

CONVERGENCE ANALYSIS FOR MATHEMATICAL PROGRAMS WITH DISTRIBUTIONALLY ROBUST CHANCE CONSTRAINT *

SHAOYAN GUO[†], HUIFU XU[‡], AND LIWEI ZHANG[§]

Abstract. Convergence analysis for optimization problems with chance constraints concerns impact of variation of probability measure in the chance constraints on the optimal value and the optimal solutions and research on this topic has been well documented in the literature of stochastic programming. In this paper, we extend such analysis to optimization problems with distributionally robust chance constraints where the true probability distribution is unknown, but it is possible to construct an ambiguity set of probability distributions and the chance constraint is based on the most conservative selection of probability distribution from the ambiguity set. The convergence analysis focuses on impact of the variation of the ambiguity set on the optimal value and the optimal solutions. We start by deriving general convergence results under abstract conditions such as continuity of the robust probability function and uniform convergence of the robust probability functions and followed with detailed analysis of these conditions. Two sufficient conditions have been derived with one applicable to both continuous and discrete probability distribution and the other to continuous distribution. Case studies are carried out for ambiguity sets being constructed through moments and samples.

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

1. Introduction. Consider the following mathematical programs with chance constraint:

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

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, $\xi : \Omega \rightarrow \Xi$ is a vector of random variables defined on a probability space (Ω, \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 .

MPCC has wide applications in engineering design, supply chain management, production planing, water management, where chance constraints may be used to describe likelihood of financial loss, power balance or system control; see [41]. Initiated by Charnes and Cooper [11] and pioneered by Prékopa [33], MPCC has witnessed significant progress in recent years both from a theoretical viewpoint and an algorithmic perspective. For example, Henrion and Strugarek [24], and Van Ackooij [2] study convexity of chance constraints which is closely related to solvability of MPCC. Van Ackooij and Henrion [3] propose an efficient method to compute the gradient and value of probability functions under multivariate Gaussian distribution so that existing NLP solvers can be applied. There are also various approximation methods proposed for solving MPCCs such as sample approximation [28], constrained bundle method and other methods with regularization [13], convex approximation [30] and sample average approximation (SAA) [31], etc. For a comprehensive overview of MPCC, we refer interested readers to [34, 35, 41] and references therein.

The literature on MPCC often assume full knowledge of the underlying probability distribution. However, in many practical applications, the true probability distribution may be unknown but it is possible to obtain some partial information such as prior moments of some reference functions of the random variables, and/or their samples and use them to construct an ambiguity set of probability distributions which contain or approximate the true probability distribution. This motivates one to consider mathematical programs with so-called distributionally robust chance constraint:

$$\text{(MPDRCC)} \quad \begin{array}{ll} \min_{x \in X} & f(x) \\ \text{s.t.} & \inf_{P \in \mathcal{P}} P(g(x, \xi) \leq 0) \geq 1 - \beta, \end{array} \quad (1.2)$$

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

[†] School of Mathematics, University of Southampton, Southampton, SO17 1BJ, UK (sg2y15@soton.ac.uk) and School of Mathematical Sciences, Dalian University of Technology, Dalian, 116024, China.

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

[§] School of Mathematical Sciences, Dalian University of Technology, Dalian, 116024, China (lwzhang@dlut.edu.cn). The research of this author is supported in part by NSFC grants 91330206, 11571059.

where \mathcal{P} is an ambiguity set. Note that if we consider (Ξ, \mathcal{B}) as a measurable space equipped with Borel sigma algebra \mathcal{B} , then \mathcal{P} may be viewed as a set of probability measures defined on (Ξ, \mathcal{B}) induced by the random variate ξ . We will use terminologies probability measure and probability distribution interchangeably throughout the paper.

In comparison with MPCC, MPDRCC requires each feasible solution of problem (1.2) to satisfy the chance constraint $P(g(x, \xi) \leq 0) \geq 1 - \beta$ for every $P \in \mathcal{P}$ no matter whether P is the true probability or not. Thus the feasible set of MPDRCC is smaller than (contained in) that of MPCC and hence the optimal value of the former is larger than that of the latter which means the minimum cost (if we interpret $f(x)$ as a cost) resulting from MPDRCC will be higher. But this is the price we have to pay for the lack of complete information of the true probability distribution because in this model we are not allowed to take a decision which can potentially violate the chance constraint. To explain why an optimal solution arising from (1.2) is sensible, let x^* denote an optimal solution and P^* the corresponding optimal probability distribution attaining the infimum (not necessarily attainable if \mathcal{P} is not weakly compact). Let \hat{P} denote the true unknown probability distribution. Then $P^*(g(x^*, \xi) \leq 0) \geq 1 - \beta$ and

$$\hat{P}(g(x^*, \xi) \leq 0) \geq P^*(g(x^*, \xi) \leq 0) \geq 1 - \beta \quad (1.3)$$

so long as $\hat{P} \in \mathcal{P}$. This shows x^* is feasible to MPCC and moreover, if the first inequality in (1.3) is strict, then it means the distributionally robust chance constraint is more conservative than necessary.

MPDRCC is an important topic in distributionally robust optimization, a number of papers have appeared on this topic. For instance, Calafiore and El Ghaoui [10] show that MPDRCC can be converted into a tractable mathematical program with second-order cone constraint provided that $g(\cdot, \cdot)$ is bilinear and the ambiguity set is characterized by the known mean and variance. Without the linear property of g , Zymler et al [49] prove that CVaR approximation of individual robust chance constraint becomes exact if the constraint functions are either concave or (possibly nonconcave) quadratic in ξ and the first and second order moments, as well as the support set of ξ are known. The quality of this approximation depends on certain scaling parameters when they are confronted with joint chance constraints. Yang and Xu [46] establish that if g is concave in x , and quasi-convex in ξ and the ambiguity set is characterized by its mean and variance, then MPDRCC is also tractable. Erdoğan and Iyengar [16] show that if the ambiguity set contains distributions that are within a certain distance of a nominal distribution in terms of the Prohorov metric, then MPDRCC can be well approximated by a set of sample-based robust optimization constraints. Jiang and Guan [26] consider a family of density-based confidence sets based on general ϕ -divergences, and derive an equivalent reformulation for MPDRCC and show that it is equivalent to a classical chance constraint with a perturbed risk level. Similar results have been proved in [25]. More recently, Hanasusanto et al [20] investigate tractability of a richer class of ambiguity sets defined through moment conditions and structural information such as symmetry, unimodality, and independence patterns for MPDRCC.

One of the main issues in MPDRCC is construction of the ambiguity set \mathcal{P} . 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 [10, 12, 17, 26, 48] propose data-driven distributionally robust optimization that the ambiguity set is defined through samples, however, they mainly investigate tractability and the finite sample guarantee of the resulting reformulation. More recently, Bertsimas et al [6] propose a modification of SAA, termed robust SAA to the data-driven settings. Using the goodness-of-fit (GoF) hypothesis test, they discuss the finite sample guarantee 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 [6] 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 a theoretical perspective, the analysis may be viewed as convergence 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 convergence analysis of distributionally robust formulation of a one stage stochastic program by Sun and Xu [42] where the impact of the optimal value and the 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.

- **Convergence 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 converges to \mathcal{P} in some pseudo-metric (Theorem 3.1). We also investigate convergence of the optimal value and the optimal solutions of MPDRCC with \mathcal{P}_N replaced by \mathcal{P} under some general conditions (Theorem 3.2). The convergence results extend similar results in stochastic programming [37] to MPDRCC.
- **Continuity of the robust probability function.** We present a new alternative condition ((C2) in Condition 4.1) for the continuity of probability function (Theorem 4.1) and discuss in detail the relationship between the new condition and the well-known condition in the literature of stochastic programming ((C1) in Condition 4.1) through a proposition (Proposition 4.1) and an example (Example 4.1). Moreover, under (C1) or (C2), we derive pointwise continuity of the robust probability function (Theorem 4.2) and give a simple example to illustrate (Example 4.2).
- **Convergence of approximate ambiguity set under the pseudo-metric.** Since the uniform convergence of the robust probability function in Theorem 3.1 relies heavily on the convergence of \mathcal{P}_N to \mathcal{P} , we derive sufficient conditions for the latter. Specifically, we show convergence of \mathcal{P}_N to \mathcal{P} under the pseudo-metric when \mathcal{P}_N converges to \mathcal{P} weakly and \mathcal{P} satisfies (C1) or (C2) (Lemma 5.1), the result has independent interest in stochastic programming. With the lemma, we move on to establish semi-convergence of \mathcal{P}_N to \mathcal{P} (Propositions 5.1 and 5.2).
- **Applications of the convergence results.** We present three applications of the established convergence results with different motivations. First (Section 6.1), we consider a case where the ambiguity set \mathcal{P} is defined by some general moment conditions. Our concern is that the resulting MPDRCC may be very difficult to solve and we propose a piecewise approximation scheme. We show that our general convergence results can be applied to this kind of computationally oriented approximation scheme (Theorem 6.1). Second (Section 6.2), we consider a situation where the ambiguity set \mathcal{P} is defined through some specific moment conditions and it is possible to obtain samples of ξ to calculate the empirical average mean and absolute deviation. We demonstrate that our general convergence results in Section 3 can also be applied to this kind of circumstance as sample size increases (Proposition 6.2). Finally (Section 6.3), we give another application where we use empirical data to construct a so-called kernel density estimator and then use Kullback-Leibler (KL) divergence to construct an ambiguity set. We demonstrate how our convergence results can also be effectively applied to that setting (Proposition 6.4).

Throughout the paper, 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 sets $A, B \subset \mathbb{R}^n$, $\text{cl} A$, $\text{int} A$ and $\text{bd} A$ denote respectively the closure, interior and boundary of A ; $A \setminus B$ denotes the set of points which lie in set A but not in set B , $A + B := \{a + b : a \in A, b \in B\}$ denotes Minkowski sum. If A and B are two compact sets, 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 . Finally, we use \mathbf{B} to denote closed unit ball in finite dimensional Hilbert spaces.

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 \bar{x} is the set

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

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

$$\liminf_{x \rightarrow \bar{x}} \Psi(x) := \left\{ y \in \mathcal{Y} : \forall x^k \rightarrow \bar{x}, \exists N \in \mathbb{N}_\infty, y^k \xrightarrow{N} y \text{ with } y^k \in \Psi(x^k) \right\},$$

where $\mathbb{N}_\infty := \{N \subseteq \mathbb{N} : \mathbb{N} \setminus N \text{ is finite}\}$ (\mathbb{N} denotes natural numbers), $y^k \xrightarrow{N} y$ denotes the convergence of a subsequence designed by an index N [36].

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

$$\limsup_{x \rightarrow \bar{x}} \Psi(x) \subset \Psi(\bar{x}),$$

or equivalently, $\limsup_{x \rightarrow \bar{x}} \Psi(x) = \Psi(\bar{x})$. Ψ is said to be *inner semicontinuous* (isc for short) at $\bar{x} \in \mathcal{X}$ if

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

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 [36, Proposition 5.12]). *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

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

2.2. Pseudo-metric, Kolmogorov metric and total variation metric. Let $\mathcal{P}(\Xi)$ denote the set of all probability measures in the space (Ξ, \mathcal{B}) . We need appropriate metrics to characterize convergence of probability measures in $\mathcal{P}(\Xi)$.

For each fixed $x \in X$, let

$$H(x) := \{z \in \Xi : g(x, z) \leq 0\}, \quad (2.1)$$

and

$$\mathbb{1}_{H(x)}(z) := \begin{cases} 1 & \text{for } z \in H(x), \\ 0 & \text{for } z \in \Xi \setminus H(x), \end{cases}$$

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

$$P(\xi \in \Xi : g(x, \xi) \leq 0) = \mathbb{E}_P [\mathbb{1}_{H(x)}(\xi)].$$

We consider the following set of random indicator functions

$$\mathcal{G} := \{\mathbb{1}_{H(x)}(\xi(\cdot)) : x \in X\}. \quad (2.2)$$

For $P, Q \in \mathcal{P}(\Xi)$, let

$$\mathcal{D}(P, Q) := \sup_{g \in \mathcal{G}} |\mathbb{E}_P[g] - \mathbb{E}_Q[g]| = \sup_{x \in X} |P(H(x)) - Q(H(x))|. \quad (2.3)$$

We call $\mathcal{D}(P, Q)$ *pseudo-metric* in that it satisfies all properties of a metric except that $\mathcal{D}(P, Q) = 0$ does not necessarily imply $P = Q$ unless the set of functions \mathcal{G} is sufficiently large. This type of pseudo-metric is widely used for stability analysis in stochastic programming; see similar definitions in a review paper by Römisch [37, Page 529].

