mathbb{R}^n_+$. The proposition below states the approximation of RMPSICC (3.1) to MPSICC (1.1) in terms of the optimal value and the optimal solutions as τ is driven to 0.

PROPOSITION 3.1. **(Stability of RMPSICC (3.1))** *Let $X(\tau)$ and X^* denote the sets of the optimal solutions to the problems (3.1) and (1.1) respectively. Let $v(\tau)$ and v^* denote the corresponding optimal values. Then $\overline{\lim}_{\tau \to 0} X(\tau) \subset X^*$ and $\lim_{\tau \to 0} v(\tau) = v^*$.*

Proof. Let $\mathcal{F}(\tau)$ and $\mathcal{F}$ denote the feasible sets of the problems (3.1) and (1.1) respectively. Since $\mathcal{F} \subset \mathcal{F}(\tau)$, it follows by [43, Lemma 4.2(i)], $\lim_{\tau \to 0} \mathbb{H}(\mathcal{F}(\tau), \mathcal{F}) = 0$. Let $x_\tau \in X(\tau)$ and $y^* \in X^*$. Assume for the simplicity of exposition (by taking a subsequence if necessary) that $x_\tau \to x^*$. Since $x^* \in \mathcal{F}$ and $y^* \in \mathcal{F}(\tau)$, then

$$f(y^*) \leq f(x^*) = \lim_{\tau \to 0} f(x_\tau) \leq f(y^*).$$

This shows $f(x^*) = f(y^*)$ hence $x^* \in X^*$. The convergence of the optimal value also follows. ■

In what follows, we regard t as a random variable with support set T and approximate the supermum with the entropic risk measure, namely,

$$\text{(EA-RMPSICC)} \qquad \begin{array}{ll} \min\limits_{x\in X_+} & f(x) \\ \text{s.t.} & \left.\begin{array}{l} e_\gamma^j(x) \le 0 \\ \bar{e}_\gamma^j(x) \le \tau \end{array}\right\} \text{for } j \in \{1,\cdots,n\}, \end{array} \tag{3.2}$$

where γ is a fixed positive parameter,

$$e_\gamma^j(x) := e_\gamma(g_j(x,t)), \qquad \bar{e}_\gamma^j(x) := e_\gamma(-x_j \cdot g_j(x,t)),$$

and $e_\gamma(\cdot)$ is defined in equation (2.1). Compared to RMPSICC (3.1), EA-RMPSICC (3.2) consists of two ordinary stochastic constraints where the underlying functions are continuously differentiable and can be approximated through sampling. Moreover, the stability results to be established in this section do not depend on the probability distribution of t, which means t can be any random variable whose support set is T. Our focus in this section is to look into the approximation of the optimal value and the optimal solutions of EA-RMPSICC (3.2) to those of RMPSICC (3.1) as γ increases. Observe that the two problems have identical objective functions. Therefore it suffices to investigate the approximation of feasible constraints/solutions and its impact on the optimal value and the optimal solutions.

Let $\mathcal{F}(\tau)$ and $\mathcal{F}_\gamma(\tau)$ denote the feasible solution sets of problems (3.1) and (3.2) respectively. Obviously $\mathcal{F}(\tau) \subseteq \mathcal{F}_\gamma(\tau)$. In other words, $\mathcal{F}_\gamma(\tau)$ provides an outer bound for $\mathcal{F}(\tau)$. The following proposition gives a quantitative description of the excess of $\mathcal{F}_\gamma(\tau)$ over $\mathcal{F}(\tau)$.

PROPOSITION 3.2. **(Error bound of the feasible solution set of EA-RMPSICC (3.2))** *Let $F_x^{1j}(t)$ and $F_x^{2j}(t)$ denote the cumulative distributions of $g_j(x,t)$ and $x_j g_j(x,t)$ respectively for $j = 1,\cdots,n$. Assume there exists a positive constant C such that*

$$d(x,\mathcal{F}(\tau)) \le C\sum_{j=1}^{n}\left(\left(\sup_{t\in T} -g_j(x,t)\right)_+ + \left(\sup_{t\in T} x_j\cdot g_j(x,t) - \tau\right)_+\right) \tag{3.3}$$

and for any fixed positive number ϵ, there exists $\delta(\epsilon) \in (0,1)$ such that for $j = 1,\cdots,n$,

$$1 - F_x^{1j}\left(\sup_{t\in T} g_j(x,t) - \epsilon\right) \ge \delta(\epsilon), \ \ \forall x \in X_\epsilon^{1j}$$
$$1 - F_x^{2j}\left(\sup_{t\in T} x_j g_j(x,t) - \epsilon\right) \ge \delta(\epsilon), \ \ \forall x \in X_\epsilon^{2j},$$

where $X_\epsilon^{1j} := \{x \in X_+ : \mathrm{Diam}(\Omega_x^{1j}) > 2\epsilon\}$, $X_\epsilon^{2j} := \{x \in X_+ : \mathrm{Diam}(\Omega_x^{2j}) > 2\epsilon\}$ and Ω_x^{1j} and Ω_x^{2j} are the support sets of random variables $g_j(x,\xi)$ and $x_j g_j(x,\xi)$ respectively. Then

(i) for any $\epsilon^* > 0$, there exists a positive number γ_{ϵ^*} such that $\mathbb{H}(\mathcal{F}_\gamma(\tau),\mathcal{F}(\tau)) \le \epsilon^*$ when $\gamma \ge \gamma_{\epsilon^*}$;
(ii) $\mathbb{H}(\mathcal{F}_\gamma(\tau),\mathcal{F}(\tau)) \le C\Delta_\gamma$, where $\Delta_\gamma := 4n\epsilon + \frac{2n}{\gamma}|\ln\delta(\epsilon)|$, and n is the dimension of variable x.

The condition (3.3) is an error bound for the system of inequalities in the constraints. This type of conditions has been well studied in the past decade, see survey papers [3, 31] for more details.

Proof. Part (i). Let ϵ^* be a fixed small positive number. Define

$$R_j(\epsilon^*) := \inf_{\substack{x\in X_+ \\ d(x,\,\mathcal{F}(\tau))\ge\epsilon^*}} \sup_{t\in T} -g_j(x,t),$$
$$\bar{R}_j(\epsilon^*) := \inf_{\substack{x\in X_+ \\ d(x,\,\mathcal{F}(\tau))\ge\epsilon^*}} \sup_{t\in T} x_j\cdot g_j(x,t) - \tau,$$

and $R(\epsilon^*) := \max\limits_{j\in\{1,\cdots,n\}} \max\{R_j(\epsilon^*), \bar{R}_j(\epsilon^*)\}$. Let $\delta := R(\epsilon^*)/2$. Then $\delta > 0$. By Proposition 2.1, $e_\gamma^j(x)$ and $\bar{e}_\gamma^j(x)$, converge to $\sup_{t\in T} -g_j(x,t)$ and $\sup_{t\in T} x_j\cdot g_j(x,t)$ uniformly on compact set X_+ as $\gamma \to \infty$. Therefore,

there exists a sufficiently large γ_{ϵ^*} such that

$$\sup_{\substack{x\in X_+\\ j\in\{1,\cdots,n\}}} \left\{\sup_{t\in T} -g_j(x,t) - e^j_\gamma(x)\right\} \le \delta$$

and

$$\sup_{\substack{x\in X_+\\ j\in\{1,\cdots,n\}}} \left\{\sup_{t\in T} x_j\cdot g_j(x,t) - \bar{e}^j_\gamma(x)\right\} \le \delta,$$

when $\gamma \ge \gamma_{\epsilon^*}$.

Let $x \in X_+$ be such that $d(x, \mathcal{F}(\tau)) \ge \epsilon^*$ and $\gamma \ge \gamma_{\epsilon^*}$. There exists $j \in \{1,\cdots,n\}$ such that at least one of the following inequalities holds

$$\begin{aligned}\bar{e}^j_\gamma(x) - \tau &= \sup_{t\in T} x_j\cdot g_j(x,t) - \tau + \bar{e}^j_\gamma(x) - \sup_{t\in T} x_j\cdot g_j(x,t) \ge R(\epsilon^*) - R(\epsilon^*)/2 = R(\epsilon^*)/2 > 0,\\ e^j_\gamma(x) &= \sup_{t\in T} -g_j(x,t) + e^j_\gamma(x) - \sup_{t\in T} -g_j(x,t) \ge R(\epsilon^*) - R(\epsilon^*)/2 = R(\epsilon^*)/2 > 0,\end{aligned}$$

which means $x \notin \mathcal{F}_\gamma(\tau)$, or equivalently, $d(x, \mathcal{F}(\tau)) < \epsilon^*$ for every $x \in \mathcal{F}_\gamma(\tau)$. This shows $\mathbb{D}(\mathcal{F}_\gamma(\tau), \mathcal{F}(\tau)) \le \epsilon^*$ and hence $\mathbb{H}(\mathcal{F}(\tau), \mathcal{F}_\gamma(\tau)) \le \epsilon^*$ given $\mathcal{F}(\tau) \subseteq \mathcal{F}_\gamma(\tau)$.

