Stability Analysis for Mathematical Programs with Distributionally Robust Chance Constraint *

Shaoyan Guo[†], Huifu Xu[‡] and Liwei Zhang[§]

September 24, 2015

Abstract. Stability analysis for optimization problems with chance constraints concerns impact of variation of probability measure in the chance constraints on the optimal value and optimal solutions and research on the topic has been well documented in the literature of s-tochastic programming. In this paper, we extend such analysis to optimization problems with distributionally robust chance constraints where the true probability is unknown, but it is possible to construct an ambiguity set of distributions and the chance constraint is based on the most conservative selection of probability distribution from the ambiguity set. The stability analysis focuses on impact of the variation of the ambiguity set on the optimal value and optimal solutions. We start by looking into continuity of the robust probability function and followed with a detailed analysis of approximation of the function. Sufficient conditions have been derived for continuity of the optimal value and outer semicontinuity of optimal solution set. Case studies are carried out for ambiguity sets being constructed through moments and samples.

Key words. Distributionally robust chance constraint, approximation of ambiguity set, approximation of robust probability function, stability analysis

1 Introduction

Consider the following mathematical programs with chance constraint (MPCC):

$$\begin{array}{ll} \min\limits_{x} & f(x) \\ \text{s.t.} & x \in X, \\ & P(g(x,\xi) \leq 0) \geq 1-\beta, \end{array} \tag{1.1}$$

where X is a compact set of $\mathbb{R}^n$, f and g are continuous functions which map from $\mathbb{R}^n$ and $\mathbb{R}^n \times \mathbb{R}^k$ to $\mathbb{R}$ and $\mathbb{R}^m$ respectively, $\xi : \Omega \to \Xi$ is a vector of random variables defined on

*The research is supported by EPSRC grant EP/M003191/1.

[†]Institute of Operations Research and Control Theory, School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China. (syguomaths@mail.dlut.edu.cn). The work of this author is carried out while she was visiting the second author in the School of Mathematics, Computer Science and Engineering at City University London sponsored by China Scholarship Council.

[‡]School of Mathematics, University of Southampton, Southampton, SO17 1BJ, UK. (h.xu@soton.ac.uk). Haitian Scholar, Dalian University of Technology.

[§]Institute of Operations Research and Control Theory, School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China. (lwzhang@dlut.edu.cn).

probability space $(\Omega, \mathcal{F})$ with closed support set $\Xi \subset \mathbb{R}^k$, $\beta \in (0,1)$ is a given positive scalar, and P represents the probability distribution of ξ.

MPCCs (1.1) have wide applications in engineering, finance and management sciences where the chance constraint may be used to describe likelihood of financial loss, power balance or system control. MPCCs are first discussed by Charnes et al [6], Miller and Wagner [23] and Prékopa [26]. Since then, both theories and applications of MPCCs have been studied extensively. There are generally three major difficulties in solving MPCCs. First, MPCCs might not be convex, and consequently it is difficult to find the global optimal solution. To address this difficulty, previous research identifies some special cases where MPCCs can be reformulated as a tractable convex constraint; see [5, 26]. Second, in order to evaluate the chance constraint, the probability distribution P of random vector ξ must be known. In most applications, however, the true distribution P may be unknown, and only a series of historical data points sampled from the true distribution P are available. Third, even if the true distribution is precisely known, the computation of probability would require evaluation of an integral over a potentially high-dimensional polyhedron, which could be computationally expensive.

One way to overcome some of the aforementioned challenges is to adopt a distributionally robust modelling paradigm where an ambiguity set of distributions, denoted by $\mathcal{P}$, is constructed through empirical data, computer simulation or subjective judgement [32] and the chance constraint is based on the worst-case probability distribution from $\mathcal{P}$, that is,

$$\text{(MPDRCC)} \quad \begin{array}{ll} \min\limits_{x} & f(x) \\ \text{s.t.} & x \in X, \\ & \inf\limits_{P \in \mathcal{P}} P(g(x,\xi) \le 0) \ge 1 - \beta, \end{array} \tag{1.2}$$

The distributionally robust chance constraint (1.2) requires that the chance constraint hold for every $P \in \mathcal{P}$. Construction of the ambiguity set in (1.2) is one of the main topics in MPDRCC. Over the past few decades, various ambiguity sets have been studied in the literature of distributionally robust optimization including those defined through moment conditions ([5, 8, 14, 35, 38, 42]), Wasserstein ball of empirical distribution ([11, 41]), Kullback-Leibler divergence ([18, 19]), Prohorov metric [10] and maximum likelihood ratio [34]. One of the general principles is that $\mathcal{P}$ constitutes the true probability distribution in good faith. In a recent paper, Gupta [13] requires a robust feasible solution to be feasible in problem (1.1) with a specified confidence. Here we focus on problem (1.2) and refer readers to [13] and reference therein for the relationship between problems (1.1) and (1.2). Note that if we consider $(\Xi, \mathscr{B})$ as a measurable space equipped with Borel sigma algebra $\mathscr{B}$, then $\mathcal{P}$ may be viewed as a set of probability measures defined on $(\Xi, \mathscr{B})$ induced by the random variate ξ. We will use terminologies probability measure and probability distribution interchangeably throughout the paper. Moreover, to ease the notation, we will use ξ to denote either the random vector $\xi(\omega)$ or an element of $\mathbb{R}^k$ depending on the context.

An important issue concerning MPDRCC is its numerical tractability. Calafiore and El Ghaoui [5] show that when $g(\cdot,\cdot)$ is bilinear and the ambiguity set is characterized by the known mean and variance, MPDRCC can be converted into a tractable mathematical program with second order cone constraint. Zhao and Guan [39] and Zymler et al [42] analyse the general case when $g(\cdot,\cdot)$ is nonlinear. More recently, Hanasusanto et al [14] investigate a richer class of ambiguity sets that are defined through moment constraints and structural information such as symmetry, unimodality, and independence patterns. Under some conditions, they prove that this class of MPDRCC are computationally tractable.

When MPDRCC is not tractable, different approaches have been proposed to solve its approximation. Nemirovski and Shapiro [24] propose to use Bonferroni's inequality to decompose joint chance constraint to individual chance constraints. However the Bonferroni method may be a poor approximation for problems with joint chance constraint. In order to improve it, Chen et al [7] first use Worst-Case Conditional Value-at-Risk (CVaR) to approximate the joint chance constraint. Recently, Zymler et al [42] obtain the Worst-Case CVaR approximation for the joint chance constraints depending on a set of scaling parameters, which is much tighter than the above two approximations in [7, 24].

Among the above literatures, most focus on the ambiguity set by fixing the first or second moment of a distribution or other structural features, without explicitly taking into account of the data-driven setting. The authors of [5, 8, 19] propose data-driven distributionally robust optimization that the ambiguity set is defined through the sample, however, they mainly investigate tractability and the finite sample guarantee of the resulting reformulation. More recently, Bertsimas et al [3] propose a modification of sample average approximation (SAA), termed robust SAA to the data-driven settings. Using the goodness-of-fit (GoF) hypothesis test, they discuss the finite sample guarantees and asymptotic convergence of robust SAA and prove that Robust SAA yields tractable reformulations for a wide class of cost functions. Here we take a different perspective from [3] to investigate the asymptotic convergence of MPDRCC.

Our focus in this paper is on the case when the ambiguity set $\mathcal{P}$ is approximated by a sequence of ambiguity sets $\{\mathcal{P}_N\}$ and we analyse the impact of the approximation on the optimal value and the optimal solutions to MPDRCC. This is driven not only by the need for appropriate quantification of the uncertainty data but also understanding of asymptotic relationship between statistical estimators of the optimum and the size of uncertainty data. From theoretical perspective, the analysis may be viewed as stability analysis of MPDRCC. Indeed, when $\mathcal{P}$ reduces to a singleton, MPDRCC collapses to MPCC and our analysis coincides with classical stability analysis of MPCC. The research can also be viewed as an extension of stability analysis of distributionally robust formulation of a one stage stochastic program by Sun and Xu [33] where the impact of the optimal value and optimal solutions is investigated against variation of the underlying ambiguity set.

As far as we are concerned, the main contributions of this paper can be summarized as follows.

- **Continuity of the distributionally robust probability function.** We derive under some moderate conditions pointwise continuity the robust probability function (Theorem 3.2), which may be viewed as an extension of similar results for probability functions in the literature of stochastic programming. Moreover, we discuss through some examples sufficient conditions for the continuity of probability function (Theorem 3.1) in comparison with known sufficient conditions for the continuity of probability function.

- **Approximation of the robust probability function and stability analysis of MPDRCC (1.2).** We consider generic approximation of the ambiguity set $\mathcal{P}$ by another ambiguity set $\mathcal{P}_N$ and establish uniform convergence of the corresponding robust probability function when $\mathcal{P}_N$ approximates $\mathcal{P}$ under the pseudo metric (Theorem 4.2). Sufficient conditions are derived for the convergence of the ambiguity set (Theorem 4.1, Propositions 4.1 and 4.2). Moreover, we investigate the impact of variation of the ambiguity set on the optimal value and the optimal solutions to MPDRCC (Theorem 4.3). The new convergence/stability results effectively extend classical stability analysis in stochastic programming.

- **Approximation schemes for some ambiguity sets.** We discuss various approximation schemes for some specific ambiguity sets which lead to the desired approximation of robust probability function and stability analysis including those constructed through moment conditions, continuous distribution with moment conditions, and KL-divergence (Section 5). In the case when the ambiguity set is defined through continuous distribution with moment conditions, we propose uniform distributions on each partition of the support set to approximate the true ambiguity set, and the resulting reformulation can be solved by SAA method (Section 5.1).

Throughout the paper, we will use the following notation. By convention, we use $\mathbb{R}^n$ to represent n dimensional Euclidean space, $\|x\|$ the Euclidean norm of a vector $x \in \mathbb{R}^n$ and $d(x, A) := \inf_{x' \in A} \|x - x'\|$ the distance from a point x to a set A. For two compact sets A and B, we write $\mathbb{D}(A, B) := \sup_{x \in A} d(x, B)$ for the deviation of A from B and $\mathbb{H}(A, B) := \max\{\mathbb{D}(A, B), \mathbb{D}(B, A)\}$ for the Hausdorff distance between A and B. For a $A \subset \mathbb{R}^n$, cl A, int A and bd A denote respectively the closure, interior and boundary of A; $A \backslash B$ denotes the set of points which lie in set A but not in set B, and S^n_+ denotes the space of all $n \times n$ positive semidefinite symmetric matrices. $M \succeq 0$ signifies positive semidefiniteness of matrix M. Finally, we use $\mathbb{B}$ and $\mathbf{B}$ to denote open and closed unit ball in Hilbert space respectively.

2 Preliminaries

2.1 Set-valued mapping

Let $\mathcal{X}, \mathcal{Y}$ be finite dimensional Hilbert spaces and $\Psi : \mathcal{X} \rightrightarrows \mathcal{Y}$ be a set-valued mapping. The *outer limit* of Ψ at $\overline{x}$ is the set

$$\limsup_{x \to \overline{x}} \Psi(x) := \left\{ y \in \mathcal{Y} : \exists x^k \to x, \exists y^k \to y \text{ with } y^k \in \Psi(x^k) \right\},$$

while the *inner limit* of Ψ at $\overline{x}$ is the set

$$\liminf_{x \to \overline{x}} \Psi(x) := \left\{ y \in \mathcal{Y} : \forall x^k \to x, \exists N \in \mathcal{N}_\infty, y^k \overset{N}{\to} y \text{ with } y^k \in \Psi(x^k) \right\},$$

where $\mathcal{N}_\infty := \{N \subseteq \mathcal{N} : \mathcal{N} \backslash N \text{ is finite}\}$ ($\mathcal{N}$ is the natural numbers)[27].

Ψ is said to be *closed-valued* if $\Psi(x)$ is a closed set for each $x \in \mathcal{X}$. A set-valued mapping Ψ is said to be *outer semicontinuous* (osc for short) at $\bar{x} \in \mathcal{X}$ if

$$\limsup_{x \to \overline{x}} \Psi(x) \subset \Psi(\bar{x}).$$

Ψ is said to be *inner semicontinuous* (isc for short) at $\bar{x} \in \mathcal{X}$ if

$$\liminf_{x \to \overline{x}} \Psi(x) \supset \Psi(\bar{x}).$$

Ψ is said to be *continuous* at $\bar{x}$ if it is both osc and isc at $\bar{x}$.

Proposition 2.1 (Characterization of osc and isc properties, see[27]) *Let Ψ be a closed-valued mapping. Ψ is osc at $\bar{x} \in \mathcal{X}$ if and only if for every $\rho > 0$ and $\epsilon > 0$ there is a neighbourhood V of $\bar{x}$ such that*

$$\Psi(x) \cap \rho\mathbf{B} \subset \Psi(\bar{x}) + \epsilon\mathbf{B}, \text{ for all } x \in \mathcal{X} \cap V.$$

Ψ is isc at $\bar{x}$ if and only if for every $\rho > 0$ and $\epsilon > 0$ there is a neighborhood V of $\bar{x}$ such that

$$\Psi(\bar{x}) \cap \rho\mathbf{B} \subset \Psi(x) + \epsilon\mathbf{B}, \text{ for all } x \in \mathcal{X} \cap V.$$

For $f : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$, each function-valued mapping $u \to f(\cdot, u)$ can be associated with a certain set-valued mapping from $\mathbb{R}^m$ to sets in $\mathbb{R}^n \times \mathbb{R}$, namely its *epigraphical mapping* $u \to \text{epi} f(\cdot, u)$, where

$$\text{epi} f(\cdot, u) := \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} : f(x, u) \leq \alpha\}.$$

The epi-continuity of a function-valued mapping is described as follows; see [27, Definition 7.39] and [27, Exercise 7.40] for details.

Definition 2.1 (Epi-lsc) *For $f : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$, the function-valued mapping $u \to f(\cdot, u)$ is epi-lsc at $\overline{u}$ if $u \to \text{epi} f(\cdot, u)$ is outer semicontinuous at $\overline{u}$, or equivalently*

$$\limsup_{u \to \overline{u}} \text{epi} f(\cdot, u) = \text{epi} f(\cdot, \overline{u}).$$

Proposition 2.2 (Characterization of epi-lsc property) *The function-valued mapping $u \to f(\cdot, u)$ is epi-lsc at $\overline{u}$ if and only if for every sequence $u_n \to \overline{u}$ and point $\overline{x} \in \mathbb{R}^n$,*

$$\liminf_{n \to \infty} f(x_n, u_n) \geq f(\overline{x}, \overline{u})$$

for every sequence $x_n \to \overline{x}$.