Let $P \in \mathcal{P}(\Xi)$ be a probability measure and $\mathcal{A}_i \subset \mathcal{P}(\Xi)$, $i = 1, 2$, 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 $\mathcal{D}(P, \mathcal{A}_1) := \inf_{Q \in \mathcal{A}_1} \mathcal{D}(P, Q)$, the deviation (excess) of \mathcal{A}_1 from (over) \mathcal{A}_2 as

$$\mathcal{D}(\mathcal{A}_1, \mathcal{A}_2) := \sup_{P \in \mathcal{A}_1} \mathcal{D}(P, \mathcal{A}_2) \quad (2.4)$$

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

$$\mathcal{H}(\mathcal{A}_1, \mathcal{A}_2) := \max \{\mathcal{D}(\mathcal{A}_1, \mathcal{A}_2), \mathcal{D}(\mathcal{A}_2, \mathcal{A}_1)\}. \quad (2.5)$$

DEFINITION 2.1 (Kolmogorov metric [18]). *For two probability measures $P, Q \in \mathcal{P}(\Xi)$, the Kolmogorov metric is defined by*

$$\mathbb{D}_K(P, Q) := \sup_{\eta \in \mathbb{R}^k} |F_P(\eta) - F_Q(\eta)|, \quad (2.6)$$

where F_P and F_Q denote the 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 2.2 (Total variation metric [18]). Let $P, Q \in \mathcal{P}(\Xi)$. The total variation metric between P and Q is defined as

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

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

2.3. Weak convergence. Let $\{P_N\} \in \mathcal{P}(\Xi)$ be a sequence of probability measures. Recall that $\{P_N\}$ is said to converge to $P \in \mathcal{P}(\Xi)$ weakly if

$$\lim_{N \rightarrow \infty} \int_{\Xi} h(\xi) P_N(d\xi) = \int_{\Xi} h(\xi) P(d\xi), \quad (2.7)$$

for each bounded and continuous function $h : \Xi \rightarrow \mathbb{R}$; see [7, Theorem 2.1].

For a family of probability measures \mathcal{A} on (Ξ, \mathcal{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. The set \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 $P \in \mathcal{A}$. The set \mathcal{A} is said to be *weakly compact* if every sequence $\{P_N\} \subset \mathcal{A}$ contains a subsequence $\{P_{N'}\}$ and moreover there exists $P \in \mathcal{A}$ such that $P_{N'}$ converges to P weakly; see Billingsley [7] for a similar notion called relative compactness.

By the well-known Prokhorov's theorem (see [4]), a closed set \mathcal{A} (under the weak topology) of probability measures is *weakly compact* if it is tight. In particular, if Ξ is a compact set, then the set of all probability measures on (Ξ, \mathcal{B}) is weakly compact; see [40].

LEMMA 2.1 (Uniform integrability). 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 $h : Z \rightarrow \mathbb{R}$ be a measurable function with $P(D_h) = 0$, where $D_h := \{z \in Z : h \text{ is discontinuous at } z\}$. Then it holds

$$\lim_{N \rightarrow \infty} \int_Z h(\xi) P_N(d\xi) = \int_Z h(\xi) P(d\xi) \quad (2.8)$$

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

$$\lim_{r \rightarrow \infty} \sup_{N \in \mathbb{N}} \int_{\{z \in Z : |h(z)| \geq r\}} |h(\xi)| P_N(d\xi) = 0. \quad (2.9)$$

A sufficient condition for (2.9) is

$$\sup_{N \in \mathbb{N}} \int_Z |h(\xi)|^{1+\epsilon} P_N(d\xi) < \infty \text{ for some } \epsilon > 0. \quad (2.10)$$

Lemma 2.1 is given in [42, Lemma 1]. It is noted that the conclusion is drawn from [7, Theorem 3.5] (or Theorem 5.4 in an earlier version of the book). Here we provide a proof in the appendix in that no proof is given in [42] and there are indeed some subtle arguments behind the claim.

Recall that P_N converges to P weakly if and only if the limit (2.7) holds for all continuous and bounded functions. Lemma 2.1 gives sufficient conditions for the limit to hold for some discontinuous and unbounded functions.

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

$$\begin{aligned} (\text{MPDRCC}_N) \quad & \min_{x \in X} f(x) \\ & \text{s.t.} \quad \inf_{P \in \mathcal{P}_N} P(g(x, \xi) \leq 0) \geq 1 - \beta. \end{aligned} \quad (2.11)$$

Our purpose is to analyse convergence of the optimal value and the optimal solutions of problem (2.11) 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 [22, 23, 37, 38] 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 6.

Note that for the constant β , we require the feasible set of problem (2.11) to be nonempty for each N and to ensure well definedness of the problem. A necessary and sufficient condition is that for each N there exists at least one point $x^N \in X$ such that

$$\sup_{P \in \mathcal{P}_N} P(\Xi \setminus H(x^N)) \leq \beta, \quad (2.12)$$

which means any probability measure in \mathcal{P}_N must not mass outside $H(x^N)$ above level β . A similar comment applies to the true problem (1.2).

For each fixed $x \in X$, let

$$v(x) := \inf_{P \in \mathcal{P}} P(g(x, \xi) \leq 0) \equiv \inf_{P \in \mathcal{P}} P(H(x)), \quad (2.13)$$

and

$$v_N(x) := \inf_{P \in \mathcal{P}_N} P(g(x, \xi) \leq 0) \equiv \inf_{P \in \mathcal{P}_N} P(H(x)). \quad (2.14)$$

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

3. Main convergence results. In this section, we turn to the central theme of this paper, that is, impact of the variation of the ambiguity set 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 convergence 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 approximates \mathcal{P} . To this end we make the following assumptions on \mathcal{P} and \mathcal{P}_N .

ASSUMPTION 3.1 (Approximation of the ambiguity set under the pseudo-metric). *The ambiguity sets \mathcal{P} and \mathcal{P}_N satisfy the following conditions.*

- (a) $\lim_{N \rightarrow \infty} \mathcal{D}(\mathcal{P}_N, \mathcal{P}) = 0$;
- (b) $\lim_{N \rightarrow \infty} \mathcal{D}(\mathcal{P}, \mathcal{P}_N) = 0$.

Part (a) requires that \mathcal{P}_N upper semiconverges to \mathcal{P} under the pseudo-metric whereas part (b) requires that \mathcal{P}_N lower semiconverges to \mathcal{P} . Parts (a) and (b) imply $\mathcal{H}(\mathcal{P}_N, \mathcal{P}) \rightarrow 0$ as $N \rightarrow \infty$. When \mathcal{P}_N and \mathcal{P} have a specific structure, Assumption 3.1 may be verified directly. We will come back to this in Section 5.

Under Assumption 3.1, 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 3.1 (Uniform approximation of the robust probability function). *Under Assumption 3.1, $v_N(x)$ converges to $v(x)$ uniformly over X as N tends to ∞ , that is,*

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

Proof. Let $x \in X$ be fixed. By the definition of the pseudo-metric, we have

$$\begin{aligned} v_N(x) - v(x) &= \inf_{P \in \mathcal{P}_N} P(H(x)) - \inf_{P' \in \mathcal{P}} P'(H(x)) \\ &= \sup_{P' \in \mathcal{P}} \inf_{P \in \mathcal{P}_N} [P(H(x)) - P'(H(x))] \\ &\leq \sup_{P' \in \mathcal{P}} \inf_{P \in \mathcal{P}_N} \sup_{x' \in X} |P'(H(x')) - P(H(x'))| \\ &= \mathcal{D}(\mathcal{P}, \mathcal{P}_N). \end{aligned}$$

Likewise, we can obtain $v(x) - v_N(x) \leq \mathcal{D}(\mathcal{P}_N, \mathcal{P})$, hence $|v_N(x) - v(x)| \leq \mathcal{H}(\mathcal{P}, \mathcal{P}_N)$. The conclusion follows as the right hand side of the inequality above is independent of x and it goes to 0 under Assumption 3.1. \blacksquare

ASSUMPTION 3.2. *We assume that v is continuous over X .*

With the uniform convergence of the robust probability function, we are ready to discuss the convergence of the optimal values and the optimal solutions of problem (2.11). To ease the exposition, let

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

denote the feasible sets of MPDRCC (1.2) and MPDRCC_N (2.11) respectively. We can rewrite problems (1.2) and (2.11) as

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & x \in \mathcal{F}, \end{aligned} \tag{3.1}$$

and

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & x \in \mathcal{F}_N. \end{aligned} \tag{3.2}$$

Let $\vartheta := \inf\{f(x) : x \in \mathcal{F}\}$ denote the optimal value of problem (3.1), and S the corresponding set of optimal solutions, that is, $S := \{x \in \mathcal{F} : \vartheta = f(x)\}$. Likewise, let

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

be defined similarly for problem (3.2). Let \mathcal{F}° denote the set of strict feasible solutions of problem (3.1), i.e.,

$$\mathcal{F}^\circ := \{x \in X : v(x) > 1 - \beta\}. \tag{3.3}$$

The following theorem states the convergence of problem (3.2) to problem (3.1) in terms of the optimal values and the optimal solutions.

THEOREM 3.2 (Convergence analysis of MPDRCC (3.1)). *Suppose: (a) Assumptions 3.1 and 3.2 hold; (b) there exists at least one optimal solution $x^* \in S$ which falls in the closure of \mathcal{F}° , that is,*

$$\text{cl } \mathcal{F}^\circ \cap S \neq \emptyset. \tag{3.4}$$

Then the optimal value and the optimal solutions of problem (3.2) converge to that of problem (3.1), i.e.,

$$\lim_{N \rightarrow \infty} \mathbb{D}(S_N, S) = 0, \quad \lim_{N \rightarrow \infty} \vartheta_N = \vartheta. \tag{3.5}$$

Condition (b) requires problem (3.1) to have at least one optimal solution which is connected to the interior of \mathcal{F} . In the case when S is a singleton, say $\{x^*\}$, the condition requires x^* to be either located in the interior of \mathcal{F} or at the boundary of its closure (the closure of \mathcal{F}°). The latter means that there is a sequence in \mathcal{F}° approximating x^* . It is fulfilled if the feasible set \mathcal{F} is convex or connected. In particular, it holds when the robust chance constraint $v(x) \geq 1 - \beta$ satisfies the Slater constraint qualification. This condition is well adopted for asymptotic convergence in stochastic programming; see [29] and references therein.

Proof of Theorem 3.2. Let us begin by proving convergence of the feasible set \mathcal{F}_N , i.e.,

$$\lim_{N \rightarrow \infty} \mathbb{D}(\mathcal{F}_N, \mathcal{F}) = 0. \tag{3.6}$$

It follows from Theorem 3.1 that $v_N(\cdot)$ converges to $v(\cdot)$ uniformly over X , together with continuity of $v(\cdot)$, we can establish (3.6) by [44, Lemma 4.2(i)].

We now move on to prove the first equation in (3.5). Let $x^N \in S_N$ be an optimal solution to problem (3.2). Since the sequence $\{x^N\}$ is contained in the compact set X , by taking a subsequence if necessary, we may assume for the simplicity of notation that $x^N \rightarrow \bar{x}$, for some $\bar{x} \in X$. The closedness of \mathcal{F} ensures $\bar{x} \in \mathcal{F}$. In what follows, we show that $\bar{x} \in S$. By the continuity of f ,

$$\lim_{N \rightarrow \infty} \vartheta_N = \lim_{N \rightarrow \infty} f(x^N) = f(\bar{x}) \geq \vartheta. \tag{3.7}$$

On the other hand, condition (b) ensures that there exist $x^* \in S$ and $\{y^t\} \subset \mathcal{F}^\circ$ such that $y^t \rightarrow x^*$. Together with the continuity of $f(\cdot)$, this implies that for any fixed $\varepsilon > 0$, we can find $y^t \in \mathcal{F}^\circ$ such that

$$f(y^t) - \vartheta = f(y^t) - f(x^*) \leq \varepsilon.$$

Moreover, since $y^t \in \mathcal{F}^\circ \subset X$ and $v_N(x)$ converges to $v(x)$ uniformly over X , then $y^t \in \mathcal{F}_N$ for N sufficiently large. The optimality of x^N means $f(y^t) \geq f(x^N)$. Therefore,

$$\vartheta \geq f(y^t) - \varepsilon \geq f(x^N) - \varepsilon = \vartheta^N - \varepsilon.$$

By driving N to infinity, we arrive at $\vartheta \geq f(\bar{x}) - \varepsilon$, which implies $\vartheta \geq f(\bar{x})$ in that ε can be chosen arbitrarily small. Together with (3.7), this shows $f(\bar{x}) = \vartheta$, hence $\bar{x} \in S$. Since x^N is arbitrarily selected from S_N , this shows that any cluster point of S_N is contained in S , which is the first limit in (3.5).

The second limit of (3.5) follows straightforwardly from the first limit and the continuity of f . \blacksquare

4. Continuity of robust probability function. The main convergence results in the preceding section are established under two general assumptions: Assumption 3.1 and Assumption 3.2. In this section, we investigate under which conditions Assumption 3.2 is fulfilled, that is, $v(\cdot)$ defined in (2.13) is continuous over X .