Part (ii). Observe that $\mathcal{F}(\tau) \subseteq \mathcal{F}_\gamma(\tau)$. Hence, it suffices to prove $\mathbb{D}(\mathcal{F}_\gamma(\tau), \mathcal{F}(\tau)) \le C\Delta_\gamma$. Let $\hat{x} \in \mathcal{F}_\gamma(\tau)$. For $j = 1,\cdots,n$, $e^j_\gamma(\hat{x}) \le 0$ and $\bar{e}^j_\gamma(\hat{x}) - \tau \le 0$. Through (2.8), we have

$$\left|\sup_{t\in T} -g_j(\hat{x},t) - e^j_\gamma(\hat{x})\right| \le 2\epsilon + \frac{1}{\gamma}\left|\ln(1 - F^{1j}_{\hat{x}}(\sup_{t\in T} -g_j(\hat{x},t) - \epsilon))\right|$$

and

$$\left|\sup_{t\in T} \hat{x}_j\cdot g_j(\hat{x},t) - \bar{e}^j_\gamma(\hat{x})\right| \le 2\epsilon + \frac{1}{\gamma}\left|\ln(1 - F^{2j}_{\hat{x}}(\sup_{t\in T} \hat{x}_j\cdot g_j(\hat{x},t) - \epsilon))\right|.$$

By exploiting the inequalities above and the condition (3.3), we arrive at

$$\begin{aligned} d(\hat{x}, \mathcal{F}(\tau)) &\le C\sum_{j=1}^n \left(\left(\sup_{t\in T} -g_j(\hat{x},t)\right)_+ + \left(\sup_{t\in T} x_j\cdot g_j(\hat{x},t) - \tau\right)_+\right)\\ &\le C\left(\sum_{j=1}^n \left(\sup_{t\in T} -g_j(\hat{x},t) - e^j_\gamma(\hat{x})\right)_+ + \left(\sup_{t\in T} x_j\cdot g_j(\hat{x},t) - \bar{e}^j_\gamma(\hat{x})\right)_+\right)\\ &\le 4n\epsilon + \frac{2n}{\gamma}\ln|\delta(\epsilon)|.\end{aligned}$$

The proof is complete. ■

Proposition 3.2 says that the feasible set mapping of EA-RMPSICC (3.2) converges to the feasible set of RMPSICC (3.1) as $\gamma \to \infty$. Using this property, we can establish the convergence of the optimal value and the optimal solutions.

THEOREM 3.1. **(Stability of EA-RMPSICC (3.2))** *Let $X_\gamma(\tau)$ and $X(\tau)$ denote the sets of optimal solutions of the problems (3.2) and (3.1) respectively, $v_\gamma(\tau)$ and $v(\tau)$ be the corresponding optimal values. Assume the conditions of Proposition 3.2. Then*

(i) $\lim_{\gamma\to\infty} v_\gamma(\tau) = v(\tau)$ and $\overline{\lim}_{\gamma\to\infty} X_\gamma(\tau) \subseteq X(\tau)$;
(ii) if, in addition, $f(x)$ is Lipschitz continuous with modulus L, then $|v_\gamma(\tau) - v(\tau)| \le LC\Delta_\gamma$, where Δ_γ and C are given as in Proposition 3.2.

Proof. Part (i). It is sufficient to show the convergence of the optimal solutions as $f(\cdot)$ is continuous and independent of γ. Assume a contradiction that there exists $x^N \in X_{\gamma_N}(\tau)$ such that $x^N \to x^*$ and $x^* \notin X(\tau)$, where $\gamma_N \to +\infty$. By Proposition 3.2, $x^* \in \mathcal{F}(\tau)$. Moreover, there exists $\bar{x} \in X(\tau)$ such that $f(x^*) - f(\bar{x}) > 0$. On the other hand, since $\bar{x} \in \mathcal{F}(\tau) \subset \mathcal{F}_{\gamma_N}(\tau)$, $f(x^N) - f(\bar{x}) \leq 0$. This contradicts an earlier inequality when N is sufficiently large.

Part (ii). The conclusion essentially follows from Part (i) and [22, Theorem 1]. Here we include a brief proof for completeness. Let x_γ and x_0 be the optimal solutions of the problems (3.2) and (3.1) respectively. By part (ii) of the proposition 3.2 , there exists $\bar{x}_\gamma \in \mathcal{F}(\tau)$ such that $\|x_\gamma - \bar{x}_\gamma\| \leq C\Delta_\gamma$. Moreover

$$v(\tau) \leq f(\bar{x}_\gamma) \leq f(x_\gamma) + |f(\bar{x}_\gamma) - f(x_\gamma)| \leq v_\gamma(\tau) + LC\Delta_\gamma.$$

Under the symmetry of the Hausdorff distance between $\mathcal{F}_\gamma(\tau)$ and $\mathcal{F}(\tau)$, we can show $v_\gamma(\tau) \leq v(\tau) + LC\Delta_\gamma$. The conclusion follows. ∎

In the classical stability analysis of nonlinear programming, it is often assumed some kind of growth conditions for the objective function in order to derive the stability of the optimal solution, see for instance Klatte [22, Theorem 2]. In this context, the growth condition would be

$$f(x') \geq \min_{x \in \mathcal{F}(\tau)} f(x) + \alpha d(x', X(\tau)).$$

Since problems (3.1) and (3.2) have identical objective functions while the feasible set of the former is contained in that of the latter, the growth condition would force the set of optimal solutions to the problem (3.2) to stay on the set of optimal solutions to the problem (3.1). To see this more clearly, let $\bar{x} \in X_\gamma(\tau)$ be an optimal solution to the problem (3.2). Since $X(\tau) \subset \mathcal{F}_\gamma(\tau)$, the growth condition means

$$0 \geq v_\gamma(\tau) - v(\tau) = f(\bar{x}) - v(\tau) \geq \alpha d(\bar{x}, X(\tau)),$$

which yields $\bar{x} \in X(\tau)$ and hence $X_\gamma(\tau) \equiv X(\tau)$.

3.2. Sample average approximation. In some circumstances, it might be numerically too expensive to calculate the expected values in the entropic function. A well-known approximation method to deal with the mathematical expectation in stochastic programming is the sample average approximation (SAA) which is also known under various name such as Monte Carlo method, sample path optimization method, see [35] for a comprehensive review. The basic idea of SAA can be described as follows. Suppose that we have an independent and identically distributed (iid) sample $t^1, \cdots, t^N$ of random vector t. This may be obtained through random sampling over set T or we have a way to obtain samples of t (e.g. empirical data) in the case when t is a random parameter in the original problem. With the samples, we can construct the sample average approximation:

$$\text{(SAA-EA-RMPSICC)} \qquad \begin{array}{ll} \min\limits_{x \in X_+} & f(x) \\ \text{s.t.} & \left.\begin{array}{l} e_\gamma^{j,N}(x) \leq 0 \\ \bar{e}_\gamma^{j,N}(x) \leq \tau \end{array}\right\} \text{for } j \in \{1, \cdots, n\}, \end{array} \tag{3.4}$$

where

$$e_\gamma^{j,N}(x) := \frac{1}{\gamma} \ln \left(\frac{1}{N} \sum_{i=1}^N e^{-\gamma g_j(x, t^i)} \right), \quad \bar{e}_\gamma^{j,N}(x) := \frac{1}{\gamma} \ln \left(\frac{1}{N} \sum_{i=1}^N e^{\gamma x_j \cdot g_j(x, t^i)} \right).$$

In what follows, we investigate the convergence of the optimal value and the optimal solutions obtained from solving SAA-EA-RMPSICC (3.4) as the sample size increases. To this end, we consider the following general stochastic inequality constrained minimization problem

$$\begin{array}{ll} \min\limits_{x \in D} & \mathbb{E}[\psi_0(x, \xi)] \\ \text{s.t.} & \mathbb{E}[\psi_j(x, \xi)] \leq 0, \quad \text{for } j = 1, \cdots, m, \end{array} \tag{3.5}$$

where $\psi_j(x, \xi) : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$, $j = 0, \cdots, m$, are continuous functions. Let $\xi^1, \cdots, \xi^N$ be independent random vectors following a distribution identical to that of ξ and

$$\psi_j^N(x) := \frac{1}{N} \sum_{i=1}^N \psi(x, \xi^i), \quad \text{for } j = 0, \cdots, m.$$

By replacing $\mathbb{E}[\psi_j(x,\xi)]$ with $\psi_j^N(x)$, we can construct sample average approximation of the problem (3.5) as follows:

$$\begin{array}{lll} \text{(SAA)} & \min\limits_{x\in D} & \psi_0^N(x) \\ & \text{s.t.} & \psi_j^N(x) \le 0, \quad \text{for } j=1,\cdots,m. \end{array} \tag{3.6}$$

Let $\mathcal{F}^N$, S^N and v^N denote the set of feasible solutions, the set of optimal solutions, and the optimal value of problem (3.6) respectively, let $\mathcal{F}$, S^* and v^* denote the set of feasible solutions, the set of optimal solutions, and the optimal value of problem (3.5). Let $\mathcal{F}^s$ denote the set of strictly feasible solutions of problem (3.5), that is,

$$\mathcal{F}^s := \{x \in D : \mathbb{E}[\psi_j(x,\xi)] < 0, \text{ for } j=1,\cdots,m\}. \tag{3.7}$$

Note that $\mathcal{F}^s$ should be distinguished from the interior of the feasible set $\mathcal{F}$ because what we need here is the strict inequality in the constraints.

The following lemma summarizes the convergence of problem (3.6) to problem (3.5) in terms of the optimal value and the optimal solutions as the sample size N increases.

LEMMA 3.1. **(Convergence of SAA (3.6))** *Assume: (a) D is a compact set; (b) $\mathrm{cl}\mathcal{F}^s \cap S^* \neq \emptyset$; (c) $\psi_j(x,\xi)$, $j=0,\cdots,m$, is integrably bounded. Then,*

*(i) with probability one (*w.p.1*)*

$$\overline{\lim_{N\to\infty}} S^N \subseteq S^* \quad \text{and} \quad \lim_{N\to\infty} v^N = v^*.$$

(ii) Assume, in addition, that (d) the objective $\mathbb{E}[\psi_0(\cdot,\xi)]$ satisfies some growth condition, that is, there exists $\epsilon_0 > 0$ such that

$$R(\epsilon) := \inf_{\substack{x\in\mathcal{F} \\ d(x,S^*)\ge\epsilon}} \mathbb{E}[\psi_0(\cdot,\xi)] - v^* > 0, \quad \forall \epsilon \in (0,\epsilon_0];$$

(e) the constraints $\mathbb{E}[\psi_j(\cdot,\xi)]$ $j=1,\cdots,m$ satisfy some growth condition, that is

$$\hat{R}(\epsilon) := \inf_{\substack{x\in X,\ j\in\{1,\cdots,m\} \\ d(x,\mathcal{F})\ge\epsilon}} \mathbb{E}[\psi_j(\cdot,\xi)] > 0, \quad \forall \epsilon \in (0,\epsilon_0];$$

(f) for every $x \in D$, the moment function $M_x^j(s) := \mathbb{E}\left[e^{s(\psi_j(x,\xi))}\right]$, $j=0,\cdots,m$, is finite valued for all s in a neighborhood of zero; (g) there exist a measurable function $L(\xi)$ and a constant ν such that

$$|\psi_j(x',\xi) - \psi_j(x,\xi)| \le L(\xi)\|x'-x\|^\nu, \quad j=0,\cdots,m,$$

for all $\xi \in \Xi$ and all $x', x \in D$, and the moment function $M_L(s)$ of $L(\xi)$ is finite for s in a neighborhood of zero. Then for any $\epsilon \in (0,\epsilon_0]$ and sequence $\{x^N\}$ with $x^N \in S^N$, there exist positive constants $C(\epsilon)$ and $\alpha(\epsilon)$ (independent of N) such that

$$\mathrm{Prob}\left\{d(x^N,S^*) \ge \epsilon\right\} \le C(\epsilon)e^{-\alpha(\epsilon)N} \tag{3.8}$$

for N sufficiently large.