2.2 Pseudo metric, Kolmogorov metric and total variation metric

Let $\mathscr{P}$ denote the set of all probability measures in the space $(\Xi, \mathscr{B})$. We need appropriate metrics to characterize convergence of probability measures in $\mathscr{P}$.

For each fixed $x \in X$, let

$$H(x) := \{\xi \in \Xi : g(x, \xi) \leq 0\},$$

and

$$\mathbb{I}_{H(x)}(\xi) = \begin{cases} 1 & \text{for } \xi \in H(x), \\ 0 & \text{for } \xi \notin H(x), \end{cases}$$

denote the indicator function of $H(x)$. Then

$$P(g(x, \xi) \leq 0) = \mathbb{E}_P \left[\mathbb{I}_{H(x)}(\xi)\right].$$

We consider the following set of random indicator functions

$$\mathscr{G} := \{\mathbb{I}_{H(x)}(\xi(\cdot)) : x \in X\}. \tag{2.3}$$

For $P, Q \in \mathscr{P}$, let

$$\mathscr{D}(P,Q) := \sup_{g\in\mathscr{G}} \big|\mathbb{E}_P[g] - \mathbb{E}_Q[g]\big| = \sup_{x\in X} \big|P(H(x)) - Q(H(x))\big|. \tag{2.4}$$

We call $\mathscr{D}(P,Q)$ *pseudo metric* in that it satisfies all properties of a metric except that $\mathscr{D}(P,Q) = 0$ does not necessarily imply $P = Q$ unless the set of functions $\mathscr{G}$ is sufficiently large. This type of pseudo metric is widely used for stability analysis in stochastic programming; see an excellent review by Römisch [28].

Let $P \in \mathscr{P}$ be a probability measure and $\mathcal{A}_i \subset \mathscr{P}$, $i = 1, 2$, be two sets of probability measures. With the pseudo metric, the distance from a single probability measure P to a set of probability measures $\mathcal{A}_1$ is defined as $\mathscr{D}(P, \mathcal{A}_1) := \inf_{Q\in\mathcal{A}_1} \mathscr{D}(P,Q)$, the deviation (excess) of $\mathcal{A}_1$ from (over) $\mathcal{A}_2$ as

$$\mathscr{D}(\mathcal{A}_1, \mathcal{A}_2) := \sup_{P\in\mathcal{A}_1} \mathscr{D}(P, \mathcal{A}_2) \tag{2.5}$$

and Hausdorff distance between $\mathcal{A}_1$ and $\mathcal{A}_2$ as

$$\mathscr{H}(\mathcal{A}_1, \mathcal{A}_2) := \max\left\{ \sup_{P\in\mathcal{A}_1} \mathscr{D}(P, \mathcal{A}_2),\ \sup_{Q\in\mathcal{A}_2} \mathscr{D}(Q, \mathcal{A}_1) \right\}. \tag{2.6}$$

Definition 2.2 (Kolmogorov metric) *For two probability measures $P, Q \in \mathscr{P}$, the Kolmogorov metric [12] is defined by*

$$\mathbb{D}_K(P,Q) := \sup_{\eta\in\mathbb{R}^k} |F_P(\eta) - F_Q(\eta)|, \tag{2.7}$$

where F_P and F_Q denote the corresponding cumulative distribution functions (c.d.f) of P and Q respectively. Let $\mathbb{H}_K(\cdot,\cdot)$ denote the Hausdorff distance of the sets of probability measures under the Kolmogorov metric.

Definition 2.3 (Total variation metric) *Let $P, Q \in \mathscr{P}$ and $\mathcal{H}$ denote the set of measurable functions defined in the probability space $(\Xi, \mathscr{B})$. The total variation metric between P and Q is defined as*

$$\mathbb{D}_{TV}(P,Q) := \sup_{h\in\mathcal{H}} \big\{\mathbb{E}_P[h(\xi)] - \mathbb{E}_Q[h(\xi)]\big\},$$

where $\mathcal{H} := \left\{h : \mathbb{R}^k \to \mathbb{R} : h \text{ is measurable}, \sup_{\xi\in\Xi} |h(\xi)| \le 1\right\}$. Let $\mathbb{H}_{TV}(\cdot,\cdot)$ denote the Hausdorff distance of the sets of probability measures under the total variation metric.

2.3 Weak compactness

Let $\{P_N\} \in \mathscr{P}$ be a sequence of probability measures. Recall that $\{P_N\}$ is said to converge to $P \in \mathscr{P}$ *weakly* if

$$\lim_{N\to\infty} \int_\Xi h(\xi) P_N(d\xi) = \int_\Xi h(\xi) P(d\xi), \tag{2.8}$$

for each bounded and continuous function $h : \Xi \to \mathbb{R}$.

For a set of probability measures $\mathcal{A}$ on $(\Xi, \mathscr{B})$, $\mathcal{A}$ is said to be *tight* if for any $\epsilon > 0$, there exists a compact set $\Xi_\epsilon \subset \Xi$ such that $\inf_{P\in\mathcal{A}} P(\Xi_\epsilon) > 1-\epsilon$. In the case when $\mathcal{A}$ is a singleton, it reduces to the tightness of a single probability measure. $\mathcal{A}$ is said to be *closed* (under the weak topology) if for any sequence $\{P_N\} \subset \mathcal{A}$ with P_N converging to P weakly, we have $P \in \mathcal{A}$. $\mathcal{A}$ is said to be *weakly compact* if it is closed and bounded.

By the well-known Prokhorov's theorem (see [1]), a closed set $\mathcal{A}$ (under the weak topology) of probability measures is *compact* if it is tight. In particular, if Ξ is a compact metric space, then the set of all probability measures on $(\Xi, \mathscr{B})$ is compact in that Ξ is in a finite dimensional space; see [31].

Lemma 2.1 (Uniform integrability) *([33, Lemma 1]) Let Z be a separable metric space, P and $\{P_N\}$ be Borel probability measures on Z such that P_N converges to P weakly, let $h : Z \to \mathbb{R}$ be a measurable function with $P(D_h) = 0$, where $D_h = \{z \in Z : h \text{ is discontinuous at } z\}$. Then it holds*

$$\lim_{N\to\infty} \int_Z h(z)P_N(dz) = \int_Z h(z)P(dz),$$

if the sequence $\{P_N h^{-1}\}$ is uniformly integrable*, i.e.,*

$$\lim_{r\to\infty} \sup_{N\in\mathcal{N}} \int_{\{z\in Z: |h(z)|\geq r\}} |h(z)|P_N(dz) = 0.$$

A sufficient condition for the uniform integrability is

$$\sup_{N\in\mathcal{N}} \int_Z |h(z)|^{1+\varepsilon} P_N(dz) < \infty, \text{ for some } \varepsilon > 0. \tag{2.9}$$

Recall that P_N converges to P weakly if and only if the limit (2.8) holds for all continuous and bounded functions h. Lemma 2.1 gives sufficient conditions for the limit to hold for some discontinuous and unbounded function.

2.4 Problem setup

Let $\mathcal{P}_N \subset \mathscr{P}$ be a set of probability distributions which approximates $\mathcal{P}$ in some sense (to be specified later) as N tends to ∞. We consider the following mathematical program with distributionally robust chance constraint:

$$(\text{MPDRCC}_\text{N}) \quad \begin{array}{ll} \min\limits_x & f(x) \\ \text{s.t.} & x \in X, \\ & \inf\limits_{P\in\mathcal{P}_N} P(g(x,\xi) \leq 0) \geq 1-\beta. \end{array} \tag{2.10}$$

Our purpose is to analyse convergence of the optimal value and the optimal solutions of problem (2.10) when $\mathcal{P}_N$ converges to $\mathcal{P}$. In the case when $\mathcal{P}$ reduces to a singleton of the true probability measure, the convergence analysis is well documented in the literature of stochastic programming; see [15, 16, 28, 29] and references therein. Our focus here is the case when $\mathcal{P}$ is a set, e.g., constructed through some moment conditions and $\mathcal{P}_N$ is an approximation regime with some parameters being estimated through empirical data or samples. We will discuss this in detail in Section 5.

Note that for each fixed N and β, we require the feasible set of problem (2.10) to be nonempty to ensure well definedness of the problem. A necessary and sufficient condition is that there exists at least one point $x_0 \in X$ such that

$$\sup_{P \in \mathcal{P}_N} P(\Xi \backslash H(x_0)) \leq \beta, \tag{2.11}$$

which means any probability measure in $\mathcal{P}_N$ must not mass outside $H(x_0)$ above level β. Similar comment applies to the true problem (1.2). We will come back to this in Section 4.

For each fixed $x \in X$, let

$$v(x) := \inf_{P \in \mathcal{P}} P(g(x,\xi) \leq 0) \equiv \inf_{P \in \mathcal{P}} P(H(x)), \tag{2.12}$$

and

$$v_N(x) := \inf_{P \in \mathcal{P}_N} P(g(x,\xi) \leq 0) \equiv \inf_{P \in \mathcal{P}_N} P(H(x)). \tag{2.13}$$

These are *robust probability functions* in the robust chance constraints which determine the feasible set of solutions to MPDRCC and MPDRCC$_N$. A key step towards the desired convergence analysis is to establish uniform convergence of $v_N(x)$ to $v(x)$ over X as N tends to ∞. To this end, we need to derive sufficient conditions for continuity of $v(\cdot)$ and convergence of $\mathcal{P}_N$ to $\mathcal{P}$.

3 Continuity of the robust probability function

An important condition to be used throughout this section is continuity of the set-valued mapping $H(\cdot)$. Unless specified otherwise, the set-valued mapping $H(\cdot)$ may be unbounded. The following result on continuity of $H(\cdot)$ is well known; see e.g. [27, Example 5.10] or [40, Lemma 2.2].

Proposition 3.1 *Suppose that Ξ is convex, and for each fixed $x \in X$, $g_i(x,\cdot)$, $i = 1,\ldots,m$, is convex w.r.t. ξ. If for any $x' \in X$, there exists $\bar{\xi} \in \Xi$ such that $g(x',\bar{\xi}) < 0$, then $H(\cdot)$ is continuous on X.*

The Slater condition in Proposition 3.1 is only a sufficient condition. Let us explain this through an example.

Example 3.1 Consider $g(x,\xi) = \max\{x,\xi\}$, where $X = [-1,0]$ and ξ is uniformly distributed over $[-1,0]$. For fixed $x \in [-1,0]$, it is easy to figure out that $H(x) = [-1,0]$. Therefore $H(\cdot)$ is continuous on X but for any $x \in X$, there does not exist $\xi' \in \Xi$ such that $g(x,\xi') < 0$. However, if we extend X to $[-1,1]$, then $H(x) = \emptyset$ for $x \in (0,1]$. In that case $H(\cdot)$ is not continuous at $x = 0$.

3.1 Continuity of probability function $P(H(\cdot))$

With continuity of set-valued mapping $H(\cdot)$, we are able to establish outer semicontinuity of the set-valued mappings $\Xi\backslash\text{int } H(\cdot)$ and $\text{bd } H(\cdot)$.

Proposition 3.2 *Suppose that $H(\cdot)$ is continuous on X. Then both $\Xi\backslash\text{int } H(\cdot)$ and $\text{bd } H(\cdot)$ are outer semicontinuous on X, i.e., for any $\bar{x} \in X$,*

$$\limsup_{x\to\overline{x}} \Xi\backslash\text{int } H(x) \subset \Xi\backslash\text{int } H(\overline{x}) \tag{3.14}$$

and

$$\limsup_{x\to\overline{x}} \text{bd } H(x) \subset \text{bd } H(\overline{x}). \tag{3.15}$$

Proof. We only prove (3.14) in that (3.15) can be proved analogously. Assume for the sake of a contradiction that (3.14) fails to hold. Then there exists $\bar{u}$ such that

$$\overline{u} \in \left(\limsup_{x\to\overline{x}} \Xi\backslash\text{int } H(x)\right)\backslash(\Xi\backslash\text{int } H(\overline{x})).$$

The relationship enables us to find a sequence $\{x_N\}$ converging to $\bar{x}$ and $u_N \in \Xi\backslash\text{int } H(x_N)$ such that

$$u_N \to \overline{u} \notin \Xi\backslash\text{int } H(\overline{x}).$$

The latter entails $\overline{u} \in \text{int } H(\overline{x})$ and existence of a positive constant δ such that $\overline{u}+\delta\mathbb{B} \subset \text{int } H(\overline{x})$. Since $H(\cdot)$ is continuous on X, for any $\rho > 0$, there exists $N_1 > 0$ such that

$$H(\overline{x}) \cap \rho\mathbf{B} \subset H(x_N) + \frac{\delta}{4}\mathbf{B}$$

for $N \geq N_1$. Let ρ be chosen such that $\rho > \|\bar{u}\| + \delta$. Then

$$\overline{u} + \delta\mathbb{B} \subset H(\overline{x}) \cap \rho\mathbf{B} \subset H(x_N) + \frac{\delta}{4}\mathbf{B}$$

and hence $\overline{u} + \frac{\delta}{8}\mathbf{B} \subset \text{int } H(x_N)$ for $N \geq N_1$. On the other hand, since u_N converges to $\overline{u}$, there exists $N_2 \geq N_1$ such that $u_N \in \overline{u} + \frac{\delta}{8}\mathbf{B} \subset \text{int } H(x_N)$ for $N > N_2$. This contracts the fact that $u_N \in \Xi\backslash\text{int } H(x_N)$. ■

With Proposition 3.2, we are able to derive continuity of the probability function which is one of the main results in this section.

Theorem 3.1 (Continuity of the probability function) *Assume that $H(\cdot)$ is continuous on X and P satisfies*

$$P(H(x)\backslash\text{int } H(x)) = 0, \forall x \in X. \tag{3.16}$$

Then the probability function $P(H(\cdot)) : \mathbb{R}^n \to \mathbb{R}$ is continuous on X.

Proof. We prove the conclusion by showing $P(H(\cdot))$ is both upper and lower semicontinuous. Since $H(x)$ is closed for any $x \in X$, the upper semicontinuity of $P(H(\cdot))$ follows from [29, Proposition 3.1]. Thus it suffices to show the lower semicontinuity. We do so in three steps.