Observe that $g(x, z) \leq 0$ is equivalent to $\max_{1 \leq i \leq m} g_i(x, z) \leq 0$. Thus, to ease the exposition, we regard $g(x, z)$ as a scalar function throughout this section.

We divide the investigation into two parts: (a) continuity of probability function $P(H(x))$ and (b) continuity of robust probability function $v(x)$. The former paves the way for the latter.

4.1. Continuity of probability function $P(H(\cdot))$. Discussions on continuity of probability functions are well documented in the literature of stochastic programming; see for instance [1, 21, 31]. Here we revisit the issue by proposing an alternative condition.

CONDITION 4.1. *Let $H(\cdot)$ be defined as in (2.1) and $P \in \mathcal{P}(\Xi)$.*

(C1) $P(\xi \in \Xi : g(x, \xi) = 0) = 0$ for any $x \in X$;

(C2) $H(\cdot)$ is convex-valued and continuous over X and for any $x \in X$,

$$P(\text{bd } H(x)) = 0. \tag{4.1}$$

Here $\text{bd } S$ denotes the boundary of set S in \mathbb{R}^k . Likewise, we will use $\text{int } S$ to denote the interior of S .

(C1) is well known and has been widely used in stochastic programming no matter whether Ξ is discrete or not. (C2) is new as far as we know. Note that in the case when Ξ is a discrete set, (C2) may fail because of the discontinuity of $H(\cdot)$ (here we regard Ξ as a set of \mathbb{R}^k rather than a space itself), or become irrelevant when $\text{int } H(x) = \emptyset$, $P(H(x)) = P(\text{bd } H(x))$ and equality (4.1) implies $v(x) = 0$ and hence x is infeasible to MPDRCC (1.2). Therefore (C2) is meaningful only when Ξ is not a discrete set and $\text{int } H(x) \neq \emptyset$. We include (C2) because there is an interesting and important case where (C1) may fail but (C2) holds (Example 4.1) and we believe (C2) provides an instrumental complement to (C1). Our purpose in this section is to show the continuity of $P(H(\cdot))$ under either of the conditions. Before this, let us discuss the relationship between (C1) and (C2).

Observe that since Ξ is closed and H is closed valued,

$$\text{bd } H(x) \subseteq \{z \in \text{bd } \Xi : g(x, z) \leq 0\} \cup \{z \in \text{int } \Xi : g(x, z) = 0\} \subseteq H(x). \tag{4.2}$$

In the case when $\Xi = \mathbb{R}^k$, $\text{bd } \Xi = \emptyset$ and then (4.2) reduces to

$$\text{bd } H(x) \subseteq \{z \in \Xi : g(x, z) = 0\}. \tag{4.3}$$

Equality holds when

$$\text{int } H(x) = \{z \in \Xi : g(x, z) < 0\}, \tag{4.4}$$

because $H(x) \setminus \text{int } H(x) = \text{bd } H(x)$ and $H(x) \setminus \{z \in \Xi : g(x, z) < 0\} = \{z \in \Xi : g(x, z) = 0\}$. In what follows, we discuss conditions for equality holding in (4.3), continuity of $H(\cdot)$ and equivalence between (C1) and (C2).

PROPOSITION 4.1 (Relationship between (C1) and (C2)). *Assume: (a) Ξ is a convex set, and for each fixed $x \in X$, $g(x, \cdot)$ is convex over Ξ ; (b) for any $x' \in X$, there exists $z' \in \Xi$ such that $g(x', z') < 0$. Then the following assertions hold.*

- (i) $H(\cdot)$ is continuous over X ;
- (ii) the first equality in (4.2) holds;
- (iii) if, in addition, $\Xi = \mathbb{R}^k$, then equality holds in (4.3) and hence (C1) is equivalent to (C2).

Proof. Part (i): Conditions (a) and (b) combined is the well-known Slater constraint qualification. Under that condition, it follows from [36, Example 5.10] or [47, Lemma 2.2] that $H(\cdot)$ is continuous over X .

Parts (ii) and (iii): By [36, Proposition 2.34], the Slater constraint qualification also ensures

$$\text{int } H(x) = \{z \in \text{int } \Xi : g(x, z) < 0\}.$$

Hence the first equality in (4.2) holds. Moreover, when $\Xi = \mathbb{R}^k$, the equality above implies

$$\text{bd } H(x) = \{z \in \Xi : g(x, z) = 0\}.$$

The proof is complete. ■

Proposition 4.1 gives rise to sufficient conditions for the continuity of $H(\cdot)$ and the equivalence of (C1) and (C2). There are some circumstances where (C2) holds whereas (C1) fails. Here is a simple example.

EXAMPLE 4.1. Let $\xi : \Omega \rightarrow \mathbb{R}$ be a random variable defined on a probability space (Ω, \mathcal{F}, P) with support set $\Xi = \mathbb{R}$. Let

$$g(x, z) := \begin{cases} z + x & \text{for } z \geq -x, \\ 0 & \text{for } z \in [-x - 1, -x], \\ z + x + 1 & \text{for } z \leq -x - 1. \end{cases}$$

Observe that $H(x) = \{z \in \mathbb{R} : g(x, z) \leq 0\} = (-\infty, -x]$ and $\{z \in \mathbb{R} : g(x, z) = 0\} = [-x - 1, -x]$. Obviously $H(\cdot)$ is convex-valued and continuous over X and $P(\text{bd } H(x)) = 0$, therefore (C2) is satisfied. On the other hand, $P([-x - 1, -x]) \neq 0$ since P is a continuous probability measure. This shows that (C1) is failed.

Note that all conditions of Proposition 4.1 are satisfied except the convexity of $g(x, \cdot)$, which means the Slater constraint qualification fails. ■

It might be helpful to make some comments on condition (4.1). In probability theory, $H(x)$ is a P-continuity set; see [7]. When $H(x)$ is a convex set, the Lebesgue measure of $\text{bd } H(x)$ is zero; see [15, Lemma 2.4.3]. Thus, if the probability measure P is absolutely continuous w.r.t. the Lebesgue measure on \mathbb{R}^k , then (4.1) holds.

We are now ready to discuss the continuity of $P(H(\cdot))$.

THEOREM 4.1 (Continuity of the probability function). *Let $P \in \mathcal{P}(\Xi)$. Then $P(H(\cdot))$ is continuous over X when either (C1) or (C2) is fulfilled.*

Proof. The continuity is well known under (C1), see for instance [21].

In what follows, we prove the continuity under (C2). We do so by showing that $P(H(\cdot))$ is both upper and lower semicontinuous. Since $H(x)$ is closed for any $x \in X$, $P \rightarrow \{x \in X : P(H(x)) \geq p\}$ has a closed graph for fixed $p \in \mathbb{R}$ [38, Proposition 3.1]. This means that the set $\{x \in X : P(H(x)) \geq p\}$ is closed, hence the upper semicontinuity of $P(H(\cdot))$ holds [36, Page 13]. Thus it suffices to show the lower semicontinuity.

First we claim that for any fixed $\bar{x} \in X$,

$$\liminf_{x \rightarrow \bar{x}} \mathbb{1}_{\text{int } H(x)}(z) \geq \mathbb{1}_{\text{int } H(\bar{x})}(z) \tag{4.5}$$

holds for every $z \in \Xi$. To see this, let us consider two cases.

Case (i): $z \in \Xi \setminus \text{int } H(\bar{x})$. In this case, $\mathbb{1}_{\text{int } H(\bar{x})}(z) = 0$. The inequality (4.5) holds trivially.

Case (ii): $z \in \text{int } H(\bar{x})$. In this case, $\mathbb{1}_{\text{int } H(\bar{x})}(z) = 1$, and there is a positive constant $\delta > 0$ such that $z + 2\delta \mathbf{B} \in H(\bar{x})$, where \mathbf{B} is the unit ball in \mathbb{R}^k . Since $H(\cdot)$ is continuous, then for any $\rho > 0$, there exists a

neighbourhood $\mathcal{N}_{\bar{x}}$ of \bar{x} such that

$$H(\bar{x}) \cap \rho \mathbf{B} \subset H(x) + \delta \mathbf{B}$$

for all $x \in \mathcal{N}_{\bar{x}}$. Let ρ be chosen such that $\rho > \|z\| + 2\delta$. Then for any $x \in \mathcal{N}_{\bar{x}}$,

$$z + 2\delta \mathbf{B} \subset H(\bar{x}) \cap \rho \mathbf{B} \subset H(x) + \delta \mathbf{B}.$$

Since $H(\cdot)$ is convex-valued, by the cancellation law [32, Theorem 3.2.1], we can obtain $z + \delta \mathbf{B} \subset H(x)$ which implies $z \in \text{int } H(x)$ for any $x \in \mathcal{N}_{\bar{x}}$. Hence $\mathbb{1}_{\text{int } H(x)}(z) = 1$ for all $x \in \mathcal{N}_{\bar{x}}$ and (4.5) holds.

By Fatou's lemma, (4.5) implies

$$\liminf_{x \rightarrow \bar{x}} \mathbb{E}_P \left[\mathbb{1}_{\text{int } H(x)}(\xi) \right] \geq \mathbb{E}_P \left[\liminf_{x \rightarrow \bar{x}} \mathbb{1}_{\text{int } H(x)}(\xi) \right] \geq \mathbb{E}_P \left[\mathbb{1}_{\text{int } H(\bar{x})}(\xi) \right]. \quad (4.6)$$

Observe that the difference between $\mathbb{1}_{H(x)}(\cdot)$ and $\mathbb{1}_{\text{int } H(x)}(\cdot)$ occurs only over the set $\text{bd } H(x)$. Since P satisfies $P(\text{bd } H(x)) = 0$ for all $x \in X$, then $\mathbb{E}[\mathbb{1}_{\text{int } H(x)}(\xi)] = P(H(x))$ and $\mathbb{E}[\mathbb{1}_{\text{int } H(\bar{x})}(\xi)] = P(H(\bar{x}))$. Substituting these relations into (4.6), we immediately get

$$\liminf_{x \rightarrow \bar{x}} P(H(x)) \geq P(H(\bar{x})),$$

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

4.2. 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.

LEMMA 4.1. *Suppose $H(\cdot)$ is convex-valued and continuous over X , and $\text{int } H(x) \neq \emptyset$ for any $x \in X$. Then $\text{bd } H(\cdot)$ is outer semicontinuous over X , i.e., for any $\bar{x} \in X$,*

$$\limsup_{x \rightarrow \bar{x}} \text{bd } H(x) \subset \text{bd } H(\bar{x}). \quad (4.7)$$

Proof. The results may be well known in variational analysis, we include a proof as we are unable to identify a reference. Assume for the sake of a contradiction that there exists \bar{u} such that

$$\bar{u} \in \left(\limsup_{x \rightarrow \bar{x}} \text{bd } H(x) \right) \setminus \text{bd } H(\bar{x}).$$

By the definition of \limsup , we can find a sequence $\{x_N\}$ converging to \bar{x} and $u_N \in \text{bd } H(x_N)$ such that $u_N \rightarrow \bar{u} \notin \text{bd } H(\bar{x})$. The later entails $\bar{u} \in \text{int } H(\bar{x})$ due to the fact that H is outer semicontinuous on X and $\bar{u} \notin \text{bd } H(\bar{x})$, and hence there exists a positive constant δ such that $\bar{u} + \delta \mathbf{B} \subset H(\bar{x})$. By the inner semicontinuity of H , for any $\rho > 0$, there exists $N_1 > 0$ such that

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

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

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

By the cancellation law [32], we can obtain $\bar{u} + \frac{3\delta}{4} \mathbf{B} \subset H(x_N)$ for $N \geq N_1$. On the other hand, since u_N converges to \bar{u} , there exists $N_2 \geq N_1$ such that $u_N \in \bar{u} + \frac{\delta}{4} \mathbf{B} \subset \text{int } H(x_N)$ for $N > N_2$. This contradicts the fact that $u_N \in \text{bd } H(x_N)$. ■

REMARK 4.1. It might be helpful to comment on our assumption $\text{int } H(x) \neq \emptyset$ for $x \in X$.

- Observe first that if $\text{int } H(x) = \emptyset$ for $x \in X$, then $H(x) = \text{bd } H(x)$. In that case, (4.7) reduces to outer semicontinuity of $H(\cdot)$ at \bar{x} which is subsumed by the continuity of $H(\cdot)$. Moreover, if the Lebesgue measure of $H(x)$ is nonzero, then the convexity of $H(x)$ implies $\text{int } H(x) \neq \emptyset$.