Before presenting a proof, it might be helpful to make some comments on the conditions of the lemma. Condition (b) requires that there exists an optimal solution which is an accumulation point of a sequence of strict feasibility points. Condition (d) is weaker than the second order growth condition which is widely adopted in the literature of stability analysis for nonlinear programming. Condition (e) is a kind of growth condition for the constraint functions which is needed for characterizing the error bound of the feasible solutions in terms of the constraint functions. The condition may be dropped when the constraint functions are convex, see [2, Theorem 5] for the case where the constraint function is CVaR of a convex random function. Conditions (f) and (g) are the moment conditions which require the tail of the random variables die at exponential rate and they are satisfied when ξ has bounded support set. These conditions are standard for deriving exponential rate of convergence, see for example [41].

Proof of Lemma 3.1. Part (i). Let

$$\Psi^N(x) := \begin{pmatrix} \psi_1^N(x) \\ \vdots \\ \psi_m^N(x) \end{pmatrix} \qquad \text{and} \qquad \Psi(x) := \begin{pmatrix} \mathbb{E}[\psi_1(x,\xi)] \\ \vdots \\ \mathbb{E}[\psi_m(x,\xi)] \end{pmatrix}.$$

The feasible sets $\mathcal{F}^N$ and $\mathcal{F}$ can be represented respectively as the sets of solutions to the following generalized equations

$$0 \in \Psi^N(x) + \mathbb{R}^m_+ \qquad \text{and} \qquad 0 \in \Psi(x) + \mathbb{R}^m_+$$

restricted to the set D. Under condition (c), it follows by [35, Chapter 6, Proposition 7] that $\psi_j^N(\cdot)$ converges to $\mathbb{E}[\psi_j(\cdot,\xi)]$ uniformly over D w.p.1, which means that $\Psi^N(x)$ converges to $\Psi(x)$ uniformly over D. By [43, Lemma 4.2 (i)],

$$\lim_{N\to\infty} \mathbb{D}(\mathcal{F}^N, \mathcal{F}) = 0, \qquad \text{w.p.1.} \tag{3.9}$$

Let $\{x^N\}$ be a sequence of the optimal solutions to problem (3.6). Since the sequence is contained in the compact set D, by taking a subsequence if necessary we may assume for the simplicity of notation that $x^N \to x^*$. By (3.9) and the closedness of $\mathcal{F}$, $x^* \in \mathcal{F}$. In what follows, we show that $x^* \in S^*$ w.p.1. Observe first that the uniform convergence of $\psi_0^N(\cdot)$ ensures

$$\lim_{N\to\infty} v^N = \lim_{N\to\infty} \psi_0^N(x^N) = \mathbb{E}[\psi_0(x^*,\xi)] \geq v^*.$$

Under condition (b), there exists a $y^* \in S^*$ such that $y^* \in \mathrm{cl}\mathcal{F}^s$. By the continuity of $\mathbb{E}[\psi_0(\cdot,\xi)]$, for any small positive number ϵ, there exists $y^\epsilon \in \mathcal{F}^s$ such that $\mathbb{E}[\psi_0(y^\epsilon,\xi)] - v^* \leq \epsilon$. Since $y^\epsilon \in \mathcal{F}^s$, and $\Psi^N(x)$ converges to $\Psi(x)$ uniformly over D, it is easy to show that there exists $y^N \in \mathcal{F}^N$ such that $||y^N - y^\epsilon|| \to 0$, w.p.1. Therefore

$$v^* \geq \mathbb{E}[\psi_0(y^\epsilon,\xi)] - \epsilon = \lim_{N\to\infty} \psi_0^N(y^N) - \epsilon \geq \lim_{N\to\infty} \psi_0^N(x^N) - \epsilon = \mathbb{E}[\psi_0(x^*,\xi)] - \epsilon$$

w.p.1. Since ϵ is chosen arbitrarily, we arrive at $v^* \geq \mathbb{E}[\psi_0(x^*,\xi)]$ which means that $x^* \in S^*$.

Part (ii). To make it easier to follow, we divide the proof of this part into 5 steps.

Step 1. Let σ be a small positive number. Under condition (e), it follows by [41, Theorem 5.1] that there exist positive constants $C(\sigma)$, $\alpha(\sigma)$ and $N(\sigma)$ such that

$$\mathrm{Prob}\left\{\sup_{x\in D} \left|\psi_j^N(x) - \mathbb{E}[\psi_j(x,\xi)]\right| \geq \sigma\right\} \leq C(\sigma)e^{-\alpha(\sigma)N}, \quad \text{for } j \in \{0,\cdots,m\}, \tag{3.10}$$

when $N \geq N(\sigma)$.

Step 2. Let $\epsilon \leq \epsilon_0$ be a positive number and $y^\epsilon \in \mathcal{F}^s$ such that $\mathbb{E}[\psi_0(y^\epsilon)] - v^* \leq R(\epsilon)/4$, where ϵ_0 is given in condition (d). The existence of y^ϵ is ensured by condition (b). We estimate $\mathrm{Prob}\left\{y^\epsilon \notin \mathcal{F}^N\right\}$.

$$\begin{aligned} \mathrm{Prob}\left\{y^\epsilon \notin \mathcal{F}^N\right\} &= \mathrm{Prob}\left\{\max_{j\in\{1,\cdots,m\}} \psi_j^N(y^\epsilon) > 0\right\} \\ &= \mathrm{Prob}\left\{\max_{j\in\{1,\cdots,m\}} \psi_j^N(y^\epsilon) - \max_{j\in\{1,\cdots,m\}} \mathbb{E}[\psi_j(y^\epsilon,\xi)] > - \max_{j\in\{1,\cdots,m\}} \mathbb{E}[\psi_j(y^\epsilon,\xi)]\right\}. \end{aligned}$$

Since $y^\epsilon \in \mathcal{F}^s$, there exists a positive number λ such that

$$- \max_{j\in\{1,\cdots,m\}} \mathbb{E}[\psi_j(y^\epsilon,\xi)] \geq \lambda.$$

Under condition (f), there exist positive constants $C(\lambda)$ and $\alpha(\lambda)$ (independent of N) such that

$$\mathrm{Prob}\left\{y^\epsilon \notin \mathcal{F}^N\right\} \leq C(\lambda)e^{-\alpha(\lambda)N} \tag{3.11}$$

for N sufficiently large.

Step 3. When $y^\epsilon \in \mathcal{F}^N$, $\psi_0^N(x^N) \le \psi_0^N(y^\epsilon)$, which implies

$$\psi_0^N(x^N) - v^* \le \psi_0^N(x^N) - \mathbb{E}[\psi_0(y^\epsilon,\xi)] + R(\epsilon)/4 \le \psi_0^N(y^\epsilon) - \mathbb{E}[\psi_0(y^\epsilon,\xi)] + R(\epsilon)/4.$$

Therefore

$$\begin{aligned}
&\text{Prob}\left\{\mathbb{E}[\psi_0(x^N,\xi)] - v^* \ge R(\epsilon) \text{ and } y^\epsilon \in \mathcal{F}^N\right\}\\
&\le \text{Prob}\left\{\mathbb{E}[\psi_0(x^N,\xi)] - \psi_0^N(x^N) \ge R(\epsilon)/2\right\} + \text{Prob}\left\{\psi_0^N(x^N) - v^* \ge R(\epsilon)/2\right\}\\
&\le \text{Prob}\left\{\mathbb{E}[\psi_0(x^N,\xi)] - \psi_0^N(x^N) \ge R(\epsilon)/2\right\} + \text{Prob}\left\{\psi_0^N(y^\epsilon) - \mathbb{E}[\psi_0(y^\epsilon,\xi)] \ge R(\epsilon)/4\right\}.
\end{aligned}$$

The uniform exponential convergence in (3.10) ensures that $\mathbb{E}[\psi_0(x^N,\xi)] - \psi_0^N(x^N)$ and $\psi_0^N(y^\epsilon) - \mathbb{E}[\psi_0(y^\epsilon,\xi)]$ converge to zero at exponential rate as $N \to \infty$. Then

$$\text{Prob}\left\{\mathbb{E}[\psi_0(x^N,\xi)] - v^* \ge R(\epsilon) \text{ and } y^\epsilon \in \mathcal{F}^N\right\} \le 2C(R(\epsilon)/4)e^{-\alpha(R(\epsilon)/4)N}.$$

Step 4. Let $\hat\epsilon \le \epsilon_0$ be a positive number. By the growth condition (e)

$$\begin{aligned}
\text{Prob}(d(x^N,\mathcal{F}) > \hat\epsilon) &\le \text{Prob}\left(\min_{j\in\{1,\cdots,m\}} \mathbb{E}[\psi_j(x^N,\xi)] \ge \hat R(\hat\epsilon)\right)\\
&\le \text{Prob}\left(\min_{j\in\{1,\cdots,m\}} |\mathbb{E}[\psi_j(x^N,\xi)] - \psi_j^N(x^N)| \ge \hat R(\hat\epsilon)\right)\\
&\le \text{Prob}\left(\sup_{x\in D}\min_{j\in\{1,\cdots,m\}} |\mathbb{E}[\psi_j(x,\xi)] - \psi_j^N(x)| \ge \hat R(\hat\epsilon)\right)\\
&\le C(\hat R(\hat\epsilon))e^{-\alpha(\hat R(\hat\epsilon))N}.
\end{aligned}$$

Step 5. We are now ready to estimate $\text{Prob}\left\{d(x^N,S^*) \ge \epsilon\right\}$. Observe that

$$\text{Prob}\left\{d(x^N,S^*) \ge \epsilon\right\} \le \text{Prob}\left\{d(x^N,S^*) \ge \epsilon \text{ and } y^\epsilon \in \mathcal{F}^N\right\} + \text{Prob}\left\{y^\epsilon \notin \mathcal{F}^N\right\}. \tag{3.12}$$

By (3.11), $\text{Prob}\left\{y^\epsilon \notin \mathcal{F}^N\right\}$ goes to zero at an exponential rate. Thus, it is sufficient to estimate the first term at the right hand side of (3.12). Let z^N be a project of x^N on $\mathcal{F}$. Then $\|x^N - z^N\| = d(x^N,\mathcal{F})$. Taking advantage of the step 4, it is sufficient to consider the case that $\|x^N - z^N\| \le \hat\epsilon$ where $0 < \hat\epsilon < \epsilon \le \epsilon_0$ and such that $R(\epsilon - \hat\epsilon) - \mathbb{E}[L(\xi)]\hat\epsilon > 0$.