Step 1. For any fixed x, consider the indicator function $\mathbb{I}_{\text{int } H(x)}(\cdot)$. This function is lower semicontinuous over Ξ and

$$\text{epi } \mathbb{I}_{\text{int } H(x)}(\cdot) = H(x) \times [1, +\infty) \bigcup \Xi \backslash \text{int } H(x) \times [0, +\infty). \tag{3.17}$$

Step 2. By the continuity of $H(\cdot)$ and outer semicontinuity of $\Xi \backslash \text{int } H(\cdot)$ established in Proposition 3.2, we have for any fixed $\overline{x} \in X$,

$$\begin{aligned} &\limsup_{x \to \overline{x}} H(x) \times [1, +\infty) \bigcup \Xi \backslash \text{int } H(x) \times [0, +\infty) \\ &\subset H(\overline{x}) \times [1, +\infty) \bigcup \Xi \backslash \text{int } H(\overline{x}) \times [0, +\infty). \end{aligned} \tag{3.18}$$

Since $\bar{x}$ is taken arbitrarily from X, we conclude from a combination of (3.17) and (3.18) that the set-valued mapping $x \to \text{epi } (\mathbb{I}_{\text{int } H(x)}(\cdot))$ is outer semicontinuous, i.e.,

$$\limsup_{x \to \bar{x}} \text{epi } (\mathbb{I}_{\text{int } H(x)}(\cdot)) = \text{epi } (\mathbb{I}_{\text{int } H(\overline{x})}(\cdot)).$$

Through Proposition 2.2, the latter gives rise to

$$\liminf_{x \to \overline{x}} \mathbb{I}_{\text{int } H(x)}(\xi) \geq \mathbb{I}_{\text{int } H(\overline{x})}(\xi)$$

for almost every $\xi \in \Xi$. In other words, $\mathbb{I}_{\text{int } H(\cdot)}(\xi)$ is lower semicontinuous at $\bar{x}$ for almost every ξ. By Fatou's lemma, the above inequality implies

$$\liminf_{x \to \overline{x}} \mathbb{E}\left[\mathbb{I}_{\text{int } H(x)}(\xi)\right] \geq \mathbb{E}\left[\liminf_{x \to \overline{x}} \mathbb{I}_{\text{int } H(x)}(\xi)\right] \geq \mathbb{E}\left[\mathbb{I}_{\text{int } H(\overline{x})}(\xi)\right]. \tag{3.19}$$

Step 3. Observe that the difference between $\mathbb{I}_{H(x)}(\cdot)$ and $\mathbb{I}_{\text{int } H(x)}(\cdot)$ occurs only over the set $H(x) \backslash \text{int } H(x)$. Since P satisfies $P(H(x) \backslash \text{int } H(x)) = 0$ for all $x \in X$, then $\mathbb{E}[\mathbb{I}_{\text{int } H(x)}(\xi)] = P(H(x))$ and $\mathbb{E}[\mathbb{I}_{\text{int } H(\overline{x})}(\xi)] = P(H(\overline{x}))$. Substituting these relations into (3.19), we immediately get

$$\liminf_{x \to \overline{x}} P(H(x)) \geq P(H(\overline{x})),$$

the lower semicontinuity of $P(H(\cdot))$ at $\bar{x}$ holds. ■

It might be helpful to make some comments about Theorem 3.1. First, continuity of probability function has been well investigated in stochastic programming; see for example [25, Proposition 2.1] for a recent result in this regard. A widely used sufficient condition for the continuity is

$$P(\xi \in \Xi : g(x, \xi) = 0) = 0, \ \forall x \in X. \tag{3.20}$$

To understand this condition, we note that

$$P(H(x)) = \mathbb{E}[\mathbb{I}_{H(x)}(\xi)] = \mathbb{E}[\mathbb{I}_{(-\infty, 0]}(g(x, \xi))].$$

It is easy to see that discontinuity occurs when the random indicator function $\mathbb{I}_{(-\infty,0]}(g(x, \xi))$ switches from 1 to 0 with positive measure and this is effectively ruled out by condition (3.20).

Second, in some circumstances, our conditions (namely (3.16) and continuity of $H(\cdot)$) is weaker than condition (3.20). To see this, let us revisit Example 3.1. When $X = [-1, 0]$, int $H(0) = (-1, 0)$, hence condition (3.16) is fulfilled. Moreover $P(H(\cdot))$ is continuous at 0. On the other hand, $P(g(x, \xi) = 0) = P([-1, 0]) = 1$, hence condition (3.20) fails at 0. Third, condition (3.16) holds when probability measure P is absolutely continuous w.r.t. the Lebesgue measure on $\mathbb{R}^k$. This is satisfied by many practically interesting continuous probability measures/distributions. The claim of the proposition says that if $H(\cdot)$ is continuous and the probability measure P at the boundary of $H(x)$ is zero, then $P(H(\cdot))$ is continuous at x. This paves the way for us to look into continuity of $v(\cdot)$.

Example 3.2 Consider $g(x, \xi) = \max\{0, x + \xi\}$, where $X = [1, 2]$ and ξ is normally distributed over $\mathbb{R}$. For fixed $x \in [1, 2]$, it is easy to work out that $H(x) = (-\infty, -x]$ and $\Xi \backslash \text{int } H(x) = [-x, +\infty)$. Therefore both $H(\cdot)$ and $\Xi \backslash \text{int } H(\cdot)$ are continuous on X. Condition (3.16) requires $P(\{-x\}) = 0$ for any $x \in X$, whereas condition (3.20) requires $P(g(x, \xi) = 0) = P(H(x)) = 0$ for any $x \in X$, which is stronger than condition (3.16). In this case, under the condition (3.20), the feasible set of chance constraint, $P(H(x)) \geq \beta$, is empty.

Example 3.3 Consider $g(x, \xi) = x + \xi$, where $X = [-1, 0]$ and ξ is uniformly distributed over $[0, 2]$. For any fixed $x \in X$, $H(x) = [0, -x]$ is continuous at x. Observe that $H(-1) = [0, 1]$ and $H(-1) \backslash \text{int } H(-1) = \{0, 1\}$. On the other hand, $\{\xi \in \Xi : x + \xi = 0\} = \{1\}$ for $x = -1$. In this case, both condition (3.16) and condition (3.20) are satisfied.

3.2 Continuity of robust probability function $v(\cdot)$

We now proceed to discuss pointwise continuity of the robust probability function $v(\cdot)$ over X. We need the following intermediate technical results.

Proposition 3.3 *Let $\{P_N\}$ be a sequence of probability measures converging to P weakly, let $\overline{x}$ be any fixed point in X and $\{x_N\} \subset X$ be any sequence converging to $\bar{x}$. Assume: (a) $\{P_N\}$ and P are tight, and (b) equality (3.16) holds for the probability measure P. Then*

$$\lim_{N \to \infty} P_N(H(x_N) \backslash H(\overline{x})) = 0 \tag{3.21}$$

and

$$\lim_{N \to \infty} P_N(H(\overline{x}) \backslash H(x_N)) = 0. \tag{3.22}$$

Proof. The thrust of the proof is to exploit weak convergence of P_N to P, the continuity of $H(\cdot)$ and condition (3.16). For any $\epsilon > 0$, the tightness condition (a) ensures existence of a sufficiently large number ρ such that

$$\sup_N P_N(\Xi \backslash (\Xi \cap \rho \mathbf{B})) \leq \epsilon \text{ and } P(\Xi \backslash (\Xi \cap \rho \mathbf{B})) \leq \epsilon. \tag{3.23}$$

Observe that

$$H(x) \backslash (H(x) \cap \rho \mathbf{B}) \subset \Xi \backslash (\Xi \cap \rho \mathbf{B}), \ \forall x \in X.$$

Thus (3.23) entails

$$\sup_N \sup_{x\in X} P_N(H(x)\backslash(H(x)\cap\rho\mathbf{B}))\le\epsilon \text{ and } \sup_{x\in X} P(H(x)\backslash(H(x)\cap\rho\mathbf{B}))\le\epsilon. \tag{3.24}$$

Let us first prove (3.21). Since $H(\overline{x})\backslash\text{int } H(\overline{x})$ is closed, then

$$H(\overline{x})\backslash\text{int } H(\overline{x}) = \bigcap_{\delta>0}(H(\overline{x})\backslash\text{int } H(\overline{x})+\delta\mathbf{B})$$

which implies

$$P(H(\overline{x})\backslash\text{int } H(\overline{x})) = \inf_{\delta>0} P(H(\overline{x})\backslash\text{int } H(\overline{x})+\delta\mathbf{B}).$$

Therefore, for the specified $\epsilon>0$, there exists $\delta_0>0$ such that

$$P(H(\overline{x})\backslash\text{int } H(\overline{x})+\delta_0\mathbf{B})\le P(H(\overline{x})\backslash\text{int } H(\overline{x}))+\epsilon=\epsilon,$$

where the equality holds due to our assumption that P satisfies (3.16). On the other hand, the weak convergence of P_N to P and closedness of set $H(\overline{x})\backslash\text{int } H(\overline{x})+\delta_0\mathbf{B}$ enable us to obtain through [4, Theorem 2.1]

$$\limsup_{N\to\infty} P_N\big(H(\overline{x})\backslash\text{int } H(\overline{x})+\delta_0\mathbf{B}\big)\le P\big(H(\overline{x})\backslash\text{int } H(\overline{x})+\delta_0\mathbf{B}\big)\le\epsilon,$$

which in turn means that there exists $N_1>0$ such that

$$P_N\big(H(\overline{x})\backslash\text{int } H(\overline{x})+\delta_0\mathbf{B}\big)\le 2\epsilon, \tag{3.25}$$

when $N\ge N_1$. For the specified ρ and δ_0, the continuity of $H(\cdot)$ ensures existence of a positive number N_2 such that

$$H(x_N)\cap\rho\mathbf{B}\subset H(\overline{x})+\delta_0\mathbf{B}$$

for $N\ge N_2$. Thus

$$(H(x_N)\cap\rho\mathbf{B})\backslash H(\overline{x})\subset(H(x_N)\cap\rho\mathbf{B})\backslash\text{int } H(\overline{x})\subset H(\overline{x})\backslash\text{int } H(\overline{x})+\delta_0\mathbf{B}. \tag{3.26}$$

Combining (3.24), (3.25) and (3.26), we obtain

$$\begin{aligned}P_N(H(x_N)\backslash H(\overline{x})) &\le P_N\left((H(x_N)\cap\rho\mathbf{B})\backslash H(\overline{x})\right)+\epsilon\\ &\le P_N\left((H(\overline{x})\backslash\text{int } H(\overline{x}))+\delta_0\mathbf{B}\right)+\epsilon\\ &\le 3\epsilon \end{aligned}\tag{3.27}$$

for $N\ge\max\{N_1,N_2\}$. The conclusion follows as ϵ can be arbitrarily small.

We now turn to prove (3.22). Let ρ, ϵ and δ_0 be fixed as in Part (i). It follows from Part (ii) of Proposition 3.2 that $H(\cdot)\backslash\text{int } H(\cdot)$ is outer semicontinuous. Together with the continuity of $H(\cdot)$, we can find $N_3>0$ such that

$$\begin{aligned}(H(\overline{x})\cap\rho\mathbf{B})\backslash H(x_N) &\subset (H(\overline{x})\cap\rho\mathbf{B})\backslash\text{int }(H(x_N))\\ &\subset (H(x_N)+\delta_0/2\mathbf{B})\backslash\text{int }(H(x_N))\cap\rho\mathbf{B}\\ &\subset (H(x_N)\backslash\text{int }(H(x_N)))\cap(\rho+\delta_0/2)\mathbf{B}+\delta_0/2\mathbf{B}\\ &\subset H(\overline{x})\backslash\text{int } H(\overline{x})+\delta_0\mathbf{B},\end{aligned}$$

for $N \geq N_3$, where the second inclusion holds due to the continuity of $H(\cdot)$ and the fourth inclusion holds due to the outer semicontinuity of bd $H(\cdot)$ (since $H(x)\backslash\text{int } H(x) = \text{bd } H(x)$). Similar to the analysis in Part (i), we deduce

$$\begin{aligned} P_N(H(\bar{x})\backslash H(x_N)) &\leq P_N\left(H(\bar{x}) \cap \rho\mathbf{B})\backslash H(x_N)\right) + \epsilon \\ &\leq P_N(H(\bar{x})\backslash\text{int } H(\bar{x}) + \delta_0\mathbf{B}) + \epsilon \\ &\leq 3\epsilon \end{aligned}$$

for N being sufficiently large. The proof is complete. ■

With Proposition 3.3, we are ready to derive continuity of the robust probability function $v(\cdot)$.

Theorem 3.2 (Pointwise continuity of the robust probability function) *Suppose that $H(\cdot)$ is continuous on X, $\mathcal{P}$ is weakly compact and for each $P \in \mathcal{P}$, equality (3.16) holds. Then $v(\cdot)$ is continuous on X.*

Proof. By Theorem 3.1, for each $P \in \mathcal{P}$, $P(H(\cdot))$ is continuous on X. Let $x \in X$ be fixed and $\{x_N\} \subset X$ be a sequence such that $x_N \to x$ as $N \to \infty$. Since $\mathcal{P}$ is weakly compact and $P \circ \mathbb{I}_{H(x)}^{-1}$ is uniformly integrable, it follows by [33, Proposition 1] that

$$\mathcal{V}_x := \left\{\mathbb{E}_P[\mathbb{I}_{H(x)}(\xi)] : P \in \mathcal{P}\right\}$$

is a compact set. Thus there exists $P_x \in \mathcal{P}$ such that

$$v(x) = \min_{P \in \mathcal{P}} \mathbb{E}_P[\mathbb{I}_{H(x)}(\xi)] = \mathbb{E}_{P_x}[\mathbb{I}_{H(x)}(\xi)].$$

Likewise, there exists $P_{x_N} \in \mathcal{P}$ such that

$$v(x_N) = \min_{P \in \mathcal{P}} \mathbb{E}_P[\mathbb{I}_{H(x_N)}(\xi)] = \mathbb{E}_{P_{x_N}}[\mathbb{I}_{H(x_N)}(\xi)].$$

With the explicit expression of $v(x)$ and $v(x_N)$ and continuity of $P_x(H(\cdot))$, we have that for any $\epsilon > 0$, there exists $N' > 0$ such that for $N \geq N'$

$$\begin{aligned} v(x_N) - v(x) &= \mathbb{E}_{P_{x_N}}[\mathbb{I}_{H(x_N)}(\xi)] - \mathbb{E}_{P_x}[\mathbb{I}_{H(x)}(\xi)] \\ &\leq \mathbb{E}_{P_x}[\mathbb{I}_{H(x_N)}(\xi)] - \mathbb{E}_{P_x}[\mathbb{I}_{H(x)}(\xi)] \leq \epsilon. \end{aligned}$$

The upper semicontinuity follows since ϵ can be arbitrarily small.