- If $\text{int } H(x) = \emptyset$ for some $x \in X$, then (C2) implies $\inf_{P \in \mathcal{P}} P(g(x, \xi) \leq 0) = 0$ in which case x is not a feasible solution to problem (1.2). Let us look at (C1). It follows by (4.2)

$$H(x) = \{z \in \text{bd } \Xi : g(x, z) \leq 0\} \cup \{z \in \text{int } \Xi : g(x, z) = 0\}.$$

Under (C1), $P(\{z \in \text{int } \Xi : g(x, z) = 0\}) = 0$, hence $P(H(x)) = 0$ if and only if $P(\{z \in \text{bd } \Xi : g(x, z) \leq 0\}) = 0$. The latter happens when $\Xi = \mathbb{R}^k$ in that the first term at the right hand side of equation is an empty set. However, if $P(\{z \in \text{bd } \Xi : g(x, z) \leq 0\}) > 0$, then $P(H(x)) > 0$ and consequently x might still be feasible.

- When Ξ is a discrete set, $\text{int } H(x) = \emptyset$. As we commented after Condition 4.1, (C2) may fail either because of discontinuity of $H(\cdot)$ or (4.1) is too strong to make x feasible to MPDRCC. Since Lemma 4.1 is prepared for showing continuity of $v(x)$ later on under (C2), we remind readers that the technical result is more relevant with Ξ being a continuous set.
- In the case when Ξ is a continuous set but \mathcal{P} consists of distribution with atoms, (C2) may still hold. To see this, let $\Xi = [-1, 1]$ and P be a probability distribution with $P(\xi = 0) = \frac{1}{2}$ and the rest of probability is evenly spread over $[-1, 0) \cap (0, 1]$. Let $H(x) = [-x, x]$. It is easy to see that $H(\cdot)$ is continuous on $[0.5, 1]$ and $P(H(\cdot))$ is also continuous on the interval.

LEMMA 4.2. *Let $\{P_N\}$ be a sequence of probability measures converging to P weakly. Let \bar{x} be any fixed point in X and $\{x_N\} \subset X$ be any sequence converging to \bar{x} . Assume (C2) holds for the probability measure P , that is, $H(\cdot)$ is convex-valued and continuous at \bar{x} , and P satisfies (4.1). Then*

$$\lim_{N \rightarrow \infty} P_N(H(x_N) \setminus H(\bar{x})) = 0 \quad (4.8)$$

and

$$\lim_{N \rightarrow \infty} P_N(H(\bar{x}) \setminus H(x_N)) = 0. \quad (4.9)$$

Proof. For any $\epsilon > 0$, there exists a sufficiently large number ρ such that

$$\sup_N P_N(\Xi \setminus (\Xi \cap \rho \mathbf{B})) \leq \epsilon \text{ and } P(\Xi \setminus (\Xi \cap \rho \mathbf{B})) \leq \epsilon. \quad (4.10)$$

This is because each Borel probability measure P on Ξ is tight and the weak convergence of P_N to P ensures that $\{P_N\}$ is tight; see [4, Theorem 9.2.5]. Observe that

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

Thus (4.10) entails

$$\sup_N \sup_{x \in X} P_N(H(x) \setminus (H(x) \cap \rho \mathbf{B})) \leq \epsilon \text{ and } \sup_{x \in X} P(H(x) \setminus (H(x) \cap \rho \mathbf{B})) \leq \epsilon. \quad (4.11)$$

Let us first prove (4.8). Since for any fixed $\bar{x} \in X$, $\text{bd } H(\bar{x})$ is closed, then by [19, Theorem 3.1]

$$\text{bd } H(\bar{x}) = \bigcap_{\delta > 0} (\text{bd } H(\bar{x}) + \delta \mathbf{B})$$

implies

$$P(\text{bd } H(\bar{x})) = \inf_{\delta > 0} P(\text{bd } H(\bar{x}) + \delta \mathbf{B}).$$

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

$$P(\text{bd } H(\bar{x}) + \delta_0 \mathbf{B}) \leq P(\text{bd } H(\bar{x})) + \epsilon = \epsilon,$$

where the equality holds due to our assumption that P satisfies (4.1). On the other hand, the weak convergence of P_N to P and closedness of set $\text{bd } H(\bar{x}) + \delta_0 \mathbf{B}$ enable us to obtain through [7, Theorem 2.1]

$$\limsup_{N \rightarrow \infty} P_N(\text{bd } H(\bar{x}) + \delta_0 \mathbf{B}) \leq P(\text{bd } H(\bar{x}) + \delta_0 \mathbf{B}) \leq \epsilon,$$

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

$$P_N(\text{bd } H(\bar{x}) + \delta_0 \mathbf{B}) \leq 2\epsilon, \quad (4.12)$$

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

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

for $N \geq N_2$. Thus

$$(H(x_N) \cap \rho \mathbf{B}) \setminus H(\bar{x}) \subset (H(x_N) \cap \rho \mathbf{B}) \setminus \text{int } H(\bar{x}) \subset \text{bd } H(\bar{x}) + \delta_0 \mathbf{B}. \quad (4.13)$$

Combining (4.11), (4.13) and (4.12), we obtain

$$\begin{aligned} P_N(H(x_N) \setminus H(\bar{x})) &\leq P_N((H(x_N) \cap \rho \mathbf{B}) \setminus H(\bar{x})) + \epsilon \\ &\leq P_N(\text{bd } H(\bar{x}) + \delta_0 \mathbf{B}) + \epsilon \\ &\leq 3\epsilon \end{aligned} \quad (4.14)$$

for $N \geq \max\{N_1, N_2\}$. Equation (4.8) follows as ϵ can be arbitrarily small.

We now turn to prove (4.9). Let ρ , ϵ and δ_0 be fixed as above. It follows from Lemma 4.1 that $\text{bd } H(\cdot)$ is outer semicontinuous. Together with the continuity of $H(\cdot)$, we can find $N_3 > 0$ such that

$$\begin{aligned} (H(\bar{x}) \cap \rho \mathbf{B}) \setminus H(x_N) &\subset (H(\bar{x}) \cap \rho \mathbf{B}) \setminus \text{int } H(x_N) \\ &\subset (H(\bar{x}) \cap \rho \mathbf{B}) \setminus \text{int } H(x_N) \cap \rho \mathbf{B} \\ &\subset (H(x_N) + \frac{\delta_0}{2} \mathbf{B}) \setminus \text{int } H(x_N) \cap \rho \mathbf{B} \\ &\subset \text{bd } H(x_N) \cap (\rho + \frac{\delta_0}{2}) \mathbf{B} + \frac{\delta_0}{2} \mathbf{B} \\ &\subset \text{bd } H(\bar{x}) + \delta_0 \mathbf{B}, \end{aligned}$$

for $N \geq N_3$, where the third inclusion holds due to the continuity of $H(\cdot)$ and the fifth inclusion holds due to the outer semicontinuity of $\text{bd } H(\cdot)$ (since $H(x) \setminus \text{int } H(x) = \text{bd } H(x)$). Hence, we deduce together with (4.11) and (4.12) that

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

for N being sufficiently large. The proof is complete. \blacksquare

LEMMA 4.3. *Let $\{P_N\} \subset \mathcal{P}(\Xi)$ be a sequence of probability measures and $P \in \mathcal{P}(\Xi)$. Suppose P_N converges to P weakly. Then for $x \in X$,*

$$\lim_{N \rightarrow \infty} \mathbb{E}_{P_N}[\mathbb{1}_{H(x)}(\xi)] = \mathbb{E}_P[\mathbb{1}_{H(x)}(\xi)] \quad (4.15)$$

under (C1) or (4.1) holds for P .

Proof. Observe first that if Ξ is discrete, then the continuity of $\mathbb{1}_{H(x)}(\cdot)$ over Ξ holds trivially without (C1) or (4.1). The limit then follows from the definition of weak convergence. In what follows, we consider the case when Ξ is not a discrete set.

Let Ξ_d denote the set of points in Ξ where $\mathbb{1}_{H(x)}(\cdot)$ is discontinuous over Ξ . Observe that the discontinuity occurs at the boundary of $H(x)$ located in the interior of Ξ ¹. Thus

$$\Xi_d \subseteq \{z \in \text{int } \Xi : g(x, z) = 0\}.$$

Under (C1), $P(\Xi_d) = 0$. Moreover, since $\Xi_d \subseteq \text{bd } H(x)$, $P(\Xi_d) = 0$ under (4.1) too. On the other hand, $P_N \circ \mathbb{1}_{H(x)}^{-1}$ is uniformly integrable as easily observed from the definition. Since P_N converges to P weakly, then the conclusion follows by Lemma 2.1. \blacksquare

With Lemmas 4.2 and 4.3, we are ready to derive the continuity of the robust probability function $v(\cdot)$.

THEOREM 4.2 (Pointwise continuity of the robust probability function). *Suppose \mathcal{P} is weakly compact and one of the following conditions holds:*

¹When $H(x)$ contains the boundary of Ξ , $\mathbb{1}_{H(x)}(\cdot)$ is regarded as continuous on $\text{bd } \Xi \cap H(x)$.

- (a) (C1) holds for each $P \in \mathcal{P}$ and for each $x \in X$, $g(\cdot, \xi)$ is continuous at x uniformly w.r.t. $\xi \in \Xi$;
 (b) (C2) holds for each $P \in \mathcal{P}$.

Then $v(\cdot)$ is continuous on X .

Proof. By Theorem 4.1, 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 $x_N \rightarrow x$ as $N \rightarrow \infty$. Since \mathcal{P} is weakly compact and $P \circ \mathbb{1}_{H(x)}^{-1}$ is uniformly integrable, it follows by Lemma 4.3 and [42, Proposition 1] that

$$\mathcal{V}_x := \{\mathbb{E}_P[\mathbb{1}_{H(x)}(\xi)] : P \in \mathcal{P}\}$$

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

$$v(x) = \min_{P \in \mathcal{P}} \mathbb{E}_P[\mathbb{1}_{H(x)}(\xi)] = \mathbb{E}_{P_x}[\mathbb{1}_{H(x)}(\xi)].$$

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

$$v(x_N) = \min_{P \in \mathcal{P}} \mathbb{E}_P[\mathbb{1}_{H(x_N)}(\xi)] = \mathbb{E}_{P_{x_N}}[\mathbb{1}_{H(x_N)}(\xi)].$$

Part (a): By the definitions of $v(x)$ and $v(x_N)$,

$$\begin{aligned} v(x_N) - v(x) &= \mathbb{E}_{P_{x_N}}[\mathbb{1}_{H(x_N)}(\xi)] - \mathbb{E}_{P_x}[\mathbb{1}_{H(x)}(\xi)] \\ &\leq \mathbb{E}_{P_x}[\mathbb{1}_{H(x_N)}(\xi)] - \mathbb{E}_{P_x}[\mathbb{1}_{H(x)}(\xi)] \\ &= P_x(H(x_N)) - P_x(H(x)). \end{aligned}$$

Since (C1) holds for P_x , by Theorem 4.1, $P_x(H(\cdot))$ is continuous at x . Thus

$$\limsup_{N \rightarrow \infty} v(x_N) \leq v(x). \quad (4.16)$$

In what follows, we show

$$\liminf_{N \rightarrow \infty} v(x_N) \geq v(x). \quad (4.17)$$

Let

$$\Delta_N := \mathbb{1}_{(-\infty, 0]}(g(x, \xi)) - \mathbb{1}_{(-\infty, 0]}(g(x_N, \xi)). \quad (4.18)$$

We claim that

$$\mathbb{E}_{P_{x_N}}[\Delta_N] = P_{x_N}(\Delta_N = 1) = P_{x_N}(\xi \in \Xi : g(x, \xi) = 0, g(x_N, \xi) > 0). \quad (4.19)$$

To see this, let us consider two cases.

Case (i): $\xi \in \Xi$ with either $g(x, \xi) > 0$ or $g(x, \xi) < 0$. The uniform continuity of $g(\cdot, \xi)$ in x implies that there exists a positive number N_0 independent of ξ such that $g(x_N, \xi) > 0$ or $g(x_N, \xi) < 0$ when $N \geq N_0$. This entails $\Delta_N = 0$ for $N \geq N_0$ because the two indicator functions take identical value.

Case (ii): $\xi \in \{\xi \in \Xi : g(x, \xi) = 0\}$, that is $\mathbb{1}_{(-\infty, 0]}(g(x, \xi)) = 1$. In that case,

$$\Delta_N = \begin{cases} 1 & \text{if } g(x_N, \xi) > 0, \\ 0 & \text{if } g(x_N, \xi) \leq 0. \end{cases}$$

Thus Δ_N is nonzero if and only if $\xi \in \{\xi \in \Xi : g(x, \xi) = 0, g(x_N, \xi) > 0\}$ and consequently (4.19) follows.