Observe first that $d(x^N,S^*) \le \|x^N - z^N\| + d(z^N,S^*)$. Subsequently $d(x^N,S^*) \ge \epsilon$ implies $d(z^N,S^*) \ge \epsilon - \hat\epsilon$. Then under the growth condition (d), we have for sufficiently large N,

$$\begin{aligned}
\text{Prob}\left\{d(x^N,S^*) \ge \epsilon \text{ and } y^\epsilon \in \mathcal{F}^N\right\} &\le \text{Prob}\left\{d(z^N,S^*) \ge \epsilon - \hat\epsilon \text{ and } y^\epsilon \in \mathcal{F}^N\right\}\\
&\le \text{Prob}\left\{\mathbb{E}[\psi_0(z^N,\xi)] - v^* \ge R(\epsilon - \hat\epsilon) \text{ and } y^\epsilon \in \mathcal{F}^N\right\}\\
&\le \text{Prob}\left\{\mathbb{E}[\psi_0(x^N,\xi)] - v^* \ge R(\epsilon - \hat\epsilon) - \mathbb{E}[L(\xi)]\hat\epsilon \text{ and } y^\epsilon \in \mathcal{F}^N\right\}.
\end{aligned}$$

Through Step 3, we have

$$\begin{aligned}
&\text{Prob}\left\{\mathbb{E}[\psi_0(x^N,\xi)] - v^* \ge R(\epsilon - \hat\epsilon) - \mathbb{E}[L(\xi)]\hat\epsilon \text{ and } y^\epsilon \in \mathcal{F}^N\right\}\\
&\le 2C([R(\epsilon - \hat\epsilon) - \mathbb{E}[L(\xi)]\hat\epsilon]/4)e^{-\alpha([R(\epsilon-\hat\epsilon) - \mathbb{E}[L(\xi)]\hat\epsilon]/4)N}.
\end{aligned}$$

Summarizing the discussions above, we can find some positive constants $C(\epsilon)$, $\alpha(\epsilon)$ and $N(\epsilon)$ such that (3.8) holds for $N \ge N(\epsilon)$. ■

Note also that Shapiro [37] investigated the approximation of problem (3.6) to (3.5). He derived a δ-theorem which describes the asymptotic behavior of $\sqrt{N}(v^N - v^*)$ under the condition that the underlying functions are convex. It is unclear whether similar results can be established without convexity. Lemma 3.1 implies the convergence on probability but does not specifically describe the behavior of $\sqrt{N}(v^N - v^*)$ as N increases.

With Lemma 3.1, we are ready to state the convergence of the optimal solutions of problem (3.4).

THEOREM 3.2. **(Convergence of SAA-EA-MPSICC (3.4))** *Let $X_\gamma^N(\tau)$, $X_\gamma(\tau)$ and $X(\tau)$ denote the set of optimal solutions of problems (3.4), (3.2) and (3.1) respectively, $v_\gamma^N(\tau)$, $v_\gamma(\tau)$, and $v(\tau)$ the corresponding optimal values. Assume: (a) the conditions of Proposition 3.2 hold; (b) $X_\gamma(\tau) \cap \mathcal{F}_\gamma^s(\tau) \ne \emptyset$, where the superscript s in $\mathcal{F}_\gamma^s(\tau)$ indicates the set of strictly feasible solutions as in (3.7). Then the following assertions hold.*

(i) For fixed γ, w.p.1

$$\overline{\lim_{N\to\infty}} X_\gamma^N(\tau) \subseteq X_\gamma(\tau), \quad \lim_{N\to\infty} v_\gamma^N(\tau) = v_\gamma(\tau);$$

(ii) w.p.1

$$\overline{\lim_{\substack{N\to\infty\\ \gamma\to\infty}}} X_\gamma^N(\tau) \subseteq X(\tau), \quad \lim_{\substack{N\to\infty\\ \gamma\to\infty}} v_\gamma^N(\tau) = v(\tau).$$

Assume, in addition, (c) the objective function and the constraint functions of problem (3.2) satisfy the growth conditions similar to those of (d)-(e) in Lemma 3.1, (d) $g(x,t)$ is Lipschitz continuous on X uniformly with respect to t, that is, there exists a constant L such that

$$\|g(x',t) - g(x'',t)\| \le L\|x' - x''\|, \quad \forall x', x'' \in X,\ \forall t \in T.$$

Then $\mathbb{D}(X_\gamma^N(\tau), X(\gamma))$ converges to zero at exponential rate with increase of the sample size.

Theorem 3.2 follows from Lemma 3.1. Indeed, conditions (a)-(c) of the theorem are sufficient for the conditions (a)-(e) in Lemma 3.1. The compactness of T and condition (d) imply the conditions (f)-(g) in Lemma 3.1.

Before concluding this section, we clarify the difference between the SAA-EA-MPSICC scheme and the well known randomization scheme for mathematical programs with robust convex constraints proposed by Calafiore and Campi [12] and the CVaR approximation scheme recently considered by Anderson et al [2]. To simplify the discussion, let us consider the semi-infinite system

$$h(x,\xi) \le 0, \quad \forall \xi \in \Xi,$$

where $h : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$ is a continuous function. Let $\xi^1, \cdots, \xi^N$ be an iid sample of ξ. The SAA-EA-MPSICC scheme approximates the system with

$$\frac{1}{\gamma}\left[\ln\left(e^{\gamma h(x,\xi^1)} + \cdots + e^{\gamma h(x,\xi^N)}\right) - \ln N\right] \le 0 \tag{3.13}$$

whereas Calafiore and Campi's scheme approximates it by

$$\max_{i\in\{1,\cdots,N\}} h(x,\xi^i) \le 0. \tag{3.14}$$

On the other hand, the CVaR approximation is defined as

$$\min_{\eta\in\mathbb{R}} \eta + \frac{1}{1-\beta}\sum_{i=1}^N (h(x,\xi^i) - \eta)_+ \le 0, \tag{3.15}$$

where $\beta \in (0,1)$ is a parameter and $(a)_+ = \max(0,a)$ for a scalar a. Obviously the three approximation schemes are different in that (3.13) uses all samples while (3.14) only captures the extreme ones and the CVaR approximation utilizes the samples at the tail of the distribution of $h(x,\xi)$. When $\gamma \to \infty$ and $\beta \to 1$, the three schemes coincide. Note that Calafiore and Campi [12] presented a sample based probabilistic statement for the optimal value and feasibility of the optimal solution obtained from solving their randomization scheme. It will be interesting to investigate whether similar claims can be made without convexity (which is typical in MPECs). We leave this for future research.

4. Two stage SMPECs. In this section, we apply the approximation schemes proposed in the preceding section to the following two stage SMPECs:

$$\text{(SMPEC)} \qquad \begin{array}{ll} \min\limits_{x\in X,\, y(\cdot)\in\mathscr{Y}} & \mathbb{E}[f(x,y(\xi),\xi)] \\ \text{s.t.} & 0 \le G(x,y(\xi),\xi) \ \perp H(x,y(\xi),\xi) \ge 0, \quad \forall \xi \in \Xi, \end{array} \tag{4.1}$$

where X is a nonempty closed subset of $\mathbb{R}^n$, f, G, H are continuously differentiable functions from $\mathbb{R}^n \times \mathbb{R} \times \mathbb{R}$ to $\mathbb{R}$, $\xi : \Omega \to \Xi$ is a random variable defined on the probability space $(\Omega, \mathcal{F}, P)$ with compact support set $\Xi \subset \mathbb{R}$, and $\mathbb{E}[\cdot]$ denotes the expected value with respect to probability measure P. As before '$\perp$' denotes the

orthogonality of two vectors, $\mathscr{Y}$ denotes a space of functions $y(\cdot) : \Xi \to \mathbb{R}$ such that $\mathbb{E}[f(x, y(\xi), \xi)]$ is well defined. Note that our restriction of G and H to a scalar function is purely for the simplicity of exposition, the numerical scheme and technical results in this section can be easily applied to the case when G and H are vector-valued.

Over the past few years since the pioneering work by Patriksson and Wynter [32] on SMPEC, there have been increasing discussions on the two stage SMPECs, which cover a wide range of topics from optimality theory [42, 44] to numerical methods such as sample average approximation methods [38, 41], implicit smoothing method [24] and regularized method [26, 29]. Here we take a completely different numerical strategy: we apply the well known polynomial decision rule in robust optimization to approximate the second stage equilibrium constraint so that the two stage SMPEC effectively reduces to a one stage stochastic mathematical program with robust semi-infinite complementarity constraints, we then tackle the latter with the numerical scheme discussed in the preceding section.