Next, we show lower semicontinuity of $v(\cdot)$ at x. Observe first that

$$\begin{aligned} v(x) - v(x_N) &= \mathbb{E}_{P_x}[\mathbb{I}_{H(x)}(\xi)] - \mathbb{E}_{P_{x_N}}[\mathbb{I}_{H(x_N)}(\xi)] \\ &\leq \mathbb{E}_{P_{x_N}}[\mathbb{I}_{H(x)}(\xi)] - \mathbb{E}_{P_{x_N}}[\mathbb{I}_{H(x_N)}(\xi)] \\ &\leq P_{x_N}(H(x)\backslash H(x_N)). \end{aligned}$$

By (3.22) (see Proposition 3.3)

$$\lim_{N \to \infty} P_{x_N}(H(x)\backslash H(x_N)) = 0. \tag{3.28}$$

The conclusion follows as ϵ can be arbitrarily small. ■

Theorem 3.2 extends pointwise continuity of probability function to that of robust probability function. Compared to Theorem 3.1, it requires additional condition, namely weak compactness of $\mathcal{P}$. The condition entails tightness of probability measures in $\mathcal{P}$ needed for Proposition 3.3 as well closeness of $\mathcal{P}$ for well definedness of $v(\cdot)$.

4 Convergence analysis

In this section, we turn to the central theme of this paper, that is, approximation of the ambiguity set $\mathcal{P}$ by $\mathcal{P}_N$ and its impact on the optimal value and the optimal solutions of MPDRCC (1.2). If we regard $\mathcal{P}_N$ as a perturbation of $\mathcal{P}$, then the research is essentially about stability analysis of problem (1.2). A key step in the analysis is to establish uniform convergence of the robust probability function v_N to v over X as $\mathcal{P}_N \to \mathcal{P}$. We start by considering the case when $\mathcal{P}_N$ and $\mathcal{P}$ are singleton and then extend the discussion to general case.

To simplify exposition of discussion, we assume in the rest of this section that $H(\cdot)$ is continuous on X.

4.1 Approximation of the robust probability function

Our first main technical results concern pointwise and uniform convergence of the probability function when both $\mathcal{P}$ and $\mathcal{P}_N$ are singleton and the latter converges to the former weakly.

Theorem 4.1 (Pointwise and uniform approximation of probability function) *Let $\{P_N\} \subset \mathscr{P}$ be a sequence of probability measures and $P \in \mathscr{P}$. Suppose P_N converges to P weakly, and P satisfies equality (3.16). Then the following assertions hold.*

(i) For each fixed $x \in X$,

$$\lim_{N\to\infty} \mathbb{E}_{P_N}[\mathbb{I}_{H(x)}(\xi)] = \mathbb{E}_P[\mathbb{I}_{H(x)}(\xi)]. \tag{4.29}$$

(ii) $\lim_{N\to\infty} \mathscr{D}(P_N, P) = 0.$

The first part of the theorem concerns approximation of $P_N(H(x))$ to $P(H(x))$ when P_N converges to P weakly, it is an extension of classical results on weak convergence of probability measures in that here the integrand is an indicator function which is discontinuous. The result is needed for deriving convergence of probability measures under the pseudo metric in part (ii) of the theorem.

Proof of Theorem 4.1. Part (i). We use Lemma 2.1 to prove the conclusion. It therefore suffices to verify the conditions of the lemma. Let $D_{\mathbb{I}_{H(x)}} := H(x)\backslash \text{int } H(x)$. Then $\mathbb{I}_{H(x)}(\cdot)$ is continuous on $\Xi\backslash D_{\mathbb{I}_{H(x)}}$. Moreover, under condition (3.16), $P\left(D_{\mathbb{I}_{H(x)}}\right) = 0$. Furthermore, $P_N \circ \mathbb{I}_{H(x)}^{-1}$ is uniformly integrable because

$$\sup_{N\in\mathcal{N}} \int_{\Xi} |\mathbb{I}_{H(x)}(\xi)|^2 P_N(d\xi) < 1$$

uniformly w.r.t. $x \in X$ and P_N converges weakly to P. By Lemma 2.1, the limit in (4.29) holds.

Part (ii). By the definition of the pseudo metric

$$\mathscr{D}(P_N, P) = \sup_{g\in\mathscr{G}} \left|\mathbb{E}_{P_N}[g] - \mathbb{E}_P[g]\right| = \sup_{x\in X} \left|\mathbb{E}_{P_N}[\mathbb{I}_{H(x)}(\xi)] - \mathbb{E}_P[\mathbb{I}_{H(x)}(\xi)]\right|.$$

Therefore it suffices to prove that

$$\lim_{N\to\infty} \sup_{x\in X} \left|\mathbb{E}_{P_N}[\mathbb{I}_{H(x)}(\xi)] - \mathbb{E}_P[\mathbb{I}_{H(x)}(\xi)]\right| = 0. \tag{4.30}$$

Assume for a contradiction that (4.30) fails to hold. Then there exist a constant $\delta > 0$ and a sequence $\{x_N\} \subset X$ such that

$$\left|\mathbb{E}_{P_N}[\mathbb{I}_{H(x_N)}(\xi)] - \mathbb{E}_P[\mathbb{I}_{H(x_N)}(\xi)]\right| \geq \delta/2$$

for N being sufficiently large. Since X is a compact set, by taking a subsequence if necessary we may assume without loss of generality that $x_N \to \overline{x} \in X$. By triangle inequality,

$$\begin{aligned}\left|\mathbb{E}_{P_N}[\mathbb{I}_{H(\overline{x})}(\xi)] - \mathbb{E}_P[\mathbb{I}_{H(\overline{x})}(\xi)]\right| \geq & \left|\mathbb{E}_{P_N}[\mathbb{I}_{H(x_N)}(\xi)] - \mathbb{E}_P[\mathbb{I}_{H(x_N)}(\xi)]\right| \\ & -\left|\mathbb{E}_{P_N}[\mathbb{I}_{H(x_N)}(\xi)] - \mathbb{E}_{P_N}[\mathbb{I}_{H(\overline{x})}(\xi)]\right| \\ & -\left|\mathbb{E}_P[\mathbb{I}_{H(\overline{x})}(\xi)] - \mathbb{E}_P[\mathbb{I}_{H(x_N)}(\xi)]\right|. \end{aligned} \tag{4.31}$$

Following by Theorem 3.1, there exists N_0 sufficiently large such that

$$\left|\mathbb{E}_P[\mathbb{I}_{H(\overline{x})}(\xi)] - \mathbb{E}_P[\mathbb{I}_{H(x_N)}(\xi)]\right| \leq \delta/8. \tag{4.32}$$

Let us now estimate the second term at the right hand side of (4.31). Observe first that

$$\begin{aligned}\left|\mathbb{E}_{P_N}[\mathbb{I}_{H(x_N)}(\xi)] - \mathbb{E}_{P_N}[\mathbb{I}_{H(\overline{x})}(\xi)]\right| &= \left|P_N(H(x_N)) - P_N(H(\overline{x}))\right| \\ &\leq P_N(H(x_N)\backslash H(\overline{x})) + P_N(H(\overline{x})\backslash H(x_N)). \end{aligned} \tag{4.33}$$

On the other hand, since the probability measures P and $\{P_N\}$ are induced probability measure defined over $\mathbb{R}^k$, it follows by [4, Theorem 1.3] that P is tight. Moreover, by [1, Theorem 9.3.3] and [4, Theorem 2.6], the weak convergence of P_N to P implies tightness of $\{P_N\}$. With the tightness, we deduce from Proposition 3.3 that

$$P_N(H(x_N)\backslash H(\overline{x})) + P_N(H(\overline{x})\backslash H(x_N)) \leq \delta/8, \tag{4.34}$$

when N is sufficiently large. Combining (4.31)-(4.34), we deduce that

$$\left|\mathbb{E}_{P_N}[\mathbb{I}_{H(\overline{x})}(\xi)] - \mathbb{E}_P[\mathbb{I}_{H(\overline{x})}(\xi)]\right| \geq \delta/4,$$

which leads to a contradiction to (4.29) as desired. The proof is complete. ■

We now move on to investigate uniform convergence of $v_N(x)$ to $v(x)$ over X. To this end we make the following assumptions on $\mathcal{P}$ and $\mathcal{P}_N$.

Assumption 4.1 (Approximation of the ambiguity set under pseudo metric) *$\mathcal{P}$ and $\mathcal{P}_N$ satisfy the following conditions.*

(a) There exists a weakly compact set $\hat{\mathcal{P}} \subset \mathscr{P}$ such that $\mathcal{P} \subset \hat{\mathcal{P}}$, and $\mathcal{P}_N \subset \hat{\mathcal{P}}$ when N is sufficiently large.

(b) $\lim_{N\to\infty} \mathscr{D}(\mathcal{P}_N, \mathcal{P}) = 0$ w.p.1.

(c) $\lim_{N\to\infty} \mathscr{D}(\mathcal{P}, \mathcal{P}_N) = 0$ w.p.1.

Part (a) means both $\mathcal{P}$ and $\mathcal{P}_N$ must be tight. Part (b) requires $\mathcal{P}_N$ upper semiconverge to $\mathcal{P}$ under the pseudo metric whereas part (c) requires $\mathcal{P}_N$ lower semiconverge to $\mathcal{P}$. Parts (b) and (c) imply $\mathscr{H}(\mathcal{P}_N, \mathcal{P}) \to 0$ almost surely as $N \to \infty$. When $\mathcal{P}_N$ and $\mathcal{P}$ have a specific structure, Assumption 4.1 may be verified directly. We will come back to this in the next section. The following proposition gives rise to a sufficient condition for Assumption 4.1 (b) in the absence of concrete structure of the ambiguity sets.

Proposition 4.1 (Sufficient condition for upper semiconvergence of $\mathcal{P}_N$ to $\mathcal{P}$) Assume: (a) Assumption 4.1 (a) holds; (b) $\mathcal{P}_N$ converges to $\mathcal{P}$ weakly, i.e., for every sequence $\{P_N\} \subseteq \mathcal{P}_N$, $\{P_N\}$ has a subsequence $\{P_{N_k}\}$ converging to P with $P \in \mathcal{P}$; (c) for any $P \in \mathcal{P}$, equality (3.16) holds. Then Assumption 4.1 (b) holds.

Proof. The conclusion follows from Theorem 4.1. Indeed, assume for the sake of a contradiction that $\lim_{N\to\infty} \mathscr{D}(\mathcal{P}_N, \mathcal{P}) \neq 0$. Then there exist a positive constant ϵ_0 and a subsequence $\{\mathcal{P}_{N_k}\}$ such that

$$\mathscr{D}(\mathcal{P}_{N_k}, \mathcal{P}) \geq \epsilon_0,$$

i.e., there exists $P_k \in \mathcal{P}_{N_k}$ such that $\mathscr{D}(P_k, \mathcal{P}) \geq \epsilon_0$. Without loss of generality, we assume P_k converges to $P \in \mathcal{P}$ weakly due to condition (b). It follows from Theorem 4.1, $\lim_{k\to\infty} \mathscr{D}(P_k, P) = 0$, a contradiction as desired. ■

It is possible to derive sufficient conditions for Assumption 4.1 (c) as well.

Proposition 4.2 (Sufficient condition for lower semiconvergence of $\mathcal{P}_N$ to $\mathcal{P}$) *Assume: (a) Assumption 4.1 (a) holds and $\mathcal{P}$ is closed; (b) for any $P \in \mathcal{P}$, there exists a sequence $\{P_N\} \in \mathcal{P}_N$ such that P_N converges to P weakly; (c) for any $P \in \mathcal{P}$, equality (3.16) holds. Then Assumption 4.1 (c) holds.*

Proof. Assume for the sake of a contradiction that $\lim_{N\to\infty} \mathscr{D}(\mathcal{P}, \mathcal{P}_N) \neq 0$. Then there exist a positive constant ϵ_0 and a subsequence $\{\mathcal{P}_{N_k}\}$ such that

$$\mathscr{D}(\mathcal{P}, \mathcal{P}_{N_k}) \geq 2\epsilon_0.$$

Since $\mathcal{P}$ is weakly compact, there exists $P^k \in \mathcal{P}$ such that $\mathscr{D}(P^k, \mathcal{P}_{N_k}) \geq 2\epsilon_0$. Under condition (a), we may suppose without loss of generality that P^k converges to $P \in \mathcal{P}$ weakly. Through Theorem 4.1 (ii), the weak convergence and condition (c) imply $\mathscr{D}(P^k, P) \leq \epsilon_0$ for k sufficiently. Using the triangle inequality of the pseudo metric, we have

$$2\epsilon_0 \leq \mathscr{D}(P^k, \mathcal{P}_{N_k}) \leq \mathscr{D}(P^k, P) + \mathscr{D}(P, \mathcal{P}_{N_k}) \leq \mathscr{D}(P, \mathcal{P}_{N_k}) + \epsilon_0. \tag{4.35}$$

For the given P, condition (b) ensures (by taking a subsequence if necessarily) existence of $P_{N_k} \in \mathcal{P}_{N_k}$ such that P_{N_k} converges to P weakly. Since

$$\mathscr{D}(P, \mathcal{P}_{N_k}) \leq \mathscr{D}(P, P_{N_k}),$$

following a similar argument earlier in the proof, the weakly convergence and condition (c) imply $\mathscr{D}(P, P_{N_k}) \to 0$ and this effectively leads to a contradiction as desired through (4.35). ■

A special case is that both $\mathcal{P}$ and $\mathcal{P}_N$ are singleton, and $\mathcal{P}_N$ is constructed through empirical probability measure. The following corollary says that Assumption 4.1 holds in such a case.

Corollary 4.1 Let P be the true probability distribution of ξ satisfying condition (3.16), and $\xi_1, \ldots, \xi_N$ be an independent and identically distributed sample of ξ. Let $P^N := \frac{1}{N}\sum_{i=1}^N \delta_{\xi_i}$, where δ_ξ denotes measure of mass one at point ξ. Then Assumption 4.1 holds.