With (4.19), we are ready to prove (4.17). Observe that

$$\begin{aligned} v(x) - v(x_N) &= \mathbb{E}_{P_x}[\mathbb{1}_{H(x)}(\xi)] - \mathbb{E}_{P_{x_N}}[\mathbb{1}_{H(x_N)}(\xi)] \\ &\leq \mathbb{E}_{P_{x_N}}[\mathbb{1}_{H(x)}(\xi)] - \mathbb{E}_{P_{x_N}}[\mathbb{1}_{H(x_N)}(\xi)] \\ &= \mathbb{E}_{P_{x_N}}[\mathbb{1}_{(-\infty, 0]}(g(x, \xi))] - \mathbb{E}_{P_{x_N}}[\mathbb{1}_{(-\infty, 0]}(g(x_N, \xi))] \\ &= \mathbb{E}_{P_{x_N}}[\Delta_N] \\ &= P_{x_N}(\xi \in \Xi : g(x, \xi) = 0, g(x_N, \xi) > 0) \\ &= 0, \end{aligned}$$

where the last equality holds because P_{x_N} satisfies $P_{x_N}(\xi \in \Xi : g(y, \xi) = 0) = 0$ for every $y \in X$. This shows $v(x_N) \geq v(x)$. By taking infimum on both sides of the inequality, we obtain (4.17). The conclusion follows from a combination of (4.16) and (4.17).

Part (b): Note that inequality (4.16) relies only on the continuity of $P_x(H(\cdot))$. Thus the upper semicontinuity of v remains true.

In what follows, we show (4.17). Observe first that

$$\begin{aligned} v(x) - v(x_N) &= \mathbb{E}_{P_x}[\mathbb{1}_{H(x)}(\xi)] - \mathbb{E}_{P_{x_N}}[\mathbb{1}_{H(x_N)}(\xi)] \\ &\leq \mathbb{E}_{P_{x_N}}[\mathbb{1}_{H(x)}(\xi)] - \mathbb{E}_{P_{x_N}}[\mathbb{1}_{H(x_N)}(\xi)] \\ &\leq P_{x_N}(H(x) \setminus H(x_N)), \end{aligned}$$

where the first inequality holds due to the definition of P_{x_N} . By (4.9), we know $\lim_{N \rightarrow \infty} P_{x_N}(H(x) \setminus H(x_N)) = 0$, and then (4.17) holds. \blacksquare

Theorem 4.2 extends pointwise continuity of probability function to that of robust probability function. We give a simple example to illustrate.

EXAMPLE 4.2. Let $\mathcal{P}(\mathbb{R})$ be the set of all probability measures on \mathbb{R} and

$$\mathcal{P} := \left\{ P \in \mathcal{P}(\mathbb{R}) : \begin{array}{l} \mathbb{E}_P[\xi] = 0, \mathbb{E}_P[\xi^2] \in [\sigma_1^2, \sigma_2^2] \\ P \text{ follows a normal distribution on } \mathbb{R} \end{array} \right\},$$

where σ_1, σ_2 are positive constants with $\sigma_1 < \sigma_2$. Observe that \mathcal{P} is tight. To see this, for any $K > 0$ and $P \in \mathcal{P}$, it follows by the well-known Chebyshev inequality

$$P(|\xi - 0| \geq K\sigma_1) \leq \frac{1}{K^2}$$

which means P is tight. Moreover it is easy to verify that \mathcal{P} is closed and hence \mathcal{P} is weakly compact.

Let $g(x, z) = x - z$. Then $H(x) = \{z \in \mathbb{R} : x - z \leq 0\} = [x, +\infty)$ and $\text{bd } H(x) = \{z \in \mathbb{R} : x - z = 0\} = \{x\}$. Both (C1) and (C2) are satisfied. Thus all conditions of Theorem 4.2 are fulfilled. By that theorem, v is continuous. Indeed, we can easily obtain a closed form for v , that is,

$$v(x) = \begin{cases} \frac{1}{2} \left(1 - \frac{1}{\sqrt{\pi}} \int_{-\frac{x}{\sqrt{2}\sigma_2}}^{\frac{x}{\sqrt{2}\sigma_2}} e^{-t^2} dt \right), & \text{for } x < 0, \\ \frac{1}{2}, & \text{for } x = 0, \\ \frac{1}{2} \left(1 - \frac{1}{\sqrt{\pi}} \int_{-\frac{x}{\sqrt{2}\sigma_1}}^{\frac{x}{\sqrt{2}\sigma_1}} e^{-t^2} dt \right), & \text{for } x > 0, \end{cases}$$

which also shows that v is continuous on \mathbb{R} . \blacksquare

To conclude this section, we give a brief summary. We propose two conditions (C1) and (C2) for the continuity of probability function $P(H(\cdot))$ and robust probability function $v(\cdot)$. It seems relatively easier to establish the continuity under (C1). However, given that there are some interesting circumstances when (C2) holds whereas (C1) does not, we believe that it is useful to present both conditions for the desired continuity particularly when Ξ is a continuous set. In the case when Ξ is a discrete set, (C2) is largely unreasonable as we have discussed in Remark 4.1.

5. Sufficient conditions for Assumption 3.1 . We now move on to investigate sufficient conditions for Assumption 3.1. Let us start with a technical result which concerns the convergence of P_N to P under the pseudo-metric \mathcal{D} . We start by considering the case when \mathcal{P}_N and \mathcal{P} are singletons and then extend the discussion to general case.

LEMMA 5.1 (Convergence of $\mathcal{D}(P_N, P)$). *Let $\{P_N\} \subset \mathcal{P}(\Xi)$ be a sequence of probability measures and $P \in \mathcal{P}(\Xi)$. Suppose P_N converges to P weakly. Then*

$$\lim_{N \rightarrow \infty} \mathcal{D}(P_N, P) = 0 \tag{5.1}$$

under one of the following conditions:

- (a) $g(\cdot, \xi)$ is continuous on X uniformly w.r.t. $\xi \in \Xi$ and (C1) holds for P ;
(b) (C2) holds for P , that is, $H(\cdot)$ is convex-valued and continuous on X , and P satisfies (4.1).

Proof. By the definition of the pseudo-metric,

$$\mathcal{D}(P_N, P) = \sup_{g \in \mathcal{G}} |\mathbb{E}_{P_N}[g] - \mathbb{E}_P[g]| = \sup_{x \in X} |\mathbb{E}_{P_N}[\mathbb{1}_{H(x)}(\xi)] - \mathbb{E}_P[\mathbb{1}_{H(x)}(\xi)]|,$$

we need to prove that

$$\lim_{N \rightarrow \infty} \sup_{x \in X} |\mathbb{E}_{P_N}[\mathbb{1}_{H(x)}(\xi)] - \mathbb{E}_P[\mathbb{1}_{H(x)}(\xi)]| = 0. \quad (5.2)$$

Assume for a contradiction that (5.2) fails to hold. Then there exist a constant $\delta > 0$ and a subsequence $\{P_N\}$ such that

$$\sup_{x \in X} |\mathbb{E}_{P_N}[\mathbb{1}_{H(x)}(\xi)] - \mathbb{E}_P[\mathbb{1}_{H(x)}(\xi)]| > \delta.$$

So for each N we may pick x_N such that

$$|\mathbb{E}_{P_N}[\mathbb{1}_{H(x_N)}(\xi)] - \mathbb{E}_P[\mathbb{1}_{H(x_N)}(\xi)]| \geq \delta/2$$

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 $x_N \rightarrow \bar{x} \in X$. By the triangle inequality,

$$\begin{aligned} |\mathbb{E}_{P_N}[\mathbb{1}_{H(x_N)}(\xi)] - \mathbb{E}_P[\mathbb{1}_{H(x_N)}(\xi)]| &\leq |\mathbb{E}_{P_N}[\mathbb{1}_{H(\bar{x})}(\xi)] - \mathbb{E}_P[\mathbb{1}_{H(\bar{x})}(\xi)]| \\ &\quad + |\mathbb{E}_{P_N}[\mathbb{1}_{H(x_N)}(\xi)] - \mathbb{E}_{P_N}[\mathbb{1}_{H(\bar{x})}(\xi)]| \\ &\quad + |\mathbb{E}_P[\mathbb{1}_{H(\bar{x})}(\xi)] - \mathbb{E}_P[\mathbb{1}_{H(x_N)}(\xi)]|. \end{aligned} \quad (5.3)$$

Following Theorem 4.1, there exists N_0 sufficiently large such that

$$|\mathbb{E}_P[\mathbb{1}_{H(\bar{x})}(\xi)] - \mathbb{E}_P[\mathbb{1}_{H(x_N)}(\xi)]| \leq \delta/8. \quad (5.4)$$

Let us now estimate the second term at the right hand side of (5.3), and this is the place where conditions (a) and (b) make a difference.

Part (a): Let Δ_N be defined as in (4.18) with x being replaced by \bar{x} there. Observe that

$$\begin{aligned} |\mathbb{E}_{P_N}[\mathbb{1}_{H(x_N)}(\xi)] - \mathbb{E}_{P_N}[\mathbb{1}_{H(\bar{x})}(\xi)]| &= |\mathbb{E}_{P_N}[\mathbb{1}_{(-\infty, 0]}(g(x_N, \xi))] - \mathbb{E}_{P_N}[\mathbb{1}_{(-\infty, 0]}(g(\bar{x}, \xi))]| \\ &= |\mathbb{E}_{P_N}[-\Delta_N]| \\ &= P_N(\xi \in \Xi : g(\bar{x}, \xi) = 0, g(x_N, \xi) > 0) \\ &\leq P_N(\xi \in \Xi : g(\bar{x}, \xi) = 0), \end{aligned}$$

where the third equality follows from (4.19) (with x being replaced by \bar{x} there). Since P_N converges to P weakly and $P(\xi \in \Xi : g(\bar{x}, \xi) = 0) = 0$, the last term goes to zero as $N \rightarrow \infty$ [7, Theorem 2.1]. This shows for N sufficiently large,

$$|\mathbb{E}_{P_N}[\mathbb{1}_{H(x_N)}(\xi)] - \mathbb{E}_{P_N}[\mathbb{1}_{H(\bar{x})}(\xi)]| \leq \delta/8. \quad (5.5)$$

Part (b): Observe first that

$$\begin{aligned} |\mathbb{E}_{P_N}[\mathbb{1}_{H(x_N)}(\xi)] - \mathbb{E}_{P_N}[\mathbb{1}_{H(\bar{x})}(\xi)]| &= |P_N(H(x_N)) - P_N(H(\bar{x}))| \\ &\leq P_N(H(x_N) \setminus H(\bar{x})) + P_N(H(\bar{x}) \setminus H(x_N)). \end{aligned} \quad (5.6)$$

On the other hand, we deduce from Lemma 4.2 that

$$P_N(H(x_N) \setminus H(\bar{x})) + P_N(H(\bar{x}) \setminus H(x_N)) \leq \delta/8, \quad (5.7)$$

when N is sufficiently large. Combining (5.3)-(5.7), we deduce that

$$|\mathbb{E}_{P_N}[\mathbb{1}_{H(\bar{x})}(\xi)] - \mathbb{E}_P[\mathbb{1}_{H(\bar{x})}(\xi)]| \geq \delta/4,$$

which leads to a contradiction to the result in Lemma 4.3 as desired. The proof is complete. \blacksquare

We are now ready to discuss sufficient conditions for Assumption 3.1 (a) and (b).

PROPOSITION 5.1 (Sufficient conditions for Assumption 3.1 (a)). Let \mathcal{P}_N converge 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 $P \in \mathcal{P}$. Let condition (a) or (b) in Lemma 5.1 hold for any $P \in \mathcal{P}$. Then Assumption 3.1 (a) is fulfilled.

Proof. The conclusion follows from Lemma 5.1. Indeed, assume for the sake of a contradiction that $\lim_{N \rightarrow \infty} \mathcal{D}(\mathcal{P}_N, \mathcal{P}) \neq 0$. Then there exist a positive constant ϵ_0 and a subsequence $\{\mathcal{P}_{N_k}\}$ such that

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

i.e., there exists $P_k \in \mathcal{P}_{N_k}$ such that $\mathcal{D}(P_k, \mathcal{P}) \geq \epsilon_0/2$. Without loss of generality, we can assume that P_k converges to some $P \in \mathcal{P}$ weakly. Under condition (a) or (b), it follows from Lemma 5.1 that $\lim_{k \rightarrow \infty} \mathcal{D}(P_k, \mathcal{P}) = 0$, a contradiction as desired. \blacksquare

It is possible to derive sufficient conditions for Assumption 3.1 (b) as well.

PROPOSITION 5.2 (Sufficient conditions for Assumption 3.1 (b)). Let \mathcal{P} be weakly compact. For any $P \in \mathcal{P}$, there exists a sequence $\{P_N\} \in \mathcal{P}_N$ such that P_N converges to P weakly. Let condition (a) or (b) in Lemma 5.1 hold for any $P \in \mathcal{P}$. Then Assumption 3.1 (b) is fulfilled.