The basic idea of the polynomial decision rule is to replace the second stage equilibrium $y(\xi)$ with a polynomial function of ξ. This replacement effectively restricts the functional form of the decision variable at the second stage to a class of polynomial functions rather than measurable functions as it's supposed to be, and hence reduces the space of feasible equilibrium solutions to those which are representable by polynomials. The radical approach was proposed by Ben Tal et al [9] with a linear decision rule and was later extended by Shapiro and Nemirovski [40], Chen et al [15]. More recently, the approach has been further developed to a class of two stage stochastic programs by Kuhn et al [23] and extended to the polynomial decision rule by Bampou and Kuhn [4].

One of the main technical issues in the polynomial decision rule is the feasibility of the constraints when the second stage decision variables are restricted to a polynomial function of ξ. The equilibrium constraints at the second stage essentially involve equality constraints and the approximated problem may not have a feasible solution if we apply the polynomial decision rule to the constraints directly. Indeed, it is easy to observe from the classical implicit function point of view that the implicit function may not exist if it is restricted to polynomial class. To get around this technical difficulty, we apply the NLP relaxation as we discussed in the preceding sections to SMPEC (4.1) so that the complementarity constraints at the second stage are represented by a system of inequalities parameterized by a controllable parameter and then apply the polynomial decision rule to the latter. Specifically, we consider

$$\begin{array}{ll} \min\limits_{x\in X,\ y(\cdot)\in\mathscr{Y}} & \mathbb{E}[f(x, y(\xi), \xi)] \\ \text{s.t.} & \left.\begin{array}{l} G(x, y(\xi), \xi) \geq 0 \\ H(x, y(\xi), \xi) \geq 0 \\ G(x, y(\xi), \xi) \cdot H(x, y(\xi), \xi) \leq \tau \end{array}\right\} \forall \xi \in \Xi, \end{array} \tag{4.2}$$

where τ is a small positive parameter. Clearly the optimal value of problem (4.2) provides a lower bound for its true counterpart. Under some moderate conditions, it can be shown that problem (4.2) approximates problem (4.1) in terms of the optimal value, the optimal solutions as well as the stationary points as $\tau \to 0$, see [26, 41].

Let us now fix the relaxed parameter τ and apply the polynomial decision rule to problem (4.2). We restrict the second stage decision variable $y(\xi)$ to the space of polynomials with degree k, that is

$$y(\xi) := y_0 + y_1\xi + y_2\xi^2 + \cdots + y_k\xi^k,$$

where $y_0, \cdots, y_k$ are real numbers. For brevity, we write $\mathbf{y} = (y_0, \cdots, y_k)$. Substituting $y(\xi)$ into (4.2), we obtain a stochastic mathematical program with robust semi-infinite NLP constraints:

$$\text{(PDL-RSMPEC)} \qquad \begin{array}{ll} \min\limits_{x\in X\,\mathbf{y}\in\mathbb{R}^{k+1}} & \mathbb{E}_P[f(x, \mathbf{y}, \xi)] \\ \text{s.t.} & \left.\begin{array}{l} G(x, \mathbf{y}, \xi) \geq 0 \\ H(x, \mathbf{y}, \xi) \geq 0 \\ G(x, \mathbf{y}, \xi) \cdot H(x, \mathbf{y}, \xi) \leq \tau \end{array}\right\} \forall \xi \in \Xi. \end{array} \tag{4.3}$$

When problem (4.2) constitutes a strictly feasible solution, it is easy to show the feasibility of problem (4.3). Strict feasibility of the regularized problem (4.2) is related to the constraints of the original problem (4.1). For example, if there exist a point $\hat{x} \in X$, a feasible solution at the second stage $y(\hat{x}, \cdot)$ to problem (4.1) and positive constants C_1 and C_2 such that

$$H'_y(\hat{x}, y(\hat{x}, \xi), \xi), G'_y(\hat{x}, y(\hat{x}, \xi), \xi) \in [C_1, C_2]$$

for all $\xi \in \Xi$, then there exists a positive number δ_0 such that

$$\begin{aligned} H(\hat{x}, y(\hat{x},\xi)+\delta,\xi) &> 0,\\ G(\hat{x}, y(\hat{x},\xi)+\delta,\xi) &> 0,\\ G(\hat{x}, y(\hat{x},\xi)+\delta,\xi)\cdot H(\hat{x}, y(\hat{x},\xi)+\delta,\xi) &< \tau, \end{aligned}$$

for all $\xi \in \Xi$ and $\delta \in [0,\delta_0)$, which means $(\hat{x}, y(\hat{x},\xi))$ is a strictly feasible solution to problem (4.2). In such a case, the approximated problem (4.3) has a feasible solution for some appropriate degree of the polynomial. See [4] a similar discussion on continuous linear programming problems with inequality constraints. In what follows, we make a blanket assumption that problem (4.3) is feasible, that is, there exists at least one feasible solution.

The rest of discussions are similar to the preceding section: we write the problem above as

$$\begin{array}{ll} \min\limits_{x\in X,\, \mathbf{y}\in\mathbb{R}^{k+1}} & \mathbb{E}_P[f(x,\mathbf{y},\xi)] \\ \text{s.t.} & \sup\limits_{\xi\in\Xi} -G(x,\mathbf{y},\xi) \le 0, \\ & \sup\limits_{\xi\in\Xi} -H(x,\mathbf{y},\xi) \le 0, \\ & \sup\limits_{\xi\in\Xi} G(x,\mathbf{y},\xi)\cdot H(x,\mathbf{y},\xi) \le \tau, \end{array} \tag{4.4}$$

and then construct an entropic approximation to the constraints.

Note that problem (4.4) provides an upper bound for the optimal value of problem (4.3). Moreover, the gap between the optimal values of problems (4.4) and (4.3) decreases as the degree of the polynomial of $y(\xi)$ increases. In what follows, we discuss the entropic approximation of problem (4.4) for fixed a degree k, namely, we consider

$$\begin{array}{ll} \min\limits_{x\in X,\, \mathbf{y}\in\mathbb{R}^{k+1}} & \mathbb{E}_P[f(x,\mathbf{y},\xi)] \\ \text{s.t.} & e^1_\gamma(x,\mathbf{y}) \le 0, \\ & e^2_\gamma(x,\mathbf{y}) \le 0, \\ & e^3_\gamma(x,\mathbf{y}) \le \tau, \end{array} \tag{4.5}$$

where

$$\begin{aligned} e^1_\gamma(x,\mathbf{y}) &:= e_\gamma(G(x,\mathbf{y},\xi)),\\ e^2_\gamma(x,\mathbf{y}) &:= e_\gamma(H(x,\mathbf{y},\xi)),\\ e^3_\gamma(x,\mathbf{y}) &:= e_\gamma(-G(x,\mathbf{y},\xi)\cdot H(x,\mathbf{y},\xi)), \end{aligned}$$

and $e_\gamma(\cdot)$ is defined in Section 2. Clearly problem (4.5) is a one stage stochastic minimization problem. Similar to the discussions in Section 3, we can easily establish the convergence of the optimal value of problem (4.5) as γ increases.

Let us now turn to discuss computation of $e^j_\gamma(x,\mathbf{y})$. In practice, the distribution of ξ is often unknown or it is numerically too expensive to calculate the expected values. Instead it might be possible to obtain a sample of the random vector ξ from historical data. This motivates us to find an approximate optimal solution to problem (4.5) on the basis of the empirical data. Suppose that we have an iid sample $\xi^1,\cdots,\xi^N$ of random vector ξ. Then $\mathbb{E}_P[f(x,\mathbf{y},\xi)]$ and $e^j_\gamma(x,\mathbf{y})$ can be approximated by the sample average as follows:

$$\begin{aligned} f^N(x,\mathbf{y}) &:= \frac{1}{N}\sum_{i=1}^N f(x,\mathbf{y},\xi^i),\\ e^{1,N}_\gamma(x,\mathbf{y}) &:= \frac{1}{\gamma}\ln\left(\frac{1}{N}\sum_{i=1}^N e^{-\gamma G(x,\mathbf{y},\xi^i)}\right),\\ e^{2,N}_\gamma(x,\mathbf{y}) &:= \frac{1}{\gamma}\ln\left(\frac{1}{N}\sum_{i=1}^N e^{-\gamma H(x,\mathbf{y},\xi^i)}\right),\\ e^{3,N}_\gamma(x,\mathbf{y}) &:= \frac{1}{\gamma}\ln\left(\frac{1}{N}\sum_{i=1}^N e^{\gamma G(x,\mathbf{y},\xi^i)\cdot H(x,\mathbf{y},\xi^i)}\right). \end{aligned}$$

Consequently we may consider the following sample average and entropic approximated PDL-RSMPEC (4.3)

$$\text{(SAA-EA-PDL-RSMPEC)} \qquad \begin{array}{ll} \min\limits_{x\in X,\, \mathbf{y}\in\mathbb{R}^{k+1}} & f^N(x,\mathbf{y}) \\ \text{s.t.} & e_\gamma^{1,N}(x,\mathbf{y}) \le 0, \\ & e_\gamma^{2,N}(x,\mathbf{y}) \le 0, \\ & e_\gamma^{3,N}(x,\mathbf{y}) \le \tau. \end{array} \tag{4.6}$$

For the simplicity of notation, let $w := (x, \mathbf{y})$. The following theorem states the convergence of the optimal value and the optimal solutions of problem (4.6) as the sample size increases. We omit the proof as it follows directly from Lemma 3.1.

THEOREM 4.1. **(Convergence of SAA-EA-PDL-RSMPEC (4.6))** *Let W_γ^N and W_γ denote the sets of optimal solutions of problems (4.6) and (4.5) respectively, and v_γ^N and v_γ the corresponding optimal values. Assume that (a) W_γ^N and W_γ are nonempty and contained in a compact set W; (b) there exists $w \in W_\gamma$ which is strictly feasible, (c) the functions $f(w,\xi)$, $G(w,\xi)$, $H(w,\xi)$ and $G(w,\xi)^T H(w,\xi)$ are measurable and integrably bounded. Then the following assertions hold.*

(i) w.p.1 $\lim\limits_{N\to\infty} v_\gamma^N \to v_\gamma$;
(ii) w.p.1 $\overline{\lim\limits_{N\to\infty}} W_\gamma^N \subseteq W_\gamma$.