Proof. Observe first that when $\mathcal{P}$ and $\mathcal{P}_N$ are singleton, Assumption 4.1 (b) coincides with Assumption 4.1 (c). Therefore, it suffices to verify conditions of Theorem 4.1 (ii). By the well known Glivenko-Cantelli Theorem (see e.g. [36]), P_N converges to P weakly. Together with continuity of $H(\cdot)$, this verifies the conditions of Theorem 4.1. ■

Under Assumption 4.1, we are able to establish uniform convergence of $v_N(\cdot)$ to $v(\cdot)$, which is one of the main convergence results in this section.

Theorem 4.2 (Uniform approximation of the robust probability function) *Under Assumption 4.1, $v_N(x)$ converges to $v(x)$ uniformly over X as N tends to ∞, that is,*

$$\lim_{N\to\infty}\sup_{x\in X}|v_N(x)-v(x)|=0.$$

Proof. The proof is similar to that of [33, Theorem 1]. Here we provide some details for completeness.

Let $x \in X$ be fixed. Define $V := \{P(H(x)) : P \in \text{cl}\,\mathcal{P}\}$ and $V_N := \{P(H(x)) : P \in \text{cl}\,\mathcal{P}_N\}$. Under Assumption 4.1, both V and V_N are bounded subsets in $\mathbb{R}$. Let

$$a := \inf_{v\in V} v,\ b := \sup_{v\in V} v,\ a_N := \inf_{v\in V_N} v,\ b_N := \inf_{v\in V_N} v.$$

Let "conv" denote the convex hull of a set. Then the Hausdorff distance between conv$V$ and convV_N is

$$\mathbb{H}(\text{conv}V,\text{conv}V_N)=\max\{|b_N-b|,|a-a_N|\}.$$

Note that

$$b_N-b=\sup_{P\in\mathcal{P}_N}P(H(x))-\sup_{P\in\mathcal{P}}P(H(x)),$$

and

$$a_N-a=\inf_{P\in\mathcal{P}_N}P(H(x))-\inf_{P\in\mathcal{P}}P(H(x)).$$

Therefore,

$$\mathbb{H}(\text{conv}V,\text{conv}V_N)=\max\left\{\left|\sup_{P\in\mathcal{P}_N}P(H(x))-\sup_{P\in\mathcal{P}}P(H(x))\right|,\left|\inf_{P\in\mathcal{P}_N}P(H(x))-\inf_{P\in\mathcal{P}}P(H(x))\right|\right\}.$$

By the definition and the property of the Hausdorff distance (see [17]),

$$\mathbb{H}(\text{conv}V,\text{conv}V_N)\le\mathbb{H}(V,V_N)=\max\left\{\mathbb{D}(V,V_N),\mathbb{D}(V_N,V)\right\},$$

where

$$\begin{aligned}\mathbb{D}(V,V_N)=\sup_{v\in V}\mathbb{D}(v,V_N) &= \sup_{v\in V}\inf_{v'\in V_N}|v-v'|\\ &= \sup_{P\in\mathcal{P}}\inf_{Q\in\mathcal{P}_N}|P(H(x))-Q(H(x))|\\ &\le \sup_{P\in\mathcal{P}}\inf_{Q\in\mathcal{P}_N}\sup_{x\in X}|P(H(x))-Q(H(x))|\\ &= \mathscr{D}(\mathcal{P},\mathcal{P}_N).\end{aligned}$$

Likewise, we can obtain $\mathbb{D}(V_N, V) \le \mathscr{D}(\mathcal{P}_N, \mathcal{P})$. Therefore,

$$\mathbb{H}(\text{conv}V, \text{conv}V_N) \le \mathbb{H}(V, V_N) \le \mathscr{H}(\mathcal{P}_N, \mathcal{P}),$$

which subsequently yields

$$|v_N(x) - v(x)| = \left| \inf_{P\in\mathcal{P}_N} P(H(x)) - \inf_{P\in\mathcal{P}} P(H(x)) \right| \le \mathbb{H}(\text{conv}V, \text{conv}V_N) \le \mathscr{H}(\mathcal{P}_N, \mathcal{P}).$$

Note that x is any point in X and the right hand side of the inequality above is independent of x. By taking supremum w.r.t. x on both sides, we arrive at the conclusion. ■

4.2 Convergence analysis of MPDRCC$_N$

With the uniform convergence of the robust probability function established in the preceding subsection, we are ready to discuss the convergence of the optimal value function and optimal solutions of problem (2.10). To ease the exposition, let

$$\mathcal{F} := \{x \in X : v(x) \ge 1 - \beta\} \;\text{ and }\; \mathcal{F}_N := \{x \in X : v_N(x) \ge 1 - \beta\}$$

denote the feasible set of problems MPDRCC (1.2) and MPDRCC$_N$ (2.10) respectively. We can rewrite problems (1.2) and (2.10) as

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & x \in \mathcal{F}, \end{array} \tag{4.36}$$

and

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & x \in \mathcal{F}_N. \end{array} \tag{4.37}$$

Let $\vartheta := \inf\{f(x) : x \in \mathcal{F}\}$ denote the optimal value function of the problem (4.36), and S the corresponding set of optimal solutions, that is, $S := \{x \in \mathcal{F} : \vartheta = f(x)\}$. Likewise, let

$$\vartheta_N := \inf\{f(x) : x \in \mathcal{F}_N\} \;\text{ and }\; S_N := \{x \in \mathcal{F}_N : \vartheta_N = f(x)\}.$$

Let $\mathcal{F}^s$ denote the set of strict feasible solutions of the problem (4.36), i.e.,

$$\mathcal{F}^s := \{x \in X : v(x) > 1 - \beta\}. \tag{4.38}$$

The following theorem states convergence of problem (4.37) to problem (4.36) in terms of the feasible sets, the optimal value and the optimal solutions.

Theorem 4.3 (Stability of MPDRCC (1.2)) *Suppose: (a) Assumption 4.1 hold; (b)* $\text{cl}\, \mathcal{F}^s \cap S \neq \emptyset$*; (c)* $v(\cdot)$ *is continuous on* X*. Then*

(i) $\lim_{N\to\infty} \mathbb{D}(\mathcal{F}_N, \mathcal{F}) = 0$*;*

(ii) $\lim_{N\to\infty} \vartheta_N = \vartheta$*;*

(iii) $\lim_{N\to\infty} \mathbb{D}(S_N, S) = 0$.

Condition (b) requires problem (4.36) to have a non-isolated optimal solution. It is fulfilled if the feasible set $\mathcal{F}$ is convex or connected. This condition is well adopted for asymptotic convergence in stochastic programming; see [22] and the references therein.

Proof of Theorem 4.3. Part (i). It follows from Theorem 4.2 that $v_N(\cdot)$ converges to $v(\cdot)$ uniformly over X. Together with continuity of $v(\cdot)$, by [37, Lemma 4.2(i)],

$$\lim_{N\to\infty} \mathbb{D}(\mathcal{F}_N, \mathcal{F}) = 0. \tag{4.39}$$

Parts (ii) and (iii). Let x^N be an optimal solution to problem (4.36), i.e., $x^N \in S_N$. Since the sequence is contained in the compact set X, by taking a subsequence, if necessary, we may assume for the simplicity of notation that $x^N \to \overline{x}$. By (4.39), $\overline{x} \in \mathcal{F}$. In what follows, we show that $\overline{x} \in S$. Observe first that since f is continuous, then

$$\lim_{N\to\infty} \vartheta_N = \lim_{N\to\infty} f(x^N) = f(\overline{x}) \geq \vartheta.$$

Moreover, under condition $\mathrm{cl}\, \mathcal{F}^s \cap S \neq \emptyset$, there exists $y^* \in \mathrm{cl}\, \mathcal{F}^s \cap S$. By the continuity of $f(\cdot)$, for any small positive number ϵ, there exists $y^\epsilon \in \mathcal{F}^s$ such that

$$f(y^\epsilon) - \vartheta \leq \epsilon.$$

Since $y^\epsilon \in \mathcal{F}^s$ and $v_N(x)$ converges to $v(x)$ uniformly over X, we can find $y^N \in \mathcal{F}_N$ such that $\|y^N - y^\epsilon\| \to 0$. Therefore,

$$\vartheta \geq f(y^\epsilon) - \epsilon = \lim_{N\to\infty} f(y^N) - \epsilon \geq \lim_{N\to\infty} f(x^N) - \epsilon = f(\overline{x}) - \epsilon,$$

which implies $\vartheta \geq f(\overline{x})$ in that ϵ can be chosen arbitrarily small. This shows $\overline{x} \in S$ and $\lim_{N\to\infty} \vartheta_N = \vartheta$. ■

5 Approximations of the ambiguity set

Having established stability analysis of the robust chance constrained problem in the preceding section, we turn to discuss details of approximation of the ambiguity set $\mathcal{P}$ and examine how the required properties such as Assumption 4.1 may be fulfilled.

5.1 Piecewise uniform approximation and sample average approximation

We start by considering $\mathcal{P}$ being defined through moments conditions. Let $\Phi : \mathbb{R}^k \to \mathbb{R}^q$ be a continuous vector-valued measurable function and $\xi : \Omega \to \mathbb{R}^k$ be a random vector with support set Ξ. We consider

$$\mathcal{P} := \left\{ P \in \mathscr{P} : \begin{array}{l} \mathbb{E}_P[\Phi(\xi)] \leq 0,\ P \text{ is a continuous distribution over } \Xi \\ \text{and absolutely continuous with respect to the Lebesgue measure.} \end{array} \right\}. \tag{5.40}$$

Slightly different from classical moment problems, here we require the underlying probability distribution to be absolutely continuous w.r.t. the Lebesgue measure. A simple example is that the true probability distribution follows a parametric truncated normal distributions with some unknown parameters and the moment conditions are used to specify the range of these parameters.

For the convenience of analysis, we consider the case when the support set Ξ is a bounded rectangle set, i.e., there exist finite valued vectors $L, U \in \mathbb{R}^k$ such that

$$\Xi = \{\xi \in \mathbb{R}^k : L \le \xi \le U\}.$$

Let $\Xi_1, \ldots, \Xi_N$ be a partition of Ξ with

$$\begin{array}{l} \Xi_1 = \left\{\xi : L \le \xi \le L + \frac{1}{N}(U - L)\right\}, \\ \Xi_i = \left\{\xi : L + \frac{i-1}{N}(U - L) < \xi \le L + \frac{i}{N}(U - L)\right\} \quad \text{for } i = 2, \ldots, N. \end{array} \tag{5.41}$$

Let

$$\mathcal{P}_N := \left\{ P \in \mathscr{P} : \begin{array}{c} \mathbb{E}_{P_N}[\Phi(\xi)] \le 0, \ P_N \text{ is uniformly distributed over } \Xi_i, \\ P_N(\Xi_i) = p_i, \ \sum_{i=1}^N p_i = 1, p_i \ge 0, \text{ for } i = 1, \ldots, N. \end{array} \right\} \tag{5.42}$$

We investigate approximation of $\mathcal{P}_N$ to $\mathcal{P}$. Observe first that $\mathcal{P}_N \subset \mathcal{P}$ because the uniform distribution specified in the definition of $\mathcal{P}_N$ is a particular continuous distribution over Ξ which is absolutely continuous w.r.t. the Lebesgue measure. In what follows, we show that $\mathcal{P}_N$ converges to $\mathcal{P}$ under Kolmogorov metric and henceforth $\mathcal{P}_N$ converges to $\mathcal{P}$ weakly (see[12]).

Before proceeding to the convergence analysis, we explain why $\mathcal{P}_N$ is constructed in this particular manner. Suppose that $\mathcal{P}$ is a singleton, that is, the true probability measure is absolutely continuous w.r.t. the Lebesgue measure. In that case, it is natural to use a piecewise linear function to approximate $\mathcal{P}$ because the latter is relatively easier to calculate. Thus, what we are proposing here is to extends the above approximation scheme to the case when $\mathcal{P}$ is defined through some moment conditions. Our conjecture is that $\mathcal{P}_N$ converges to $\mathcal{P}$ under some appropriate metric.

Theorem 5.1 *Suppose that there exists $P_0 \in \mathcal{P}$ such that $\mathbb{E}_{P_0}[\Phi(\xi)] < 0$. Then*

$$\lim_{N\to\infty} \mathbb{H}_K(\mathcal{P}, \mathcal{P}_N) = 0 \tag{5.43}$$

and

$$\lim_{N\to\infty} \mathscr{H}(\mathcal{P}, \mathcal{P}_N) = 0. \tag{5.44}$$

Proof. Let us first prove (5.43). For any fixed $P \in \mathcal{P}$, since $\mathcal{P}$ is a convex set, then for any $0 < \lambda < 1$, $P^\lambda := \lambda P + (1-\lambda)P_0 \in \mathcal{P}$ satisfies $\mathbb{E}_{P^\lambda}[\Phi(\xi)] < 0$.

Let $\epsilon > 0$ be a fixed constant and F^λ be *c.d.f* of P^λ. Since $P^\lambda \in \mathcal{P}$, it is absolutely continuous w.r.t. the Lebesgue measure, then for a sufficiently large N, we have the partition of Ξ defined as in (5.41) satisfying

$$\sup_{1\le i\le N} P^\lambda(\Xi_i) = \sup_{1\le i\le N} F^\lambda\left(L + \frac{i}{N}(U-L)\right) - F^\lambda\left(L + \frac{i-1}{N}(U-L)\right) \le \epsilon/2, \tag{5.45}$$

and there exists P_N^λ uniformly distributed over each Ξ_i such that

$$\Delta_N := \sum_{i=1}^N \left|P_N^\lambda(\Xi_i) - P^\lambda(\Xi_i)\right| \le \epsilon/2$$

for N sufficiently large. In what follows, we show $\mathbb{D}_K(P^\lambda, P_N^\lambda) \leq \epsilon$.