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

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

Since \mathcal{P} is weakly compact, there exists $P^k \in \mathcal{P}$ such that $\mathcal{D}(P^k, \mathcal{P}_{N_k}) \geq 2\epsilon_0$, and we may suppose without loss of generality that P^k converges to $P \in \mathcal{P}$ weakly. Through Lemma 5.1, the weak convergence implies $\mathcal{D}(P^k, P) \leq \epsilon_0$ for k sufficiently large. Using the triangle inequality of the pseudo-metric, we have

$$2\epsilon_0 \leq \mathcal{D}(P^k, \mathcal{P}_{N_k}) \leq \mathcal{D}(P^k, P) + \mathcal{D}(P, \mathcal{P}_{N_k}) \leq \mathcal{D}(P, \mathcal{P}_{N_k}) + \epsilon_0. \quad (5.8)$$

Since for the given P , there exists $P_{N_k} \in \mathcal{P}_{N_k}$ such that P_{N_k} converges to P weakly (by taking a subsequence if necessarily), hence

$$\mathcal{D}(P, \mathcal{P}_{N_k}) \leq \mathcal{D}(P, P_{N_k}) \rightarrow 0,$$

this effectively leads to a contradiction as desired through (5.8). \blacksquare

Note that we only require the probability measures in \mathcal{P} to satisfy (C1) or (C2) but this is not required for \mathcal{P}_N . A special case is that both \mathcal{P} and \mathcal{P}_N are singletons, and \mathcal{P}_N is constructed through empirical probability measures. The following corollary says that Assumption 3.1 holds in such a case.

COROLLARY 5.1. Let P be the true probability distribution of ξ satisfying condition (a) or (b) in Lemma 5.1, and ξ_1, \dots, ξ_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 3.1 holds.

Proof. Observe first that when \mathcal{P} and \mathcal{P}_N are singletons, Assumption 3.1 (a) coincides with Assumption 3.1 (b). By the well-known Glivenko-Cantelli Theorem (see e.g. [43]), P_N converges to P weakly. This verifies the conditions of Lemma 5.1, therefore, $\mathcal{D}(P_N, P) \rightarrow 0$ as $N \rightarrow \infty$. \blacksquare

Before concluding this section, we note that Lemma 5.1 may be of independent interest in stochastic programming because it addresses uniform convergence of $P_N(H(x))$ to $P(H(x))$ when P_N converges to P weakly, Corollary 5.1 is one of the interesting examples. The combination of Propositions 5.1 and 5.2 is virtually an extension of Lemma 5.1 to the robust chance constraint.

6. Applications. In this section, we apply the convergence results to three cases with ambiguity set \mathcal{P} being defined by general prior moment conditions (Section 6.1), mean-absolute deviation condition (Section 6.2) and KL-divergence (Section 6.3).

6.1. Piecewise uniform approximation of ambiguity set based on moment condition. We start by considering \mathcal{P} being defined through moment conditions. Let $\Phi : \mathbb{R}^k \rightarrow \mathbb{R}^q$ be a continuous vector-valued function and $\xi : \Omega \rightarrow \mathbb{R}^k$ be a random vector with support set Ξ . We consider

$$\mathcal{P} := \left\{ P \in \mathcal{P}(\Xi) : \begin{array}{l} \mathbb{E}_P[\Phi(\xi)] \leq 0, P \text{ is absolutely continuous w.r.t the Lebesgue} \\ \text{measure and its density function } \rho(\xi) \geq \delta > 0 \text{ over } \Xi \end{array} \right\}, \quad (6.1)$$

where δ is a fixed constant. 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

$$\Xi := \{z \in \mathbb{R}^k : L \leq z \leq U\}.$$

Since Ξ is compact and the density function is bounded from below, \mathcal{P} is tight and closed. By Prokhorov's theorem, \mathcal{P} is weakly compact (see [4] and [42, Proposition 7] for a more recent discussions in this regard).

With \mathcal{P} constructed as in (6.1), it is difficult to solve the resulting MPDRCC (1.2) unless the underlying functions have a very specific structure (see [49]). In what follows, we propose a piecewise uniform approximation scheme for \mathcal{P} and investigate convergence of the resulting problem (2.11) to problem (1.2) in terms of the optimal values and the optimal solutions.

Let Ξ_1, \dots, Ξ_N be a partition of Ξ with

$$\begin{aligned} \Xi_1 &:= \{z : L \leq z \leq L + \frac{1}{N}(U - L)\} \text{ and} \\ \Xi_i &:= \{z : L \leq z \leq L + \frac{i}{N}(U - L)\} \setminus \{z : L \leq z \leq L + \frac{i-1}{N}(U - L)\} \quad \text{for } i = 2, \dots, N. \end{aligned} \quad (6.2)$$

Let

$$\mathcal{P}_N := \left\{ P_N \in \mathcal{P}(\Xi) : \begin{array}{l} \mathbb{E}_{P_N}[\Phi(\xi)] \leq 0, P_N \text{ is uniformly distributed over each } \Xi_i, \\ P_N(\Xi_i) = p_i, \sum_{i=1}^N p_i = 1, p_i \geq \delta V(\Xi_i) \text{ for } i = 1, \dots, N \end{array} \right\}, \quad (6.3)$$

where δ is defined as in (6.1) and $V(\Xi_i) := \int_{\Xi_i} 1 dz$ for $i = 1, \dots, N$. We investigate the 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 and constraint $p_i \geq \delta V(\Xi_i)$ ensures each density function is lower bounded by δ . In what follows, we show that \mathcal{P}_N converges to \mathcal{P} under the Kolmogorov metric and henceforth \mathcal{P}_N converges to \mathcal{P} weakly (see[18]).

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 probability distribution with piecewise linear density function to approximate \mathcal{P} because the former is relatively easier to calculate. Thus, what we are proposing here is to extend the above approximation scheme to the case when \mathcal{P} is a set defined through some moment conditions. We establish that \mathcal{P}_N converges to \mathcal{P} under some appropriate metric in the following proposition.

THEOREM 6.1. *Suppose there exists $P_0 \in \mathcal{P}$ such that $\mathbb{E}_{P_0}[\Phi(\xi)] < 0$. Then the following assertions hold.*

- (i) $\lim_{N \rightarrow \infty} \mathbb{H}_K(\mathcal{P}, \mathcal{P}_N) = 0$;
- (ii) if, in addition, $H(\cdot)$ is convex-valued and continuous over X , then $\lim_{N \rightarrow \infty} \mathcal{H}(\mathcal{P}, \mathcal{P}_N) = 0$ and v is continuous over X ;
- (iii) furthermore, if condition (3.4) is fulfilled, then we can establish convergence of the optimal values and optimal solutions as specified in (3.5).

Proof. Part (i): 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 the 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 (6.2) satisfying

$$\sup_{1 \leq i \leq N} P^\lambda(\Xi_i) = \sup_{1 \leq i \leq N} F^\lambda \left(L + \frac{i}{N}(U - L) \right) - F^\lambda \left(L + \frac{i-1}{N}(U - L) \right) \leq \epsilon, \quad (6.4)$$

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

$$P_N^\lambda(\Xi_i) = P^\lambda(\Xi_i) \quad (6.5)$$

for $i = 1, \dots, N$. Since the density function of P^λ is lower bounded by δ , we can obtain from the above equality that density function of P_N^λ also has this property. In what follows, we show $\mathbb{D}_K(P^\lambda, P_N^\lambda) \leq \epsilon$.

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

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

where the second inequality holds due to (6.4) and the last equality by (6.5). Likewise,

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

A combination of the two inequalities gives rise to

$$|F_N^\lambda(\eta) - F^\lambda(\eta)| \leq \epsilon.$$

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

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

which means that P_N^λ converges to P^λ under the Kolmogorov metric and by [18, Theorem 6] in the weak topology.

Next, we show that P_N^λ satisfies the moment condition in (6.3). Since $\Phi(\cdot)$ is a continuous function, the weak convergence guarantees

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

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

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

This implies that 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 $\lim_{N \rightarrow \infty} \mathbb{D}_K(\mathcal{P}, \mathcal{P}_N) = 0$ holds. We can change \mathbb{D}_K to \mathbb{H}_K in that $\mathcal{P}_N \subset \mathcal{P}$.

Parts (ii): We now prove the convergence of \mathcal{P}_N to \mathcal{P} under the pseudo-metric. It is well known in the literature of probability theory that convergence under the Kolmogorov metric implies weak convergence; see [18]. Using this result, we can easily obtain the weak convergence of \mathcal{P}_N to \mathcal{P} from the result that $\mathbb{H}_K(\mathcal{P}_N, \mathcal{P}) \rightarrow 0$ as $N \rightarrow \infty$. Since $H(x)$ is convex set for any $x \in X$ and P is absolutely continuous w.r.t the Lebesgue measure for any $P \in \mathcal{P}$, then $P(\text{bd } H(x)) = 0$. Together with the continuity of $H(\cdot)$, this shows that (C2) holds. Hence it follows by Propositions 5.1 and 5.2 that \mathcal{P}_N converges to \mathcal{P} under the pseudo-metric. The continuity of v follows from Theorem 4.2.

Part (iii): The conclusions follow from Theorem 3.2 as all of the required conditions are fulfilled. \blacksquare

Theorem 6.1 is established for the case when Ξ is a bounded rectangle. It might be interesting to extend the theorem to the case when Ξ is the whole space \mathbb{R}^k . Our conjecture is that this is possible when \mathcal{P} is weakly compact. We leave this for future work. Note also that the definition of the moment conditions and the Slater condition rule out equality constraints. It is an open question that whether similar results can be established for moment problems with equality constraints.

In what follows, we discuss how to solve MPDRCC $_N$ when \mathcal{P}_N is defined as in (6.3), that is

$$\begin{aligned} \min_{x \in X} \quad & f(x) \\ \text{s.t.} \quad & \inf_{P \in \mathcal{P}_N} P(H(x)) \geq 1 - \beta. \end{aligned} \quad (6.6)$$

For a fixed partition $\Xi_i, i = 1, \dots, N$ of Ξ , we have

$$\begin{aligned} \inf_{P \in \mathcal{P}_N} P(H(x)) &= \inf_{p \in \mathbb{R}^N} \sum_{i=1}^N \frac{p_i}{V(\Xi_i)} V(H(x) \cap \Xi_i) \\ \text{s.t.} \quad & \sum_{i=1}^N \frac{p_i}{V(\Xi_i)} \int_{\Xi_i} \Phi_l(\xi) d\xi \leq 0, \quad l = 1, \dots, q, \\ & \sum_{i=1}^N p_i = 1, p_i \geq \delta V(\Xi_i), \quad i = 1, \dots, N. \end{aligned} \quad (6.7)$$

The Lagrangian dual of the above problem is

$$\begin{aligned} \sup_{\lambda_0, \lambda_1, \dots, \lambda_q} \quad & \lambda_0 + \delta \sum_{i=1}^N V(\Xi_i) \left(\frac{V(H(x) \cap \Xi_i)}{V(\Xi_i)} - \lambda_0 + \sum_{l=1}^q \lambda_l \frac{\int_{\Xi_i} \Phi_l(\xi) d\xi}{V(\Xi_i)} \right) \\ \text{s.t.} \quad & \lambda_0 \in \mathbb{R}, \lambda_l \geq 0, \quad l = 1, \dots, q, \\ & \frac{V(H(x) \cap \Xi_i)}{V(\Xi_i)} - \lambda_0 + \sum_{l=1}^q \lambda_l \frac{\int_{\Xi_i} \Phi_l(\xi) d\xi}{V(\Xi_i)} \geq 0, \quad i = 1, \dots, N, \end{aligned}$$

or equivalently

$$\begin{aligned} \sup_{\lambda_0, \lambda_1, \dots, \lambda_q} \quad & \lambda_0 + \delta \sum_{i=1}^N V(\Xi_i) \mathbb{E}_{P^i} \left[\mathbb{1}_{H(x)}(\xi) - \lambda_0 + \sum_{l=1}^q \lambda_l \Phi_l(\xi) \right] \\ \text{s.t.} \quad & \lambda_0 \in \mathbb{R}, \lambda_l \geq 0, \quad l = 1, \dots, q, \\ & \mathbb{E}_{P^i} \left[\mathbb{1}_{H(x)}(\xi) - \lambda_0 + \sum_{l=1}^q \lambda_l \Phi_l(\xi) \right] \geq 0, \quad i = 1, \dots, N, \end{aligned}$$

where P^i is a uniform distribution over Ξ_i with $P^i(\Xi_i) = 1$ for $i = 1, \dots, N$. Consequently problem (6.6) can be written as

$$\begin{aligned} \min_{x, \lambda_0, \lambda_1, \dots, \lambda_q} \quad & f(x) \\ \text{s.t.} \quad & x \in X, \lambda_l \geq 0, \quad l = 1, \dots, q, \\ & \lambda_0 + \delta \sum_{i=1}^N V(\Xi_i) \mathbb{E}_{P^i} \left[\mathbb{1}_{H(x)}(\xi) - \lambda_0 + \sum_{l=1}^q \lambda_l \Phi_l(\xi) \right] \geq 1 - \beta, \\ & \mathbb{E}_{P^i} \left[\mathbb{1}_{H(x)}(\xi) - \lambda_0 + \sum_{l=1}^q \lambda_l \Phi_l(\xi) \right] \geq 0, \quad i = 1, \dots, N. \end{aligned} \quad (6.8)$$