5. Mathematical programs with distributionally robust constraints. In this section, we consider the mathematical programs with distributionally robust complementarity constraints

$$\text{(MPDRE)} \qquad \begin{array}{ll} \min\limits_{x\in X} & f(x) \\ \text{s.t.} & 0 \le x \perp \mathbb{E}_P[F(x,\xi)] \ge 0, \quad \forall P \in \mathcal{P}, \end{array} \tag{5.1}$$

where X is a nonempty closed set of $\mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$ and $F : \mathbb{R}^n \times \Xi \to \mathbb{R}^n$ are continuous functions, ξ is a random vector with support set $\Xi \subset \mathbb{R}^d$ and the mathematical expectation $\mathbb{E}_P[\cdot]$ is taken w.r.t. the distribution of ξ, $\mathcal{P}$ is a set of probability measures.

A unique feature of this model is the distributionally robust constraints. Differing from MPSICC (1.1), here the equilibrium/complementarity constraints must hold for a set of probability distributions rather than for every realization of the random variable. From a practical perspective, it may be interpreted as an equilibrium to be held for any distribution that the underlying random variable may possibly follow, and we are looking at such an equilibrium which minimizes the disutility $f(x)$. Equilibrium is also known as distributional ex post equilibrium [33]. The model may be used as an approach for new market design where the regulator sets out optimal parameters (representing regulative policies) which maximize its utility whereas market players are expected to reach an equilibrium (on the basis of expected profit maximization) under any foreseeable distribution of underlying uncertainty.

From computational point of view, MPDRE (5.1) might provide an approximation to MPSICC (1.1) if we restrict $\mathcal{P}$ to a subset of distributions that it could possibly take because the optimal value of the former would give rise to a lower bound for the optimal value of the latter. If we do so to the Lagrange dual of MPDRE (5.1), then we may obtain an upper bound for the optimal value of MPSICC (1.1). Examples for the selection of $\mathcal{P}$ include empirical probability measures and some specific distributions such as uniform distribution, normal distribution etc depending on the requirement on the accuracy of the approximation and consequently MPDRE (5.1) reduces to a deterministic MPCC or a one stage stochastic MPCC.

Two specific cases to note: when $\mathcal{P}$ consists of the Dirac probability measures, MPDRE (5.1) coincides with MPSICC (1.1), and when $\mathcal{P}$ contains a finite number of distributions, it reduces to one stage SMPEC. The latter has been well studied over the past decade, see for instance [25, 27] and references therein. Our focus here is on the case when $\mathcal{P}$ is a compact set in weak topology. A particular instance for this is when Ξ is a compact set.

At this point, we need to distinguish model (5.1) from one stage distributionally robust SMPEC. The latter focuses on the case where a decision maker is unaware of the true distribution of the underlying uncertainty and the optimal decision is taken on the basis of the worst distribution and worst equilibrium. In that case, only

a single stochastic complementarity constraint based on the worst probability distribution from the ambiguity set $\mathcal{P}$ is considered.

We now turn to discuss the construction of the ambiguity set $\mathcal{P}$. In many practical circumstances, it might be possible to construct a nominal distribution, denoted by P_0, with samples and/or historical data. It is then natural to construct $\mathcal{P}$ as a set of distributions within certain range of P_0. A popular way to quantify such a range is Kullback-Leibler divergence (or relative entropy), that is

$$\mathcal{P} := \{Q \in \mathbb{P} : H(Q|P_0) \leq c\},$$

where $\mathbb{P}$ denotes the set of all probability distributions, $H(Q|P_0)$ is defined by (2.2) and it is known as the relative entropy, c is a positive constant.

For a random variable Z, let

$$\rho_c(Z) := \sup_{\{Q\in\mathbb{P}: H(Q|P_0)\leq c\}} \mathbb{E}_Q[-Z]. \tag{5.2}$$

In what follows, we will represent the distributionally robust constraints of MPDRE (5.1) in terms of $\rho_c(\cdot)$ and then approximate the latter through the entropic risk measure $e_\gamma(\cdot)$ defined in Section 2. To this end, we consider an NLP relaxation for MPDRE (5.1):

$$\text{(RMPDRE)} \qquad \begin{array}{ll} \min\limits_{x\in X_+} & f(x) \\ \text{s.t.} & \left.\begin{array}{l} \sup\limits_{P\in\mathcal{P}} \mathbb{E}_P[-F_j(x,\xi)] \leq 0 \\ \sup\limits_{P\in\mathcal{P}} \mathbb{E}_P[x_j \cdot F_j(x,\xi)] \leq \tau \end{array}\right\} \text{for } j \in \{1,\cdots,n\}, \end{array} \tag{5.3}$$

where τ is a fixed positive number. Through (5.2), we can reformulate problem (5.3) as

$$\begin{array}{ll} \min\limits_{x\in X_+} & f(x) \\ \text{s.t.} & \left.\begin{array}{l} \rho_c(F_j(x,\xi)) \leq 0 \\ \rho_c(-x_j \cdot F_j(x,\xi)) \leq \tau \end{array}\right\} \text{for } j \in \{1,\cdots,n\}. \end{array}$$

Let $P(Z) := \text{Prob}\{Z = \text{ess inf}\, Z\}$. For given c, if $F_j(\cdot,\xi)$ and x satisfy

$$c < -\ln\left(P(F_j(x,\xi))\right), \tag{5.4}$$

then it follows by [18, Proposition 3.1]

$$\rho_c(F_j(x,\xi)) = \min_{\gamma_1^j>0} \frac{c}{\gamma_1^j} + e_{\gamma_1^j}(F_j(x,\xi)).$$

In the case when (5.4) fails to hold, [18, Proposition 3.1] ensures

$$\rho_c(F_j(x,\xi)) = \text{ess sup}(-F_j(x,\xi)) = \lim_{\gamma_1^j\to\infty} \frac{c}{\gamma_1^j} + e_{\gamma_1^j}(F_j(x,\xi)).$$

Indeed, it follows by [21, Proposition 2],

$$\lim_{\gamma_1^j\to\infty} \frac{c}{\gamma_1^j} + e_{\gamma_1^j}(F_j(x,\xi)) = \inf_{\gamma_1^j>0} \frac{c}{\gamma_1^j} + e_{\gamma_1^j}(F_j(x,\xi)).$$

Summarizing the discussions above, we conclude that

$$\rho_c(F_j(x,\xi)) = \begin{cases} \min\limits_{\gamma_1^j>0} \frac{c}{\gamma_1^j} + e_{\gamma_1^j}(F_j(x,\xi)), & \text{for } c < -\ln\left(P(F_j(x,\xi))\right), \\ \inf\limits_{\gamma_1^j>0} \frac{c}{\gamma_1^j} + e_{\gamma_1^j}(F_j(x,\xi)), & \text{for } c \geq -\ln\left(P(F_j(x,\xi))\right). \end{cases}$$

Likewise,

$$\rho_c(-x_j \cdot F_j(x,\xi)) = \begin{cases} \min\limits_{\gamma_2^j>0} \frac{c}{\gamma_2^j} + e_{\gamma_2^j}(-x_j \cdot F_j(x,\xi)), & \text{for } c < -\ln\left(P(-x_j \cdot F_j(x,\xi))\right), \\ \inf\limits_{\gamma_2^j>0} \frac{c}{\gamma_2^j} + e_{\gamma_2^j}(-x_j \cdot F_j(x,\xi)), & \text{for } c \geq -\ln\left(P(-x_j \cdot F_j(x,\xi))\right). \end{cases}$$

The discussions above show that we can recast problem (5.3) in a unified form:

$$\begin{array}{ll} \min\limits_{x\in X_+} & f(x) \\ \text{s.t.} & \left.\begin{array}{l} \inf\limits_{\gamma_1^j>0} \frac{c}{\gamma_1^j} + e_{\gamma_1^j}(F_j(x,\xi)) \le 0 \\ \inf\limits_{\gamma_2^j>0} \frac{c}{\gamma_2^j} + e_{\gamma_2^j}(-x_j\cdot F_j(x,\xi)) \le \tau \end{array}\right\} \text{for } j\in\{1,\cdots,n\}, \end{array} \tag{5.5}$$

where $e_{\gamma_1^j}(\cdot)$ and $e_{\gamma_2^j}(\cdot)$ are defined in (2.1) and the involved mathematical expectation is taken with respect to P_0. For brevity, we write $\boldsymbol{\gamma_1} = (\gamma_1^1,\cdots,\gamma_1^n)$ and $\boldsymbol{\gamma_2} = (\gamma_2^1,\cdots,\gamma_2^n)$. It is not difficult to verify that problem (5.5) is equivalent to

$$\text{(EA-RMPDRE)} \quad \begin{array}{ll} \inf\limits_{x\in X_+,\, \boldsymbol{\gamma_1}\in\mathbb{R}^n_{++},\, \boldsymbol{\gamma_2}\in\mathbb{R}^n_{++}} & f(x) \\ \text{s.t.} & \left.\begin{array}{l} \frac{c}{\gamma_1^j} + e_{\gamma_1^j}(F_j(x,\xi)) \le 0 \\ \frac{c}{\gamma_2^j} + e_{\gamma_2^j}(-x_j\cdot F_j(x,\xi)) \le \tau \end{array}\right\} \text{for } j\in\{1,\cdots,n\}, \end{array} \tag{5.6}$$

where $\mathbb{R}^n_{++}$ denotes the set of n-dimensional vectors whose components are strictly positive.