For any $\eta \in [L, U]$, there exists $i \in \{1, \ldots, N\}$ such that $\eta \in \Xi_i$. By definition

$$\begin{aligned}
F_N^\lambda(\eta) - F^\lambda(\eta) &\leq F_N^\lambda\left(L + \frac{i}{N}(U-L)\right) - F^\lambda\left(L + \frac{i-1}{N}(U-L)\right) \\
&\leq F_N^\lambda\left(L + \frac{i}{N}(U-L)\right) - F^\lambda\left(L + \frac{i}{N}(U-L)\right) + \epsilon/2 \\
&= P_N^\lambda(\Xi_1 \cup \ldots \cup \Xi_i) - P^\lambda(\Xi_1 \cup \ldots \cup \Xi_i) + \epsilon/2 \\
&\leq \Delta_N + \epsilon/2,
\end{aligned}$$

where the second inequality holds due to (5.45). Likewise,

$$\begin{aligned}
F_N^\lambda(\eta) - F^\lambda(\eta) &\geq F_N^\lambda\left(L + \frac{i-1}{N}(U-L)\right) - F^\lambda\left(L + \frac{i}{N}(U-L)\right) \\
&\geq F_N^\lambda\left(L + \frac{i-1}{N}(U-L)\right) - F^\lambda\left(L + \frac{i-1}{N}(U-L)\right) - \epsilon/2 \\
&= P_N^\lambda(\Xi_1 \cup \ldots \cup \Xi_{i-1}) - P^\lambda(\Xi_1 \cup \ldots \cup \Xi_{i-1}) - \epsilon/2 \\
&\geq -\Delta_N - \epsilon/2.
\end{aligned}$$

A combination of the two inequalities gives rise to

$$|F_N^\lambda(\eta) - F^\lambda(\eta)| \leq \Delta_N + \epsilon/2.$$

Since the inequality holds for any $\eta \in \mathbb{R}^k$, we have

$$\mathbb{D}_K(P_N^\lambda, P^\lambda) = \sup_{\eta \in \mathbb{R}^k} |F_N^\lambda(\eta) - F^\lambda(\eta)| \leq \Delta_N + \epsilon/2 \leq \epsilon.$$

which means that P_N^λ converges to P^λ under the Kolmogorov metric and by [12, Theorem 6] in weak topology.

Next, we show that P_N^λ satisfies the moment condition in (5.42). Since $\Phi(\cdot)$ is a continuous function, the weak convergence guarantees

$$\lim_{N \to \infty} \mathbb{E}_{P_N^\lambda}[\Phi(\xi)] = \mathbb{E}_{P^\lambda}[\Phi(\xi)].$$

Moreover, since $\mathbb{E}_{P^\lambda}[\Phi(\xi)] < 0$, the limit above ensures $\mathbb{E}_{P_N^\lambda}[\Phi(\xi)] \leq 0$ for N sufficiently large, which means $P_N^\lambda \in \mathcal{P}_N$. By driving λ to one and ϵ to zero, we deduce from the discussions above that there exists a sequence $\{P_N\}$ depending on λ and ϵ with $P_N \in \mathcal{P}_N$ such that

$$\lim_{N \to \infty} \mathbb{D}_K(P, P_N) = 0.$$

This implies for any $P \in \mathcal{P}$, there exists a sequence $\{P_N\} \subset \mathcal{P}_N$ such that P_N converges to P under the Kolmogorov metric. Hence $\lim_{N\to\infty} \mathbb{D}_K(\mathcal{P}, \mathcal{P}_N) = 0$ holds. We can change $\mathbb{D}_K$ to $\mathbb{H}_K$ in that $\mathcal{P}_N \subset \mathcal{P}$.

Finally, we prove (5.44). It is well known in the literature of probability theory that convergence under the Kolmogorov metric implies weak convergence; see [12]. Using this result, we can easily show that convergence of $\mathcal{P}_N$ to $\mathcal{P}$ under the Kolmogorov metric implies weak convergence of $\mathcal{P}_N$ to $\mathcal{P}$. Moreover, since Ξ is compact, $\mathcal{P}$ is tight and closed. By Prokhorov's theorem, $\mathcal{P}$ is compact (see [1] and [33, Proposition 7] for more recent discussions in this regard). Furthermore, for any $P \in \mathcal{P}$, condition (3.16) holds, then it follows by Propositions 4.1 and 4.2 that $\mathcal{P}_N$ converges to $\mathcal{P}$ under the pseudo metric. ■

5.1.1 Sample average approximation

In what follows, we discuss how to solve MPDRCC$_N$ when $\mathcal{P}_N$ is defined as in (5.42). The approximate robust chance constrained minimization problem is

$$\begin{array}{ll} \min\limits_{x} & f(x) \\ \text{s.t.} & x \in X, \\ & \inf\limits_{P \in \mathcal{P}_N} P(H(x)) \geq 1-\beta. \end{array} \tag{5.46}$$

For the given partition $\Xi_i, i = 1, \ldots, N$ of Ξ, let $V(\Xi_i) := \int_{\Xi_i} d\xi$ and $u_i = \frac{p_i}{V(\Xi_i)}$. Then

$$\begin{aligned} \inf_{P \in \mathcal{P}_N} P(H(x)) \;=\; & \inf_{p \in \mathbb{R}^N} \; \sum_{i=1}^N u_i V(H(x) \cap \Xi_i) \\ & \quad \text{s.t.} \; \sum_{i=1}^N u_i \int_{\Xi_i} \Phi_l(\xi) d\xi \leq 0, \quad \text{for } l = 1, \ldots, q, \\ & \qquad\quad u_i = \tfrac{p_i}{V(\Xi_i)}, \\ & \qquad\quad \sum_{i=1}^N p_i = 1, p_i \geq 0, \quad \text{for } i = 1, \ldots, N. \\ \;=\; & \inf_{p \in \mathbb{R}^N} \; \sum_{i=1}^N \frac{p_i}{V(\Xi_i)} V(H(x) \cap \Xi_i) \\ & \quad \text{s.t.} \; \sum_{i=1}^N \frac{p_i}{V(\Xi_i)} \int_{\Xi_i} \Phi_l(\xi) d\xi \leq 0, \quad \text{for } l = 1, \ldots, q, \\ & \qquad\quad \sum_{i=1}^N p_i = 1, p_i \geq 0, \quad \text{for } i = 1, \ldots, N. \end{aligned}$$

The Lagrange dual of the above problem is

$$\begin{array}{ll} \sup\limits_{\lambda_0, \lambda_1, \ldots, \lambda_q} & \lambda_0 \\ \text{s.t.} & \lambda_0 \in \mathbb{R}, \lambda_l \geq 0, \quad \text{for } l = 1, \ldots, q, \\ & \dfrac{V(H(x) \cap \Xi_i)}{V(\Xi_i)} - \lambda_0 + \displaystyle\sum_{l=1}^q \lambda_l \frac{\int_{\Xi_i} \Phi_l(\xi) d\xi}{V(\Xi_i)} \geq 0, \quad \text{for } i = 1, \ldots, N. \end{array}$$

Consequently problem (5.46) can be written as

$$\begin{array}{ll} \min\limits_{x, \lambda_0, \lambda_1, \ldots, \lambda_q} & f(x) \\ \text{s.t.} & x \in X, \lambda_0 \geq 1-\beta, \lambda_l \geq 0, \quad \text{for } l = 1, \ldots, q, \\ & \dfrac{V(H(x) \cap \Xi_i)}{V(\Xi_i)} - \lambda_0 + \displaystyle\sum_{l=1}^q \lambda_l \frac{\int_{\Xi_i} \Phi_l(\xi) d\xi}{V(\Xi_i)} \geq 0, \quad \text{for } i = 1, \ldots, N, \end{array} \tag{5.47}$$

or equivalently

$$\begin{array}{ll} \min\limits_{x, \lambda_0, \lambda_1, \ldots, \lambda_q} & f(x) \\ \text{s.t.} & x \in X, \lambda_0 \geq 1-\beta, \lambda_l \geq 0, \quad \text{for } l = 1, \ldots, q, \\ & \mathbb{E}_{P^i}[\mathbb{I}_{H(x)}(\xi) - \lambda_0 + \displaystyle\sum_{l=1}^q \lambda_l \Phi_l(\xi)] \geq 0, \quad \text{for } i = 1, \ldots, N, \end{array} \tag{5.48}$$

where P^i is a uniform distribution over Ξ_i with $P^i(\Xi_i) = 1$, for $i = 1, \dots, N$.

Note that when $H(\cdot)$ takes some special structure such as polyhedron, the expected values in problem (5.48) might be computed easily. In general, it might be numerically expensive to calculate these expected values. The well-known SAA method might be used to tackle the challenge.

For fixed partition $\{\Xi_1, \dots, \Xi_N\}$ of Ξ, let $\xi_i^1, \dots, \xi_i^{M_i}$ be an independent and identically distributed random variables uniformly distributed over Ξ_i for $i = 1, \dots, N$. We use $\frac{1}{M_i}\sum_{j=1}^{M_i} \mathbb{I}_{H(x)}(\xi_i^j)$, $\frac{1}{M_i}\sum_{j=1}^{M_i} \Phi(\xi_i^j)$ to approximate $\mathbb{E}_{P^i}[\mathbb{I}_{H(x)}(\xi)]$ and $\mathbb{E}_{P^i}[\Phi(\xi)]$ respectively. The resulting SAA scheme of problem (5.48) can be written as

$$\begin{array}{cl} \min\limits_{x,\lambda_0,\lambda_1,\dots,\lambda_q} & f(x) \\ \text{s.t.} & x \in X, \lambda_0 \geq 1-\beta, \lambda_l \geq 0, \quad \text{for } l = 1, \dots, q, \\ & \dfrac{1}{M_i}\displaystyle\sum_{j=1}^{M_i}\left[\mathbb{I}_{H(x)}(\xi_i^j) - \lambda_0 + \sum_{l=1}^{q} \lambda_l \Phi_l(\xi_i^j)\right] \geq 0, \quad \text{for } i = 1, \dots, N. \end{array} \tag{5.49}$$

To justify the approximation scheme, we discuss briefly convergence of program (5.49) to program (5.48) in terms of the optimal value and the optimal solution as M_i increases. To ease notation, let

$$\psi^i(x,\lambda) := \mathbb{E}_{P^i}\left[\mathbb{I}_{H(x)}(\xi)] - \lambda_0 + \sum_{l=1}^{q} \lambda_l \Phi_l(\xi)\right],$$

for $i = 1, \dots, N$, with $\lambda = (\lambda_0, \lambda_1, \dots, \lambda_q) \in \mathbb{R}^{q+1}$.

Proposition 5.1 *Let N be fixed. Let S^N and S_M^N denote the set of the optimal solutions of problems (5.48) and (5.49) respectively, ϑ^N and ϑ_M^N the corresponding optimal values, where $M := \min\{M_1, \cdots, M_N\}$. Assume: (a) $H(\cdot)$ is continuous on X; (b) $\mathrm{cl}\, \mathcal{F}_N^s \cap S^N \neq \emptyset$, where $\mathcal{F}_N^s$ signifies the set of strictly feasible solutions as in (4.38). Then*

$$\lim_{M\to\infty} \vartheta_M^N = \vartheta^N \text{ and } \lim_{M\to\infty} \mathbb{D}(S_M^N, S^N) = 0.$$

Proof. We use Theorem 4.3 to prove this proposition. Therefore it suffices to verify the conditions of the theorem. First, by virtue of Corollary 4.1, Assumption 4.1 holds for each $P^{M_i} := \frac{1}{M_i}\sum_{j=1}^{M_i} \delta_{\xi_i^j}$, which means P^{M_i} converges to P_i under the pseudo metric. Second, we show continuity of $\psi_i(x,\lambda)$ for $i = 1, \dots, N$. Since $H(\cdot)$ is continuous on X, and P_i satisfies condition (3.16), by Theorem 3.2, $\psi_i(x,\lambda)$ is continuous on $X \times \mathbb{R}^{q+1}$. Together with condition (b), all of the conditions of Theorem 4.3 are fulfilled. ■

5.2 Approximation of mean-absolute deviation ambiguity set and numerical tractability

In this subsection, we consider the ambiguity set $\mathcal{P}$ being defined through mean-absolute deviation, namely

$$\mathcal{P} := \{P \in \mathscr{P} : \mathbb{E}_P[\xi] = \mu,\ \mathbb{E}_P[|\xi - \mu|] \leq d\}, \tag{5.50}$$

where $\mu \in \mathbb{R}^k$ and $d \in \mathbb{R}^k_+$ denote the mean value and absolute deviation of ξ respectively.

Let $\{\xi^i\}_{i=1}^N$ be an independent and identically distributed sample drawn from the true distribution P of the random vector ξ. Let

$$\mu_N := \frac{1}{N}\sum_{i=1}^N \xi^i, \quad d_N := \frac{1}{N}\sum_{i=1}^N |\xi^i - \mu_N|.$$

Let

$$\mathcal{P}_N := \{P \in \mathscr{P} : \mathbb{E}_P[\xi] = \mu_N,\ \mathbb{E}_P[|\xi - \mu_N|] \le d_N\} \tag{5.51}$$

be an approximation of $\mathcal{P}$. By the Hoffman's lemma for moment problem in [33], we can quantify the approximation of $\mathcal{P}_N$ to $\mathcal{P}$ through total variation metric.

Proposition 5.2 *If $\Xi = \mathbb{R}^k$, then there exists a positive constant C such that*

$$\mathbb{H}_{TV}(\mathcal{P}_N, \mathcal{P}) \le C\left(2\|\mu_N - \mu\| + \max\{\|(d_N - d)_+\|, \|(d - d_N)_+\|\}\right), \tag{5.52}$$

where $(a)_+ = \max\{a, 0\}$ and the maximum is taken componentwise for a vector a.

Proof. Since $\Xi = \mathbb{R}^k$, the set $\{\mathbb{E}_P[\xi] : P \in \mathscr{P}\} = \mathbb{R}^k$, there exist probability distributions P_0 and P_1 such that $\mathbb{E}_{P_0}[\xi] = \mu, \mathbb{E}_{P_0}[|\xi - \mu|] < d$ and $\mathbb{E}_{P_1}[\xi] = \mu_N, \mathbb{E}_{P_1}[|\xi - \mu|] < d_N$ respectively, which means the Slater condition holds. Then for any $Q \in \mathscr{P}$, it follows from [33, Lemma 2] that there exist positive constants C_1, C_2 satisfying

$$\begin{aligned}
d_{TV}(Q, \mathcal{P}) &\le C_1(\|\mathbb{E}_Q[\xi] - \mu\| + \|(\mathbb{E}_Q[|\xi - \mu|] - d)_+\|), \\
d_{TV}(Q, \mathcal{P}_N) &\le C_2(\|\mathbb{E}_Q[\xi] - \mu_N\| + \|(\mathbb{E}_Q[|\xi - \mu_N|] - d_N)_+\|).
\end{aligned}$$

Moreover, for any given $Q \in \mathcal{P}_N$,

$$\begin{aligned}
d_{TV}(Q, \mathcal{P}) &\le C_1(\|\mathbb{E}_Q[\xi] - \mu_N\| + \|\mu_N - \mu\| + \|(\mathbb{E}_Q[|\xi - \mu_N|] + |\mu_N - \mu| - d_N + d_N - d)_+\|) \\
&\le C_1(\|\mu_N - \mu\| + \|(\mathbb{E}_Q[|\xi - \mu_N|] - d_N)_+\| + \|\mu_N - \mu\| + \|(d_N - d)_+\|) \\
&\le C_1(2\|\mu_N - \mu\| + \|(d_N - d)_+\|),
\end{aligned}$$

where the second inequality holds due to the fact that $\mathbb{E}[(a+b)_+] \le \mathbb{E}[(a)_+] + \mathbb{E}[(b)_+]$. Likewise, for any given $Q \in \mathcal{P}$,

$$d_{TV}(Q, \mathcal{P}_N) \le C_2(2\|\mu_N - \mu\| + \|(d - d_N)_+\|).$$

The conclusion follows by setting $C = \max\{C_1, C_2\}$. ■

Remark 5.1 It might be helpful to make a few comments on Assumption 4.1 in this setting.