Note that the Slater condition in Theorem 6.1 ensures $\mathcal{P} \neq \emptyset$. From the proof of the theorem, we can see $\mathcal{P}_N \neq \emptyset$. Thus the feasible set of problem (6.7) is nonempty, which ensures that its Lagrangian dual has no dual gap (because it is a linear programming problem). Note also that when $H(\cdot)$ takes some special structure such as polyhedron, the expected values in problem (6.8) 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 a fixed partition $\{\Xi_1, \dots, \Xi_N\}$ of Ξ , let $\xi_i^1, \dots, \xi_i^{M_i}$ be iid samples drawn over Ξ_i for $i = 1, \dots, N$. We use $\frac{1}{M_i} \sum_{j=1}^{M_i} \mathbb{1}_{H(x)}(\xi_i^j)$, $\frac{1}{M_i} \sum_{j=1}^{M_i} \Phi(\xi_i^j)$ to approximate $\mathbb{E}_{P^i}[\mathbb{1}_{H(x)}(\xi)]$ and $\mathbb{E}_{P^i}[\Phi(\xi)]$ respectively. The resulting SAA scheme of problem (6.8) can be written as

$$\begin{aligned} \min_{x, \lambda_0, \lambda_1, \dots, \lambda_q} \quad & f(x) \\ \text{s.t.} \quad & x \in X, \lambda_l \geq 0, l = 1, \dots, q, \\ & \lambda_0 + \delta \sum_{i=1}^N \frac{V(\Xi_i)}{M_i} \sum_{j=1}^{M_i} \left(\mathbb{1}_{H(x)}(\xi_i^j) - \lambda_0 + \sum_{l=1}^q \lambda_l \Phi_l(\xi_i^j) \right) \geq 1 - \beta, \\ & \frac{1}{M_i} \sum_{j=1}^{M_i} \left(\mathbb{1}_{H(x)}(\xi_i^j) - \lambda_0 + \sum_{l=1}^q \lambda_l \Phi_l(\xi_i^j) \right) \geq 0, \quad i = 1, \dots, N. \end{aligned} \quad (6.9)$$

Note that here sampling is based on known uniform distributions rather than empirical data. The end problem (6.9) is an ordinary constrained optimization problem. In other words, to solve MPDRCC with \mathcal{P} defined in (6.1), we may solve problem (6.9) to obtain approximate optimal values and solutions.

6.2. Ambiguity set based on mean-absolute deviation. We now move on to consider the ambiguity set \mathcal{P} being defined through mean-absolute deviation, namely

$$\mathcal{P} := \{P \in \mathcal{P}(\Xi) : \mathbb{E}_P[\xi] = \mu, \mathbb{E}_P[|\xi - \mu|] \leq d\}, \quad (6.10)$$

where $\mu \in \mathbb{R}^k$ and $d \in \mathbb{R}_+^k$ denote the unknown true mean value and absolute deviation of random vector ξ respectively. Let $\{\xi^i\}_{i=1}^N$ be an independent and identically distributed sample drawn from the true probability distribution P of ξ . We use sample mean and sample absolute deviation to approximate their true counterpart. Let

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

and

$$\mathcal{P}_N := \{P \in \mathcal{P}(\Xi) : \mathbb{E}_P[\xi] = \mu_N, \mathbb{E}_P[|\xi - \mu_N|] \leq d_N\} \quad (6.11)$$

be an approximation of \mathcal{P} . By Hoffman's Lemma for moment problems in (see [42, Lemma 2] or [47, Lemma 2.1]), we can quantify the approximation of \mathcal{P}_N to \mathcal{P} through the total variation metric.

PROPOSITION 6.1. *If $\Xi = \mathbb{R}^k$, then there exists a positive constant C such that*

$$\mathbb{H}_{TV}(\mathcal{P}_N, \mathcal{P}) \leq C (2\|\mu_N - \mu\| + \max\{\|(d_N - d)_+\|, \|(d - d_N)_+\|\}), \quad (6.12)$$

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

Proof. Since $\Xi = \mathbb{R}^k$ and the set $\{\mathbb{E}_P[\xi] : P \in \mathcal{P}(\Xi)\} = \mathbb{R}^k$, there exists a probability distribution P_0 such that $\mathbb{E}_{P_0}[\xi] = \mu$, $\mathbb{E}_{P_0}[|\xi - \mu|] < d$, which means the Slater condition holds; see [45]. For any $Q \in \mathcal{P}(\Xi)$, it follows from Hoffman's Lemma [42, Lemma 2] that there exists a positive constant C_1 such that

$$D_{TV}(Q, \mathcal{P}) \leq C_1 (\|\mathbb{E}_Q[\xi] - \mu\| + \|(\mathbb{E}_Q[|\xi - \mu|] - d)_+\|).$$

Let $Q \in \mathcal{P}_N$. Then

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

where the second inequality holds due to the fact that $\mathbb{E}[(a+b)_+] \leq \mathbb{E}[(a)_+] + \mathbb{E}[(b)_+]$. Since μ_N and d_N converge to μ and d respectively, the Slater condition for the system defining \mathcal{P}_N holds when N is sufficiently large, we have by virtue of the same lemma that for any given $Q \in \mathcal{P}$,

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

The conclusion follows by combining the two inequalities above and setting $C = \max\{C_1, C_2\}$. \blacksquare

REMARK 6.1. Note that both \mathcal{P} and \mathcal{P}_N may constitute discrete probability measures, thus condition (4.1) is not guaranteed here. However, due to the specific structure of \mathcal{P} and \mathcal{P}_N , we are able to show $\mathbb{H}_{TV}(\mathcal{P}_N, \mathcal{P}) \rightarrow 0$ and hence Assumption 3.1 holds because the set \mathcal{G} defined as (2.2) is bounded in this context, i.e., $\sup_{g \in \mathcal{G}} \|g\| \leq 1$, and consequently $\mathcal{D}(P, Q) \leq \mathbb{D}_{TV}(P, Q)$. By Theorem 3.1, $v_N(\cdot)$ converges to $v(\cdot)$ uniformly over X as N increases.

In order to apply Theorem 3.2 here, we first show the continuity of $v(\cdot)$.

Consider a special case where $g(x, z)$ is linear in z , namely

$$g(x, z) := b(x) - A(x)z, \quad (6.13)$$

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

LEMMA 6.1. *Let $x \in X$ be fixed and the ambiguity set be defined as in (6.10). Let*

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

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

$$\begin{aligned} \sup_{\alpha, \rho, \gamma, \eta} \quad & \alpha + \mu^T \rho - d^T \gamma \\ \text{s.t.} \quad & \alpha + \mu^T \rho \leq 1, \\ & -\gamma \leq \rho \leq \gamma, \\ & \alpha + \mu^T (\rho - A(x)^T \eta) + b(x)^T \eta \leq 0, \\ & -\gamma \leq \rho - A(x)^T \eta \leq \gamma, \\ & \alpha \in \mathbb{R}, \rho \in \mathbb{R}^k, \gamma \in \mathbb{R}_+^k, \eta \in \mathbb{R}_+^m. \end{aligned} \quad (6.14)$$

Proof. Let \mathcal{M}_+ denote the positive linear space of all signed measures generated by $\mathcal{P}(\mathbb{R}^k)$, let

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

By the definition of \mathcal{P} in (6.10), $v(x)$ can be written as

$$\begin{aligned} v(x) = \inf_{P \in \mathcal{M}_+} \quad & \langle P, \mathbb{1}_{A(x)\xi \geq b(x)}(\xi) \rangle \\ \text{s.t.} \quad & \langle P, \xi \rangle = \mu, \\ & \langle P, |\xi - \mu| \rangle \leq d, \\ & \langle P, 1 \rangle = 1. \end{aligned}$$

Since $\Xi = \mathbb{R}^k$, then $\{\mathbb{E}_P[\xi], P \in \mathcal{P}(\mathbb{R}^k)\} = \mathbb{R}^k$, and hence there exists $P_0 \in \mathcal{P}(\mathbb{R}^k)$ such that $\mathbb{E}_{P_0}[\xi] = \mu$, $\mathbb{E}_{P_0}[|\xi - \mu|] < d$, i.e., the strong duality holds (see [45, Example 2.1]) and [40, Proposition 3.4]). The Lagrangian dual problem is

$$\begin{aligned} \sup_{\gamma \geq 0, \alpha, \rho} \quad & \alpha + \mu^T \rho - d^T \gamma \\ \text{s.t.} \quad & \alpha + z^T \rho - |z - \mu|^T \gamma \leq \mathbb{1}_{A(x)z \geq b(x)}(z), \forall z \in \Xi. \end{aligned} \quad (6.15)$$

The constraint of (6.15) is equivalent to

$$\begin{cases} \alpha + z^T \rho - |z - \mu|^T \gamma \leq 1, \forall z \in \Xi, \\ \alpha + z^T \rho - |z - \mu|^T \gamma \leq 0, \forall z \in \Xi \text{ such that } A(x)z < b(x). \end{cases} \quad (6.16)$$

The first constraint in (6.16) means the optimal value of the linear problem:

$$\begin{aligned} \sup_{z, \theta} \quad & \alpha + z^T \rho - \theta^T \gamma \\ \text{s.t.} \quad & z - \mu \leq \theta, \\ & -z + \mu \leq \theta, \end{aligned}$$

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

$$\alpha + \mu^T \rho \leq 1, \quad -\gamma \leq \rho \leq \gamma. \quad (6.17)$$

Likewise, the second constraint in (6.16) holds if there exists $\eta \in \mathbb{R}_+^m$ satisfying

$$\alpha + \mu^T(\rho - A(x)^T \eta) + b(x)^T \eta \leq 0, \quad -\gamma \leq \rho - A(x)^T \eta \leq \gamma. \quad (6.18)$$

Combining (6.15), (6.17) and (6.18), we obtain (6.14). \blacksquare

PROPOSITION 6.2. *Assume that for any fixed $x \in X$, the following system of equalities*

$$A(x)^T y = 0 \quad (6.19)$$

has an unique solution 0 and problem (6.14) satisfies the Slater constraint qualification. Then the robust probability function $v(\cdot)$ is continuous on X . If, in addition, condition (3.4) is fulfilled, then we can establish convergence of the optimal values and optimal solutions as specified in (3.5).

Proof. We use [9, Proposition 4.4] to show the continuity of $v(\cdot)$. First we will verify the conditions of the proposition in this context.

Let $\Phi(x)$ denote the feasible set of problem (6.14). It is easy to see that the objective function of (6.14) is continuous, $\Phi(\cdot)$ is closed valued and outer semicontinuous, thus conditions (i), (ii) of the proposition are fulfilled here. Moreover, the Slater condition implies condition (iv). So we are left to verify condition (iii), i.e., there exists a constant $c \in \mathbb{R}$ such that the lower level set

$$\{(\alpha, \rho, \gamma, \eta) \in \Phi(x) : -\alpha - \mu^T \rho + d^T \gamma \leq c\}$$

is uniformly bounded w.r.t x . Observe that a combination of the inequalities

$$-\alpha - \mu^T \rho + d^T \gamma \leq c, \quad \alpha + \mu^T \rho \leq 1 \quad (6.20)$$

implies $d^T \gamma \leq 1 + c$. Together with the fact that $d \geq 0$, $\gamma \geq 0$, we conclude that γ is bounded. The boundedness of ρ follows from constraint $\|\rho\| \leq \gamma$ and the boundedness of α is in turn implied by the boundedness of ρ and γ via inequalities (6.20).

Next, we show the boundedness of η . Based on the analysis above and the fourth constraint of problem (6.14), there is a positive constant K such that $\|A(x)^T \eta\| \leq K$. Assume for a contradiction that the set of solutions to the inequality above is not uniformly bounded w.r.t. x , then there exist a sequence $\{x_k\}$ and a sequence $\{\eta_k\}$ with $\|\eta_k\| \rightarrow \infty$ such that $\|A(x_k)^T \eta_k\| \leq K$. By taking a subsequence if needed, we may assume without loss of generality that $x_k \rightarrow \hat{x}$ and $\eta_k / \|\eta_k\| \rightarrow \hat{\eta}$. Dividing both sides of the inequality by $\|\eta_k\|$ and driving k to infinity, we arrive at $\|A(\hat{x})^T \hat{\eta}\| = 0$, a contradiction to condition (6.19), since $\|\hat{\eta}\| = 1$.