In many practical cases, P_0 is often constructed through samples such as empirical data. Assume for the simplicity of discussion that the samples are iid. We consider the following sample average approximation of problem (5.6)

$$\text{(SAA-EA-RMPDRE)} \quad \begin{array}{ll} \inf\limits_{x\in X_+,\, \boldsymbol{\gamma_1}\in\mathbb{R}^n_{++},\, \boldsymbol{\gamma_2}\in\mathbb{R}^n_{++}} & f(x) \\ \text{s.t.} & \left.\begin{array}{l} e_1^{j,N}(x,\gamma_1^j) \le 0, \\ e_2^{j,N}(x,\gamma_2^j) \le \tau \end{array}\right\} \text{for } j\in\{1,\cdots,n\}, \end{array} \tag{5.7}$$

where

$$e_1^{j,N}(x,\gamma_1^j) := \frac{c}{\gamma_1^j} + \frac{1}{\gamma_1^j}\ln\left(\frac{1}{N}\sum_{i=1}^N e^{-\gamma_1^j F_j(x,\xi^i)}\right),$$
$$e_2^{j,N}(x,\gamma_2^j) := \frac{c}{\gamma_2^j} + \frac{1}{\gamma_2^j}\ln\left(\frac{1}{N}\sum_{i=1}^N e^{\gamma_2^j x_j\cdot F_j(x,\xi^i)}\right).$$

The following theorem states the convergence of SAA-EA-RMPDRE (5.7) as the sample size N increases. For the simplicity of notation again, let $w := (x,\gamma_1^1,\cdots,\gamma_1^n,\gamma_2^1,\cdots,\gamma_2^n)$.

THEOREM 5.1. **(Convergence of SAA-EA-RMPDRE (5.7))** *Let W^N and W denote the sets of optimal solutions, v_γ^N and v_γ the corresponding optimal values of problems (5.7) and (5.6) respectively. Suppose that (a) W^N and W are nonempty and bounded, (b) there exists $w\in W$ which is strictly feasible; (c) X is a compact set, $F(x,\xi)$ is measurable and integrably bounded. Then the following assertions hold.*

(i) w.p.1 $\lim\limits_{N\to\infty} v_\gamma^N \to v_\gamma$;
(ii) w.p.1 $\overline{\lim\limits_{N\to\infty}} W^N \subseteq W$.

We omit the proof as it is a direct application of Lemma 3.1. Note that the distributionally robust linear optimization problems were studied by Ben-Tal et al [10] where the ambiguity set is characterized by ϕ-divergence with KL-divergence being a special case. More recently, Hu and Hong [21] investigated a distributionally robust convex optimization problem with ambiguity set being defined through KL-divergence, they reformulated the minimax distributionally robust optimization problem into a one-layer convex minimization problem via a dualization approach. Our formulation (5.5) may be regarded as an extension of the works [10, 21] to stochastic programs with equilibrium constraints.

6. Numerical tests. We have carried out some numerical experiments on the SAA-EA-PDL-RSMPEC scheme (4.6) for solving a two stage SMPEC and the EA-RMPDRE scheme (5.6) for solving a mathematical program with distributional robust constraints in Matlab R2008a installed in a PC with Windows XP operating system. In this section, we report some details of the tests and results. Throughout the tests, we employed

the random number generator `rand` in Matlab R2008a to generate the samples and solver `fmincon` to solve problems.

EXAMPLE 6.1. *Consider the following two stage SMPEC*

$$\begin{array}{ll} \min\limits_{x,y(\cdot)} & \mathbb{E}[2x - y(\xi)] \\ \text{s.t.} & x \in [2,6], \\ & 0 \le y(\xi) \perp y(\xi) + \sin\xi - x \ge 0, \quad \forall \xi \in \Xi, \end{array} \tag{6.1}$$

where x is the first stage decision variable and $y(\cdot)$ is the second stage decision variable, ξ is a random parameter following the uniform distribution over set $\Xi := [-\pi, \pi]$.

Observe that for every $x \in [2,6]$, $y(\xi) = x - \sin\xi$ is the solution of the complementary constraint. Therefore the problem is equivalent to

$$\begin{array}{ll} \min\limits_{x} & \mathbb{E}[x + \sin(\xi)] \\ \text{s.t.} & x \in [2,6]. \end{array}$$

The optimal solution to problem (6.1) is $x^* = 2$ and the optimal value is 2. The second stage optimal solution to problem (6.1) is $y^*(\xi) = x^* - \sin(\xi) = 2 - \sin(\xi)$.

Problem (6.1) is a very simple two stage SMPEC. Our intention here is to test the SAA-EA-PDL-RSMPEC scheme (4.6) with this problem and compare its numerical performance with the existing method for two stage SMPECs, namely, the NLP-regularization based SAA method for two stage SMPECs proposed by Shaprio and Xu [41]. For readers who are not familiar with the latter algorithm, let us note that it is SAA applied to the NLP regularized scheme (4.2). The differences between the regularized SAA scheme and SAA-EA-PDL-RSMPEC scheme (4.6) are two-fold: (a) the polynomial decision rule, (b) the entropic approximation. We want to get a confirmation through the tests that applying the decision rule and the entropic approximation will significantly increase the numerical efficiency.

The numerical results are displayed in Tables 6.1 and 6.2. Here, τ denotes the regularization parameter, N the sample size, γ the parameter in entropy approximation, k the degree of the polynomial rule, *Appr.Sol* the approximate optimal solution and *Appr.Val* the optimal value, t the execution time (seconds). Note that `fmincon` requires an initial point. We set the starting point to be a vector with components 1 for the entropic method and a vector with components 0 except the first component being 1 for Shapiro and Xu's method.

TABLE 6.1
SAA-EA-PDL-RSMPEC scheme for problem (6.1)

τ	N	γ	k	*Appr.Sol*		*Appr.Val*	t
				x^N	y^N (coefficients)		
10^{-1}	200	50	1	2.0025	2.1881, −0.1712	1.7980	0.0667
10^{-2}	500	100	3	2.0000	1.9599, −0.8675, −0.0112, 0.0937	2.0286	0.1450
10^{-3}	1000	150	5	2.0000	1.9990, −0.9865, 0.0003, 0.1553, −0.0001, −0.0057	2.0054	1.0775

TABLE 6.2
Shapiro and Xu's method for problem (6.1)

τ	N	*Appr.Sol*	*Appr.Val*	t
10^{-1}	200	1.0007	1.8752	0.7699
10^{-2}	500	1.3159	2.1538	9.7875
10^{-3}	1000	1.0023	1.8566	63.8623

Table 1 displays the results when the SAA-EA-PDL-RSMPEC scheme (4.6) is applied to problem (6.1). For the fixed regularization parameter $\tau = 10^{-1}$ and the sample size $N = 200$, the linear decision rule ($k = 1$) generates an optimal value of 1.7980 in execution time $t = 0.0667$. Compared to Shapiro and Xu's regularized SAA scheme with the same regularization parameter and sample size (Table 2), the optimal value is 1.8752 and the execution time is $t = 0.7699$. The difference between the optimal values is about 0.09 while the execution time of the latter is 10 times that of the former. If we increase the degree of the polynomial approximation to $k = 5$ and sample size to 1000 while the regularization parameter set 10^{-3}, the difference between the execution

times is almost 60 times in favor of the decision rule. This is fundamentally due to the fact that under the decision rule approach, the number of constraints is independent of the sample size and hence the size of the resulting NLP is fixed. In contrast, Shapiro and Xu's regularized SAA scheme does not enjoy this property. The difference of the optimal values is about 0.14 which means the SAA-EA-PDL-RSMPEC scheme (4.6) provides a reasonable upper bound for the problem. The preliminary tests results raises a promising prospect for the SAA-EA-PDL-RSMPEC scheme (4.6) although more numerical experiments are needed.

Next, we look into the EA-RMPDRE scheme (5.6).

EXAMPLE 6.2. *Consider the following mathematical program with distributionally robust equilibrium constraints*

$$\begin{array}{ll} \min & 100(x-y) \\ \text{s.t.} & x \in [2,8], \\ & \left.\begin{array}{l} \mathbb{E}_P[7\sin\xi - x] \le 0 \\ 0 \le y \perp \mathbb{E}_P[(y+\sin\xi - x)(x-2y)] \ge 0 \end{array}\right\} \forall P \in \mathcal{P}, \end{array} \tag{6.2}$$

where $\mathcal{P} := \{Q \in \mathbb{P} : H(Q|P_0) \le c\}$ *and* $P_0\left(\xi = \frac{i\pi}{8}\right) = \frac{1}{5}$, for $i = 0, 1, 2, 3, 4$. *It is easy to see that for each fixed* $P \in \mathcal{P}$ *and* $x \in [2,8]$, *both* $y = x/2$ *and* $y = \mathbb{E}_P[x - \sin\xi]$ *satisfy the complementary constraint. However, here we require the complmentarity constraint to be held for all* $P \in \mathcal{P}$. *This effectively excludes* $y = \mathbb{E}_P[x - \sin\xi]$ *because it shifts as* P *varies. Therefore the set of feasible solutions to problem (6.2) is* $\{(x, x/2) : x \in [u(c), 8]\}$, *where* $u(c) := \max_{P\in\mathcal{P}} \mathbb{E}_P[7\sin\xi]$ *is calculated from the inequality constraint. The objective function takes a value of* $50x$ *at the feasible solution point* $(x, x/2)$ *and it is minimized at* $(u(c), u(c)/2)$ *with the optimal value being* $50u(c)$.