(i) $\mathcal{P}$ is tight if there exist positive constants ϵ and C such that

$$\sup_{P \in \mathcal{P}} \int_{\xi \in \Xi} \|\xi\|^{1+\epsilon} P(d\xi) < C.$$

To show this, let $r > 1$ be sufficiently large, then

$$\begin{aligned}\sup_{P\in\mathcal{P}}\int_{\{\xi\in\Xi:\|\xi\|\geq r\}} P(d\xi) &\leq \frac{1}{r}\sup_{P\in\mathcal{P}}\int_{\{\xi\in\Xi:\|\xi\|\geq r\}} r^{1+\epsilon}P(d\xi)\\ &\leq \frac{1}{r}\sup_{P\in\mathcal{P}}\int_{\{\xi\in\Xi:\|\xi\|\geq r\}} \|\xi\|^{1+\epsilon}P(d\xi)\leq C/r,\end{aligned}$$

this means that

$$\lim_{r\to\infty}\sup_{P\in\mathcal{P}}\int_{\{\xi\in\Xi:\|\xi\|\geq r\}} P(d\xi) = 0.$$

By [1, Definition 9.2.2], $\mathcal{P}$ is tight. When Ξ is a compact set, the tightness of $\mathcal{P}$ holds trivially; see similar discussions by Sun and Xu in [33, Remark 3].

(ii) $\mathcal{P}$ is closed if there exists positive constant ϵ satisfying

$$\sup_{P\in\mathcal{P}}\int_{\xi\in\Xi}\|\xi-\mu\|^{1+\epsilon}P(d\xi) < \infty.$$

Indeed, let $\{P_k\}\subset\mathcal{P}$ and P_k converges to P weakly. By Lemma 2.1,

$$0 = \lim_{k\to\infty}\mathbb{E}_{P_k}[\xi-\mu] = \mathbb{E}_P[\xi-\mu],$$

$$d \geq \lim_{k\to\infty}\mathbb{E}_{P_k}[|\xi-\mu|] = \mathbb{E}_P[|\xi-\mu|],$$

which shows that $P\in\mathcal{P}$. When Ξ is a compact set, the closeness of $\mathcal{P}$ holds trivially.

(iii) Note that both $\mathcal{P}$ and $\mathcal{P}_N$ may constitute discrete probability measures, thus condition (3.16) is not guaranteed here. However, due to specific structure of $\mathcal{P}$ and $\mathcal{P}_N$, we are able to show $\mathbb{H}_{TV}(\mathcal{P}_N,\mathcal{P})\to 0$, i.e., Assumption 4.1 holds. This is because the set $\mathscr{G}$ defined as in (2.3) is bounded, i.e., $\sup_{g\in\mathscr{G}}\|g\|\leq 1$, hence $\mathscr{D}(P,Q)\leq\mathbb{D}_{TV}(P,Q)$. By Theorem 4.2, $v_N(\cdot)$ converges to $v(\cdot)$ uniformly over X as N increases. However, we are short of claiming continuity of $v_N(\cdot)$ or $v(\cdot)$, and hence we are unable to apply our main stability result Theorem 4.3 at this point. We will come back to this in Remark 5.2.

5.2.1 Tractability with robust liner chance constraints

To illustrate how MPDRCC$_N$ may be solved efficiently, we consider a special case where $g(x,\xi)$ is linear in ξ, namely

$$\begin{array}{ll}\min\limits_{x\in\mathbb{R}^n} & f(x)\\ \text{s.t.} & x\in X,\\ & \inf\limits_{P\in\mathcal{P}_N} P(A(x)\xi\geq b(x))\geq 1-\beta,\end{array}\tag{5.53}$$

where $A(x) := [A_1x-b_1, A_2x-b_2,\ldots,A_kx-b_k]\in\mathbb{R}^{m\times k}$, $b(x) := b_0 - A_0x\in\mathbb{R}^m$, $A_i\in\mathbb{R}^{m\times n}$, $b_i\in\mathbb{R}^m$, for $i=0,1,\ldots,k$. Here we assume $\Xi=\mathbb{R}^k$.

Proposition 5.3 *Let $x\in X$ be fixed and the ambiguity set be defined as in (5.51). Let*

$$v_N(x) := \inf_{P\in\mathcal{P}_N} P(A(x)\xi\geq b(x)).$$

Then $v_N(x)$ is the optimal value function of the following problem:

$$\begin{array}{ll} \sup\limits_{\alpha,\rho,\gamma,\eta} & \alpha + \mu_N^T\rho - d_N^T\gamma \\ \text{s.t.} & \alpha + \mu_N^T\rho \le 1, \\ & -\gamma \le \rho \le \gamma, \\ & \alpha + \mu_N^T(\rho - A(x)^T\eta) + b(x)^T\eta \le 0, \\ & -\gamma \le \rho - A(x)^T\eta \le \gamma, \\ & \alpha \in \mathbb{R}, \rho \in \mathbb{R}^k, \gamma \in \mathbb{R}^k_+, \eta \in \mathbb{R}^m_+. \end{array} \tag{5.54}$$

Proof. Let $\mathscr{M}_+$ denote the positive linear space of all signed measures generated by $\mathscr{P}$, let

$$\langle P, h(\xi)\rangle := \int_\Xi h(\xi)P(d\xi).$$

By the definition of $\mathcal{P}_N$ in (5.51), $v_N(x)$ can be written as

$$\begin{array}{lll} v_N(x) = & \inf\limits_{P\in\mathscr{M}_+} & \langle P, \mathbb{I}_{A(x)\xi\ge b(x)}(\xi)\rangle \\ & \text{s.t.} & \langle P, \xi\rangle = \mu_N, \\ & & \langle P, |\xi - \mu_N|\rangle \le d_N, \\ & & \langle P, 1\rangle = 1. \end{array}$$

Since $\Xi = \mathbb{R}^k$, then $\{\mathbb{E}_P[\xi], P \in \mathscr{P}\} = \mathbb{R}^k$, and hence there exists $P_0 \in \mathscr{P}$ such that $\mathbb{E}_{P_0}[\xi] = \mu_N$, $\mathbb{E}_{P_0}[|\xi - \mu_N|] < d_N$, i.e., the strong duality holds (see [38, Example 2.1]) and [31, Proposition 3.4]). The Lagrange dual problem is

$$\begin{array}{ll} \sup\limits_{\gamma\ge 0,\alpha,\rho} & \alpha + \mu_N^T\rho - d_N^T\gamma \\ \text{s.t.} & \alpha + \xi^T\rho - |\xi - \mu_N|^T\gamma \le \mathbb{I}_{A(x)\xi\ge b(x)}(\xi), \forall \xi \in \Xi. \end{array} \tag{5.55}$$

The constraint of (5.55) is equivalent to

$$\left\{ \begin{array}{l} \alpha + \xi^T\rho - |\xi - \mu_N|^T\gamma \le 1, \forall \xi \in \Xi, \\ \alpha + \xi^T\rho - |\xi - \mu_N|^T\gamma \le 0, \forall \xi \in \Xi \text{ such that } A(x)\xi < b(x). \end{array} \right. \tag{5.56}$$

The first constraint in (5.56) means the optimal value of problem

$$\begin{array}{ll} \sup\limits_{\xi,\theta} & \alpha + \xi^T\rho - \theta^T\gamma \\ \text{s.t.} & \xi - \mu_N \le \theta, \\ & -\xi + \mu_N \le \theta, \end{array}$$

is upper bounded by 1. Through Lagrange duality of the above problem, the constraint is equivalent to

$$\alpha + \mu_N^T\rho \le 1,\ -\gamma \le \rho \le \gamma. \tag{5.57}$$

Likewise, the second constraint in (5.56) holds if there exists $\eta \in \mathbb{R}^m_+$ satisfying

$$\alpha + \mu_N^T(\rho - A(x)^T\eta) + b(x)^T\eta \le 0,\ -\gamma \le \rho - A(x)^T\eta \le \gamma. \tag{5.58}$$

Combining (5.55), (5.57) and (5.58), we obtain (5.54). ■

Remark 5.2 Analogous to formulation (5.54), we can derive a dual formulation of $v(x)$ as

$$\begin{array}{ll} \sup\limits_{\alpha,\rho,\gamma,\eta} & \alpha + \mu^T\rho - d^T\gamma \\ \text{s.t.} & \alpha + \mu^T\rho \le 1, \\ & -\gamma \le \rho \le \gamma, \\ & \alpha + \mu^T(\rho - A(x)^T\eta) + b(x)^T\eta \le 0, \\ & -\gamma \le \rho - A(x)^T\eta \le \gamma, \\ & \alpha \in \mathbb{R}, \rho \in \mathbb{R}^k, \gamma \in \mathbb{R}^k_+, \eta \in \mathbb{R}^m_+. \end{array} \tag{5.59}$$

If for any $x \in X$, the system

$$A(x)^T\eta = 0, \quad b(x)^T\eta = 0$$

has a unique solution 0, then the set of feasible solutions of (5.59) is compact. Moreover, since $A(\cdot)$ and $b(\cdot)$ are globally Lipschitz continuous on X, then by [40, Lemma 2.1], we can show the set of feasible solutions is Lipschitz continuous on X provided that (5.59) satisfies Slater constraint qualification for every $x \in X$. By [20, Theorem 1], the optimal value function $v(x)$ is continuous. Further to Remark 5.1 (iii), we can apply Theorem 4.3 to establish convergence of the optimal value and the optimal solutions of (2.10) to their true counterpart of (1.2) in this particular setting.

It might also be interesting to mention numerical tractability of problem (5.54). By [14, Theorem 13], if $A(x)$ depends on x, the program is strongly NP-hard. Let $A_i = 0$ for $i = 1, \ldots, k$, $B := [b_1, b_2, \ldots, b_k] \in \mathbb{R}^{m\times k}$. Then $v_N(x)$ can be obtained by solving a linear second order cone programming:

$$\begin{array}{lll} v_N(x) = & \sup\limits_{\rho,\gamma,\eta} & 1 - d_N^T\gamma \\ & \text{s.t.} & -\gamma \le \rho \le \gamma, \\ & & \|(2, \eta_i - (A_0x - b_0 - B\mu_N)_i)\| \le \eta_i + (A_0x - b_0 - B\mu_N)_i, \ i = 1, \ldots, m, \\ & & -\gamma \le \rho + B^T\eta \le \gamma, \\ & & \rho \in \mathbb{R}^k, \gamma \in \mathbb{R}^k_+, \eta \in \mathbb{R}^m_+. \end{array}$$

We omit the details as this is not the main focus of this paper.

5.3 Approximation of the mean-variance ambiguity set and feasibility of the robust chance constraint

We now move on to discuss the case when the ambiguity set $\mathcal{P}$ is defined through mean and variance,

$$\mathcal{P} := \{P \in \mathscr{P} : \mathbb{E}_P[\xi] = \mu, \ \mathbb{E}_P[(\xi - \mu)(\xi - \mu)^T] = \Sigma\}, \tag{5.60}$$

and its approximation defined through samples is

$$\mathcal{P}_N := \{P \in \mathscr{P} : \mathbb{E}_P[\xi] = \mu_N, \ \mathbb{E}_P[(\xi - \mu_N)(\xi - \mu_N)^T] = \Sigma_N\}. \tag{5.61}$$

Similar to the proof of Proposition 5.2, we can derive the approximation of $\mathcal{P}_N$ to $\mathcal{P}$ in this case.

Proposition 5.4 *If $\Xi = \mathbb{R}^k$, then $\lim_{N\to\infty} \mathscr{H}(\mathcal{P}_N, \mathcal{P}) \to 0$.*

Proof. Since $\Xi = \mathbb{R}^k$, it follows from [38, Proposition 2.1], the set $\{\mathbb{E}_P[\xi], P \in \mathscr{P}\} = \mathbb{R}^k$ and $\{\mathbb{E}_P[(\xi-\mu_N)(\xi-\mu_N)^T], P \in \mathscr{P}\} = S^k_+$. The rest of the proof is analogous to that of Proposition 5.2 and followed by Remark 5.1 (iii), we omit the details. ■

Note that the equality constraint on the covariance in (5.60) may be changed to an inequality. Consequently

$$\mathcal{P} := \{P \in \mathscr{P} : \mathbb{E}_P[\xi] = \mu,\ \mathbb{E}_P[(\xi - \mu)(\xi - \mu)^T] \preceq \Sigma\}. \tag{5.62}$$

The change reflects practical need where there is an ambiguity in the true covariance. The resulting $\mathcal{P}$ is known as *Chebyshev ambiguity set*; see [14]. Let us define the approximation of $\mathcal{P}$ through samples

$$\mathcal{P}_N := \{P \in \mathscr{P} : \mathbb{E}_P[\xi] = \mu_N,\ \mathbb{E}_P[(\xi - \mu_N)(\xi - \mu_N)^T] \preceq \Sigma_N\}. \tag{5.63}$$

Similar to Proposition 5.4, we can establish convergence of $\mathcal{P}_N$ to $\mathcal{P}$ under the pseudo metric.

Proposition 5.5 *If $\Xi = \mathbb{R}^k$, then $\lim_{N\to\infty} \mathscr{H}(\mathcal{P}_N, \mathcal{P}) \to 0$.*

Following Remark 5.2, we can also derive continuity of the robust probability function $v(x)$ through dual formulation under some moderate conditions and subsequently apply main stability results Theorem 4.3 in this setting. We leave this for interested readers to verify.

5.3.1 Feasibility of the robust constraint (1.2)

Deviating slightly from our main topic on approximation of $\mathcal{P}$, here we touch a little feasibility issue of the robust constraint (1.2) with the ambiguity set being constructed as in (5.60) because this a relevant issue to the stability analysis. We do so by considering a special case, that is, program (5.53) with $\Xi = \mathbb{R}$, $m = 1$ and the ambiguity set being written as

$$\mathcal{P} := \{P \in \mathscr{P} : \mathbb{E}_P[\xi] = 0,\ \mathbb{E}_P[\xi^2] = \sigma^2\}. \tag{5.64}$$

Observe that

$$\inf_{P\in\mathcal{P}} P(A(x)\xi \geq b(x)) \geq 1 - \beta \Longleftrightarrow \sup_{P\in\mathcal{P}} P(A(x)\xi < b(x)) \leq \beta.$$

For any fixed $x \in X$, if mean value $0 \in \{\xi : A(x)\xi < b(x)\}$, it is easy to verify that $\sup_{P\in\mathcal{P}} P(A(x)\xi < b(x)) = 1$. To see this, let P_N be a discrete probability measure satisfying

$$P_N\left(\xi = \frac{\sqrt{2N}\sigma}{2}\right) = P_N\left(\xi = \frac{-\sqrt{2N}\sigma}{2}\right) = \frac{1}{N},\ P_N(\xi = 0) = 1 - \frac{2}{N},$$

where N is a positive number greater than 2. It is easy to show that

$$P_N \in \mathcal{P} \ \text{ and } \sup_N P_N(\xi = 0) = 1.$$

Note that when $b(x) = b_0 > 0$,

$$0 \in \{\xi : A(x)\xi < b_0\},\ \forall x \in X.$$

Therefore

$$\sup_{P\in\mathcal{P}} P(A(x)\xi < b_0) = 1,\ \forall x \in X,$$

which means the robust constraint is infeasible. Based on the above analysis, we conclude that a necessary condition for a point $x \in X$ to satisfy the robust chance constraint is

$$0 \notin \{\xi : A(x)\xi < b(x)\}$$

when P is defined as (5.64).

Next, we provide a sufficient condition for the feasibility of the robust chance constraint. Suppose that x satisfies $A(x) < 0$, $b(x) < 0$ and $\frac{b(x)}{A(x)} \geq \sigma\sqrt{1/\beta - 1}$ with σ being defined as in (5.64). Then

$$\sup_{P\in\mathcal{P}} P(A(x)\xi < b(x)) = \sup_{P\in\mathcal{P}} P\left(\xi > \frac{b(x)}{A(x)}\right) \leq \beta.$$

To see this, we note that by [39, Lemma 2],

$$\sup_{P\in\mathcal{P}} P\left(\xi > \frac{b(x)}{A(x)}\right) = \frac{1}{1 + (\frac{b(x)}{A(x)})^2/\sigma^2} \leq \frac{1}{1 + 1/\beta - 1} = \beta.$$

Likewise, if x satisfies $A(x) > 0$, $b(x) < 0$ and $\frac{b(x)}{A(x)} \leq -\sigma\sqrt{1/\beta - 1}$, then

$$\sup_{P\in\mathcal{P}} P(A(x)\xi < b(x)) = \sup_{P\in\mathcal{P}} P\left(\xi < \frac{b(x)}{A(x)}\right) \leq \beta.$$

5.4 Approximation of density-based ambiguity set

Finally, we discuss the ambiguity set to be constructed through KL-divergence. Let f_0 and f denote the true density function and its perturbation respectively. KL-divergence measures deviation of f from f_0, namely

$$\mathbb{D}_{KL}(f\|f_0) = \int_{R^k} f(\xi)\log\left(\frac{f(\xi)}{f_0(\xi)}\right) d\xi.$$

KL-divergence is introduced by Kullback and Leibler [21]. In practice, the true probability distribution may be unknown and therefore it is often to use a nominal distribution constructed from empirical data to approximate the true distribution. Unfortunately this kind of framework cannot be applied here directly because the density of empirical distribution is atomic hence it is not absolutely continuous w.r.t. Lebesgue probability measure. To get around the hurdle, we propose to estimate f_0 by so-called kernel density estimator (KDE) [30].

Let $\xi_1, \ldots, \xi_N$ be independent and identically distributed sample of ξ, h_N be a sequence of positive constants converging to zero, and $\Phi(\cdot)$ be a measurable kernel function satisfying $\Phi(\cdot) \geq 0, \int \Phi(\xi)d\xi = 1$. The KDE is defined as

$$f_N(\xi) = \frac{1}{Nh_N^k}\sum_{i=1}^{N} \Phi\left(\frac{\xi - \xi_i}{h_N}\right). \tag{5.65}$$

A simple example for $\Phi(\cdot)$ is the standard normal density function.

Lemma 5.1 *[9, Theorem 1] Let* $\Phi(\cdot)$ *be a measurable kernel function satisfying* $\Phi(\cdot) \geq 0$, $\int \Phi(\xi)d\xi = 1$. *Suppose that* $\{h_N\}$ *satisfies*

$$\lim_{N\to\infty} h_N = 0, \lim_{N\to\infty} Nh_N^k = \infty. \tag{5.66}$$

Then

$$\lim_{N\to\infty} \int_{\mathbb{R}^k} \left|f_N(\xi) - f_0(\xi)\right| d\xi = 0 \text{ w.p.1.}$$

Let d be a given positive constant. The ambiguity sets based on KL-divergence are defined by

$$\mathcal{P} = \{P \in \mathscr{P} : D_{KL}(f\|f_0) \leq d, f = dP/d\xi\}, \tag{5.67}$$

and

$$\mathcal{P}_N = \{P \in \mathscr{P} : D_{KL}(f\|f_N) \leq d, f = dP/d\xi\}. \tag{5.68}$$

It has been shown in [18, 19] that the robust chance constraint with the ambiguity defined in (5.67) or (5.68) is equivalent to a classical chance constraint with a perturbed confidence level.

Proposition 5.6 *The robust chance constraint*

$$\inf_{D_{KL}(f\|f_0)\leq d} P(H(x)) \geq 1 - \beta$$

can be reformulated as $P_0(H(x)) \geq 1 - \beta'$, *where*

$$1 - \beta' = \inf_{x\in(0,1)} \left\{ \frac{e^{-d}x^{1-\beta} - 1}{x - 1} \right\},$$

and P_0 *and* P_N *correspond to* f_0 *and* f_N.

By Proposition 5.6, the problems (MPRCC) and (MPRCC$_N$) can be equivalently written as

$$\begin{array}{ll} \min\limits_{x} & f(x) \\ \text{s.t.} & x \in X, \\ & P_0(H(x)) \geq 1 - \beta', \end{array} \tag{5.69}$$

and

$$\begin{array}{ll} \min\limits_{x} & f(x) \\ \text{s.t.} & x \in X, \\ & P_N(H(x)) \geq 1 - \beta'. \end{array} \tag{5.70}$$

The reformulations enable us to carry out stability of (1.2) through (5.69) and (5.70). Therefore it suffices to establish convergence of P_N to P under pseudo metric through Theorem 4.1.

Proposition 5.7 *Let* f_N *be defined as in (5.65). If the true density function* f_0 *is continuous and* $\{h_N\}$ *satisfies condition (5.66), then* $\mathscr{H}(P, P_N) \to 0$.

Proof. It suffices to verify conditions of Theorem 4.1. By Lemma 5.1, P_N converges to P weakly. Moreover, the continuity of f_0 means that the corresponding probability measure P_0 is absolutely continuous w.r.t. Lebesgue measure, which in turn ensures condition (3.16). Thus, all of the conditions of Theorem 4.1 are fulfilled. ■

Remark 5.3 When KL-divergence in (5.67) or (5.68) is replaced by ϕ-divergence measure [2], Jiang and Guan establish similar results to Proposition 5.6 in [19, Theorem 1]. This implies that when the ambiguity set is defined through the ϕ-divergence, the robust chance constraint can be reformulated as an ordinary chance constraint with revised confidence level. If the true density function is continuous and its estimation is defined as in (5.65), then we can apply Theorem 4.1 and Theorem 4.3 to programs (5.69) and (5.70).

References

[1] K. B. Athreya and S. N. Lahiri, *Measure theory and probability theory*, Springer texts in statistics, Springer, New York, 2006.

[2] A. Ben-Tal, D. den Hertog, A. De Waegenaere, B. Melenberg and G. Rennen, Robust solutions of optimization problems affected by uncertain probabilities, *Manage. Sci.*, 59: 341357, 2013.

[3] D. Bertsimas, V. Gupta and N. Kallus, Robust SAA, Preprint, arXiv:1408.4445, 2014.

[4] P. Billingsley, *Convergence of probability measures*, Wiley, New York, 1999.

[5] G. Calafiore and L. El Ghaoui, Distributionally robust chance-constrained linear programs with applications, *J. Optim. Theory Appl.*, 130: 1-22, 2006.

[6] A. Charnes, W. W. Cooper and G. H. Symonds, Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil, *Manag. Sci.*, 4: 235-263, 1958.

[7] W. Chen, M. Sim, J. Sun and C. P. Teo, From CVaR to uncertainty set: implications in joint chance-constrained optimization, *Oper. Res.*, 58: 470-485, 2010.

[8] E. Delage and Y. Ye, Distributionally robust optimization under moment uncertainty with application to data-driven problems, *Oper. Res.*, 58: 592-612, 2010.

[9] L. Devroye and L. Györfi, *Nonparametric density estimation: The l_1 View*, John Wiley & Sons Inc, 1985.

[10] E. Erdoğan and G. Iyengar, Ambiguous chance constrained problems and robust optimization, *Math. Program.*, 107: 37-61, 2006.

[11] P. M. Esfahani and D. Kuhn, Data-driven distributionally robust optimization using the wassersttein metric: performance guarantees and tractable reformulations, *Optim. Online*, 2015.

[12] A. L. Gibbs and F. E. Su, On choosing and bounding probability metrics, *Internat. Statist. Rev.*, 70: 419-435, 2002.

[13] V. Gupta, Near-Optimal Ambiguity sets for Distributionally Robust Optimization, *optimization-online*, July 2015.

[14] G. A. Hanasusanto, V. Roitch, D. Kuhn and W. Wiesemann, A distributionally robust perspective on uncertainty quanlification and chance constrained programming, *Math. Program.*, 2015, DOI 10.1007/s10107-015-0896-z.

[15] R. Henrion and W. Römisch, Metric regularity and quantitative stability in stochastic programs with probabilistic constraints, *Math. Program.*, 100: 55-88, 1999.

[16] R. Henrion and W. Römisch, Hölder and Lipschitz stability of solution sets in programs with probabilistic constraints, *Math. Program.*, 84: 589-611, 2004.

[17] C. Hess, Conditional expectation and marginals of random sets, *Pattern Recognition*, 32: 1543-1567, 1999.

[18] Z. Hu and L. J. Hong, Kullback-Leibler divergence constrained distributionally robust optimization, *Optim. Online*, 2013.

[19] R. Jiang and Y. Guan, Data-driven chance constrained stochastic program, *Math. Program.*, 2015, DOI 10.1007/s10107-015-0929-7.

[20] D. Klatte, A note on quantitative stability results in nonlinear optimization, *Seminarbericht Nr. 90*, Sektion Mathematik, Humboldt-Universität zu Berlin, Berlin, 77-86, 1987.

[21] S. Kullback and R. Leibler, On information and sufficiency, *Ann. Math. Statist.*, 79-86, 1951.

[22] Y. Liu and H. Xu, Entropic approximation for mathematical programs with robust equilibrium constraints, *SIAM J. Optim.*, 24: 933-958, 2014.

[23] L. B. Miller and H. Wagner, Chance-constrained programming with joint constraints, *Oper. Res.*, 13: 930-945, 1965.

[24] A. Nemirovski and A. Shapiro, Convex approximations of chance constrained programs, *SIAM J. Optim.*, 17: 969-996, 2006.

[25] B. K. Pagnoncelli, S. Ahmed and A. Shapiro, Sample average approximation method for chance constrained programming: theory and applications, *J. Optim. Theory Appl.*, 142: 399-416, 2009.

[26] A. Prékopa, On probabilistic constrained programming, In: Proceedings of the Princeto rinceton, 1970.

[27] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer-Verlag, Berlin, 1998.

[28] W. Römisch, Stability of stochastic programming problems, in *Stochastic Programming*, A. Ruszczyński and A. Shapiro, eds., Elsevier, Amsterdam, 483-554, 2003.

[29] W. Römisch, R. Schultz, Stability analysis for stochastic programs, *Ann. Oper. Res.*, 30: 241-266, 1991.

[30] M. Rosenblatt, Remarks on some nonparametric estimates of a density function, *Ann. Math. Stat.*, 27: 832-837, 1956.

[31] A. Shapiro, *On duality theory of conic linear problems*, Miguel A. Goberna and Marco A. López, Eds., Semi-Infinite Programming: Recent Advances, 135-165, 2001.

[32] A. Shapiro and S. Ahmed, On a class of minimax stochastic programs, *SIAM J. Optim.*, 14: 1237-1249, 2004.

[33] H. Sun and H. Xu, Convergence analysis for distributionally robust optimization and equilibrium problems, *Math. Oper. Res.*, 2015, DOI 10.1287/moor.2015.0732.

[34] Z. Wang, P. W. Glynn and Y. Ye, Likelihood robust optimization for data-driven Newsvendor problems, Preprint, arXiv: 1307.6279, 2014.

[35] W. Wiesemann, D. Kuhn and M. Sim, Distributionally robust convex optimization, *Oper. Res.*, 62: 1358-1376, 2014.

[36] J. Wolfowitz, Generalization of the theorem of glivenko-cantelli, *Ann. Math. Statist.*, 131-138, 1954.

[37] H. Xu, Uniform exponential convergence of sample average random functions under general sampling with applications in stochastic programming, *J. Math. Anal. Appl.*, 368: 692-710, 2010.

[38] H. Xu, Y. Liu and H. Sun, Distributionally robust optimization with matrix moment constraints: lagrange duality and cutting plane methods, *Optim. Online*, 2015.

[39] W. Yang and H. Xu, Distributionally robust chance constraints for non-Linear uncertainties, *Math. Program.*, 2014, DOI 10.1007/s10107-014-0842-5.

[40] J. Zhang, H. Xu and L. Zhang, Quantitative Stability Analysis of Stochastic Quasi-Variational Inequality Problems and Application, *Optim. Online*, 2015.

[41] C. Zhao and Y. Guan, Data-driven risk-averse stochastic optimization with wasserstein metric, *Optim. Online*, 2015.

[42] S. Zymler, D. Kuhn and B. Rustem, Distributionally robust joint chance constraints with second-order moment information, *Math. Program.*, 137: 167-198, 2013.