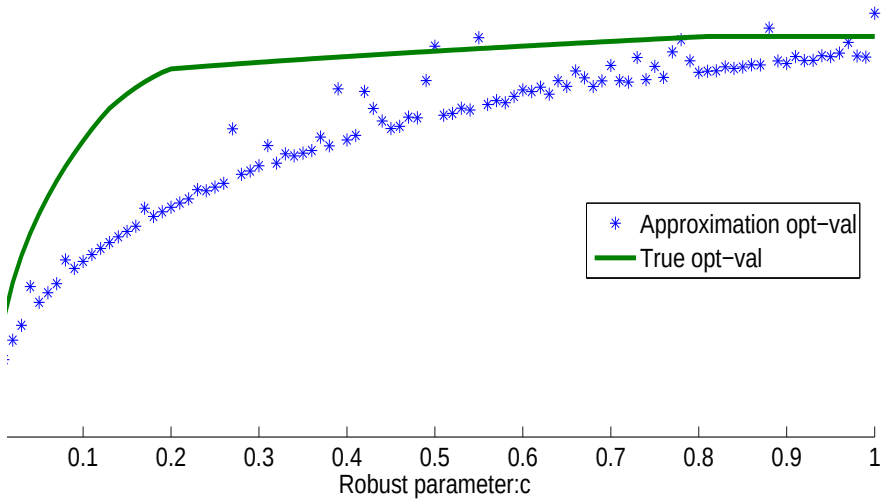

FIG. 6.1. Optimal value of problem (6.2)

We have carried out numerical tests on EA-RMPDRE scheme (5.6) by applying it to problem (6.2) with ambiguity parameter c increasing from 0 to 1. The red star dotted curve in Figure 1 displays the approximate optimal values obtained from solving the EA-RMPDRE scheme (5.6) with c taking values from 101 grid points evenly spread over the interval [0,1]. The blue curve corresponds to the *true* optimal value $50u(c)$ where $u(c)$ is the optimal value of the following convex programming problem

$$\begin{array}{ll} \min\limits_{P} & 50\mathbb{E}_P[-7\sin\xi] \\ \text{s.t.} & H(P|P_0) \le c. \end{array} \tag{6.3}$$

As it is too difficult to obtain a closed form for the optimal value of problem (6.3), we use solver CVX (version 1.2) developed by Michael Grant and Stephen Boyd [19] to solve for each fixed c.

The EA-RMPDRE scheme (5.6) is a nonlinear programming problem. We solve it with an exterior penalty function method and implement the latter through Matlab NLP solver `fmincon`[3]. As we can see from the Figure

[3]We set the maximal penalty parameter as 1000 and the tolerance as 10^{-3}.

1, the approximate optimal values obtained from latter fall below the true optimal values with a few exceptions. The underlying reason is that the regularization scheme and the penalty method enlarge the feasible set of the true problem and the "approximate optimal value" is obtained outside the feasible region. As c increases, the ambiguity set gets larger and hence the set of feasible solutions to problem (5.6) becomes smaller. This explain the overall increasing tendency of the optimal values as c increases. We have also tested the impact of regularization parameter τ. It seems that reducing the value of τ does not help to reduce the gap between the two curves. This is because the optimal solution obtained always satisfies the complementarity constraints (satisfying $y = x/2$) in which case deriving τ to 0 does not help to reinforce the complementarity constraint.

Acknowledgements. We would like to thank the two referees and the Area Editor Mikhail V. Solodov for their constructive comments which significantly help us improve the presentation of the paper.

REFERENCES

[1] F. Alvarez and M. Carrasco, *Minimization of the expected compliance as an alternative approach to multiload truss optimization*, Struct. Multidisc. Optim., 28 (2004), pp. 1-7.
[2] E. Anderson, H. Xu and D. Zhang, Approximating parametric maximization problems through CVaR with applications in minimax and robust convex optimization, manuscript, 2014.
[3] D. Azé, *A survey on error bounds for lower semicontinuous functions*, ESAIM: Proc. 13 (2003), pp. 1-17.
[4] D. Bampou and D. Kuhn, *Polynomial approximations for continuous linear programs*, SIAM J. Optim., 22 (2012), pp. 628-648.
[5] A. Ben-Tal and A. Nemirovski, *Robust truss topology design via semidefinite programming*, SIAM J. Optim., 7 (1997), pp. 991-1016.
[6] A. Ben-Tal and A. Nemirovski, *Robust convex optimization*, Math. Oper. Res., 23 (1998), pp. 769-805.
[7] A. Ben-Tal and A. Nemirovski, *Robust solutions of uncertain linear programs*, Oper. Res. Lett., 25 (1999), pp. 1-13.
[8] A. Ben-Tal, L. El Ghaoui and A. Nemirovski, *Robust Optimization*, Princeton University Press, Princeton, NJ, 2009.
[9] A. Ben-Tal, A. Goryashko, E. Guslitzer and A. Nemirovski, *Adjustable robust solutions of uncertain linear programs*, Math. Program., 99 (2004), pp. 351-376.
[10] A. Ben-Tal, D. Hertog, A.D. Waegenaere, B. Melenberg and G. Rennen, *Robust solutions of optimization problems affected by uncertain probabilities*, Manage. Sci., 59 (2013), pp. 341-357.
[11] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer Series in Operational Research, Springer, New York, 2000.
[12] G. Calafiore and M. C. Campi, *Uncertain convex programs: randomized solutions and confidence levels*, Math. Program., 102 (2005), pp. 25-46.
[13] G. Calafiore and M. C. Campi, *The scenario approach to robust control design*, IEEE Trans. Automatic Control, 51 (2006), pp. 742-753.
[14] G. Calafiore, *Uncertain convex programs*, SIAM J. Optim., 20 (2010), pp. 3427-3464.
[15] X. Chen, M. Sim, P. Sun and J. Zhang, *A linear decision-based approximation approach to stochastic programming*, Oper. Res., 56 (2008), pp. 344-357.
[16] E. Delage and Y. Ye, *Distributionally robust optimization under moment uncertainty with application to data-driven problems*, Oper. Res., 58 (2010), pp. 595-612.
[17] H. Föllmer and A. Schied, *Stochastic Finance: An Introduction in Discrete Time*, Walter de Gruyter, Berlin, 2004.
[18] H. Föllmer and T. Knispel, *Entropic risk measures: coherence vs. convexity, model ambiguity, and robust large deviations*, Sto. Dynam., 11 (2011), pp. 333-351.
[19] M. Grant and S.Boyd, CVX, for convex optimization, download from the homepage http://www.stanford.edu/ boyd/.
[20] J. Goh and M. Sim, *Distributionally robust optimization and its tractable approximations*, Oper. Res., 58 (2010), pp. 902-917.
[21] Z. Hu and J. Hong, Kullback-Leibler divergence constrained distributionally robust optimization, manuscript, 2012.
[22] D. Klatte, *A note on quantitative stability results in nonlinear optimization*, Seminarbericht Nr. 90, Sektion Mathematik, Humboldt-Universität zu Berlin, Berlin, pp. 77-86, 1987.
[23] D. Kuhn, W. Wiesemann and A. Georghiou, *Primal and dual linear decision rules in stochastic and robust optimization*, Math. Program., 130 (2011), pp. 177-209.
[24] G. H. Lin, X. Chen and M. Fukushima, *Solving stochastic mathematical programs with equilibrium constraints via approximation and smoothing implicit programming with penalization*, Math. Program., 116 (2009), pp. 343-368.
[25] G. H. Lin and M. Fukushima, *Stochastic equilibrium problems and stochastic mathematical programs with equilibrium constraints: A survey*, Pac. J. Optim., 6 (2010), pp. 455-482.
[26] Y. Liu, H. Xu and G. H. Lin, *Stability analysis of two stage stochastic mathematical programs with complementarity constraints via NLP-regularization*, SIAM J. Optim., 21 (2011), pp. 609-705.
[27] Y. Liu, H. Xu and J. J. Ye, *Penalized sample average approximation methods for stochastic mathematical programs with complementarity constraints*, Math. Oper. Res., 36 (2011), pp. 670-694.
[28] Z. Q. Luo, J. S. Pang and D. Ralph, *Mathematical Programs with Equilibrium Constraints*, Cambridge University Press, Cambridge, UK, 1996.
[29] F. Meng and H. Xu, *A regularized sample average approximation method for stochastic mathematical programs with nonsmooth equality constraints*, SIAM J. Optim., 17 (2006), pp. 891-919.
[30] J. Outrata, M. Kocvara and J. Zowe, *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints: Theory, Applications, and Numerical Results*, Kluwer Academic Publishers, Dordrect, The Netherlands, 1998.
[31] J-S. Pang, *Error bounds in mathematical programming*, Math. Program., 79 (1997), pp. 299-332.
[32] M. Patriksson and L. Wynter, *Stochastic mathematical programs with equilibrium constraints*, Oper. Res. Lett., 25 (1999), pp. 159-167.
[33] S. J. Qu and M. Goh, Distributionally robust games with an application to supply chain, manuscript, 2012.

[34] R. T. Rockafellar an S. Uryasev, *Conditional value-at-risk for general loss distributions*, J. Bank. Finan., 26 (2002), pp. 1443-1471.
[35] A. Ruszczynski and A. Shapiro, *Stochastic Programming, Handbook in Operations Research and Management Science*, Elsevier, Amsterdam, 2003.
[36] S. Scholtes, *Convergence properties of a regularization sheme for mathematical programs with complementarity constraints*, SIAM J. Optim., 11 (2001), pp. 918-936.
[37] A. Shapiro, *Asymptotic analysis of stochastic programs,* Ann. Oper. Res., 30 (1991), pp. 169-186.
[38] A. Shapiro, *Stochastic mathematical programs with equilibrium constraints*, J. Optim. Theo. Appl., 128 (2006), pp. 223-243.
[39] A. Shapiro, D. Dentcheva and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*, MPS-SIAM Series on Optimization, 2009.
[40] A. Shapiro and A. Nemirovski, *On complexity of stochastic programming problems. V. Jeyakumar, A. Rubinov (eds.) Continuous Optimization: Current Trends and Applications*, pp. 111-144, 2005.
[41] A. Shapiro and H. Xu, *Stochastic mathematical programs with equilibrium constraints, Modeling and sample average approximation*, Optimization, 57 (2008), pp. 395-418.
[42] H. Xu, *An implicit programming approach for a class of stochastic mathematical programs with complementarity constraints*, SIAM J. Optim., 16 (2006), pp. 670-696.
[43] H. Xu, *Uniform exponential convergence of sample average random functions under general sampling with applications in stochastic programming*. J. Math. Anal. Appl., 368 (2010), pp. 692-710.
[44] H. Xu and J. J. Ye, *Necessary optimality conditions for two-stage stochastic programs with equilibrium constraints*, SIAM J. Optim., 20 (2010), pp. 1685-1715.
[45] H. Xu, C. Caramanis and S. Mannor, *A distributional interpretation of robust optimization*, Math. Oper. Res., 37 (2012), pp. 95-110.