To complete the proof, we note that $v_N(x)$ converges to $v(x)$ uniformly as established in Remark 6.1. Together with the continuity of v , we can apply Theorem 3.2 to establish convergence of the optimal value and the optimal solutions of (2.11) to their true counterpart of (1.2) in this particular setting. \blacksquare

Note that in this proposition, we derive continuity of $v(x)$ without using sufficient conditions in Section 4. The convergence of \mathcal{P}_N to \mathcal{P} under the pseudo-metric is also established without applying sufficient conditions in Section 5. In other words, we are able to derive the convergence results outlined in Section 3 directly without resorting to the sufficient conditions discussed in Sections 4-5. The fundamental reason is that the ambiguity set \mathcal{P} here has a specific structure and its approximation \mathcal{P}_N is constructed in a way which allows us to apply the established Hoffman's Lemma to derive the convergence of \mathcal{P}_N to \mathcal{P} . Moreover, the linear structure of stochastic constraint $g(x, \xi) \leq 0$ enables us to derive a dual form of the robust chance constraint and continuity of v via dual formulation (6.14). With \mathcal{P}_N , we can evaluate the robust probability function $v_N(x)$ by solving linear programming problem (6.14), see Lemma 6.1.

6.3. Approximation of ambiguity set constructed through KL-divergence. 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 [27]. In practice, the true probability distribution may be unknown and therefore frequently one uses 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 the empirical distribution is atomic hence it is not absolutely continuous w.r.t. the Lebesgue measure. To get around the hurdle, we propose to estimate f_0 by so-called kernel density estimator (KDE) [39].

Let ξ_1, \dots, ξ_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 $\Phi(\cdot) \geq 0$, $\int_{\mathbb{R}^k} \Phi(\xi) d\xi = 1$. The KDE is defined as

$$f_N(z) = \frac{1}{Nh_N^k} \sum_{i=1}^N \Phi\left(\frac{z - \xi_i}{h_N}\right). \quad (6.21)$$

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

LEMMA 6.2. [14, Chapter 3, Theorem 1] Let $\Phi(\cdot)$ be a measurable kernel function satisfying $\Phi(\cdot) \geq 0$, $\int_{\mathbb{R}^k} \Phi(\xi) d\xi = 1$. Suppose that $\{h_N\}$ satisfies

$$\lim_{N \rightarrow \infty} h_N = 0, \quad \lim_{N \rightarrow \infty} Nh_N^k = \infty. \quad (6.22)$$

Then

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

Following the definition of the total variation metric and a discussion at page 29 in [7], we know that the convergence of f_N to f implies the convergence of the corresponding probability measures under the total variation metric.

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

$$\mathcal{P} = \{P \in \mathcal{P}(\mathbb{R}^k) : D_{KL}(f \| f_0) \leq d, f = dP/d\xi\}, \quad (6.23)$$

and

$$\mathcal{P}_N = \{P \in \mathcal{P}(\mathbb{R}^k) : D_{KL}(f \| f_N) \leq d, f = dP/d\xi\}. \quad (6.24)$$

It has been shown in [25, 26] that the robust chance constraint with the ambiguity set defined in (6.23) or (6.24) is equivalent to a classical chance constraint with a perturbed confidence level.

PROPOSITION 6.3. [25, 26] 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 corresponds to f_0 .

By Proposition 6.3, problems (MPDRCC) and (MPDRCC_N) can be equivalently written as

$$\begin{aligned} \min_{x \in X} & f(x) \\ \text{s.t.} & P_0(H(x)) \geq 1 - \beta', \end{aligned} \quad (6.25)$$

and

$$\begin{aligned} \min_{x \in X} & f(x) \\ \text{s.t.} & P_N(H(x)) \geq 1 - \beta'. \end{aligned} \quad (6.26)$$

The reformulations enable us to carry out stability analysis of (1.2) through (6.25) and (6.26). Therefore it suffices to establish convergence of P_N to P under the pseudo-metric through Lemma 5.1.

PROPOSITION 6.4. *Consider problems (MPDRCC) and (MPDRCC_N) with \mathcal{P} and \mathcal{P}_N being defined as in (6.23) and (6.24). Let f_N be defined as in (6.21). Assume: (a) the true density function f_0 is continuous, (b) $\{h_N\}$ satisfies condition (6.22), (c) $H(\cdot)$ is convex-valued and continuous over X . Then we have the following assertions.*

- (i) $\mathcal{H}(P, P_N) \rightarrow 0$ as N increases;
- (ii) if, in addition, condition (3.4) is fulfilled, then we can establish convergence of the optimal values and optimal solutions as specified in (3.5).

Proof. Part (i): We use Lemma 5.1 to prove the result. Let us verify conditions of the lemma. By Lemma 6.2, 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. the Lebesgue measure, which in turn ensures condition (4.1) with the convexity of $H(\cdot)$. Together with the continuity of $H(\cdot)$, all of the conditions of Lemma 5.1 are fulfilled.

Parts (ii): The conclusions follow immediately from Theorem 3.2. ■

REMARK 6.2. When KL-divergence in (6.23) or (6.24) is replaced by ϕ -divergence[5], Jiang and Guan establish similar results to Proposition 6.3 in [26, Theorem 1]. This implies that when the ambiguity set is defined through ϕ -divergences, 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 (6.21), then we can apply Lemma 5.1 and Proposition 6.4 to programs (6.25) and (6.26). As for numerical methods for solving the MPDRCC equipped with \mathcal{P}_N , we may end up by solving a chance constrained optimization problem (6.26).

7. Conclusion. This paper investigates convergence of the optimal value and the optimal solutions in an optimization problem with distributionally robust chance constraint when the ambiguity set of probability measures changes. Change of the ambiguity set may stem from increasing availability of information about the true probability distribution of the underlying uncertainty in the problem or numerical methods for solving the MPDRCC.

We began by establishing the main convergence results concerning the optimal value and the optimal solutions under general abstract conditions and then moving on to discuss in detail how these conditions may be fulfilled. One of the main theoretical issues to be tackled is the continuity of robust probability function. We present two sufficient conditions, one is well known in the literature of stochastic programming and the other is completely new. Another important issue to be dealt with is the uniform convergence of robust probability functions. Under the topology of weak convergence, we demonstrate how this may be achieved under some moderate conditions by using the theory of semi-convergence of the ambiguity set \mathcal{P}_N under the pseudo-metric.

As an application, we demonstrate how the established convergence results can be used to provide theoretical background for numerical methods for solving MPDRCC with general prior moment information (Section 6.1), MPDRCC with some specific moment information and empirical data (Section 6.2) and MPDRCC with the ambiguity set being constructed through KL-divergence (Section 6.3). While the approximation scheme in Section 6.1 is relatively new, the other reformulations in Sections 6.2-6.3 are built on the well-known results in the literature.

The work complements the mainstream research on MPDRCC in three ways: a better understanding of conditions required for the continuity of robust probability function, general convergence results, and some conceptual approximation schemes, all of which allow us to tackle a broader class of MPDRCCs from both theoretical and computational perspectives. Note that our convergence results are qualitative rather than quantitative, i.e., we are short of giving a quantitative description on various convergence results in relation to the change of the ambiguity set. Performance of the approximation schemes for solving MPDRCC when the ambiguity set takes a specific form also requires to be assessed through extensive numerical tests. We leave all these for our future research.

Acknowledgements. We would like to thank the Associate Editor, Professor D. Dentcheva, for effective handling of the review and two anonymous referees for their insightful comments which helped us significantly strengthen the paper.

8. Appendix. Proof of Lemma 2.1. Since $P(D_h) = 0$, it follows from the mapping theorem (see [7, Theorem 2.7]) that the weak convergence of P_N to P implies $P_N h^{-1}$ converges to $P h^{-1}$ weakly. On the other hand, by [8, Theorem 16.13],

$$\int_{\mathbb{R}} |y| P_N h^{-1}(dy) = \int_{\Xi} |h(x)| P_N(dx), \quad \int_{\mathbb{R}} |y| P h^{-1}(dy) = \int_{\Xi} |h(x)| P(dx).$$

For the simplicity of exposition, let $\tilde{P}_N = P_N h^{-1}$ and $\tilde{P} = P h^{-1}$. Then \tilde{P}_N converges to \tilde{P} weakly and condition (2.9) can be written as

$$\lim_{R \rightarrow \infty} \sup_{N \in \mathbb{N}} \int_{|y| \geq R} |y| \tilde{P}_N(dy) = 0. \quad (8.1)$$

Observe that the weak convergence of \tilde{P}_N to \tilde{P} implies $\tilde{P}_N(|y| > t) \rightarrow \tilde{P}(|y| > t)$ for all but countably many t . Moreover, it is well known (see i.e. [7, Theorem 3.4])

$$\int_{\mathbb{R}} |y| \tilde{P}(dy) = \int_0^\infty \tilde{P}(|y| > t) dt.$$

The rest of the proof is similar to [7, Theorem 3.5]. By Fatou's Lemma on the line,

$$\int_0^\infty \tilde{P}(|y| > t) dt \leq \liminf_{N \rightarrow \infty} \int_0^\infty \tilde{P}_N(|y| > t) dt = \liminf_{N \rightarrow \infty} \int_{\mathbb{R}} |y| \tilde{P}_N(dy).$$

By (8.1), the last term in the equation above is finite valued and this in turn ensures

$$\int_{\mathbb{R}} |y| \tilde{P}(dy) < \infty. \quad (8.2)$$

By partial integration,

$$\begin{aligned} \int_0^\infty y \tilde{P}(dy) &= \int_0^\alpha y \tilde{P}(dy) + \int_{y \geq \alpha} y \tilde{P}(dy) \\ &= \int_0^\alpha \tilde{P}(t < y < \alpha) dt + \int_{y \geq \alpha} y \tilde{P}(dy). \end{aligned} \quad (8.3)$$

Entailed by (8.2), for any positive number ϵ , the last term in the equation above is less than ϵ when α is sufficiently large. Likewise

$$\int_0^\infty y \tilde{P}_N(dy) = \int_0^\alpha \tilde{P}_N(t < y < \alpha) dt + \int_{y \geq \alpha} y \tilde{P}_N(dy). \quad (8.4)$$

By condition (8.1), the last term in the equation above is less than ϵ when α is sufficiently large. It is easy to show that the first term on the right of (8.4) converges to the corresponding part in (8.3). Since α can be chosen in such a way that $\tilde{P}(y = \alpha) = 0$ and ϵ can be arbitrarily small, it follows by the Lebesgue convergence theorem in the interval $[0, \alpha]$

$$\lim_{N \rightarrow \infty} \int_0^\infty y \tilde{P}_N(dy) = \int_0^\infty y \tilde{P}(dy).$$

Analogously we can show $\lim_{N \rightarrow \infty} \int_{-\infty}^0 y \tilde{P}_N(dy) = \int_{-\infty}^0 y \tilde{P}(dy)$. Putting them together, we obtain (2.8). The implication of (2.10) to (2.9) is obvious. ■

REFERENCES

- [1] W. van Ackooij, Chance constrained programming: with applications in Energy Management, PhD thesis, École Centrale Paris, 2013.
- [2] W. van Ackooij, Eventual convexity of chance constrained feasible sets, *Optimization*, 64:1263-1284, 2015.

- [3] W. van Ackooij and R. Henrion, Gradient formulae for nonlinear probabilistic constraints with Gaussian and Gaussian-like distributions, *SIAM J. Optim.*, 24:1864-1889, 2014.
- [4] K. B. Athreya and S. N. Lahiri, *Measure theory and probability theory*, Springer texts in statistics, Springer, New York, 2006.
- [5] 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: 341-357, 2013.
- [6] D. Bertsimas, V. Gupta and N. Kallus, Robust SAA, Preprint, arXiv:1408.4445, 2014.
- [7] P. Billingsley, *Convergence of probability measures*, Wiley, New York, 1999.
- [8] P. Billingsley, *Probability and measures*, Wiley, New York, 1995.
- [9] J. F. Bonnans and A. Shapiro, *Perturbation analysis of optimization problems*, Springer, New York, 2000.
- [10] G. Calafiore and L. El Ghaoui, Distributionally robust chance-constrained linear programs with applications, *J. Optim. Theory Appl.*, 130: 1-22, 2006.
- [11] 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.
- [12] E. Delage and Y. Ye, Distributionally robust optimization under moment uncertainty with application to data-driven problems, *Oper. Res.*, 58: 592-612, 2010.
- [13] D. Dentcheva and G. Martinez, Regularization methods for optimization problems with probabilistic constraints, *Math. Program.*, 138: 223-251, 2013.
- [14] L. Devroye and L. Györfi, *Nonparametric density estimation: The l_1 View*, John Wiley & Sons Inc, 1985.
- [15] R. Dudley, *Uniform central limit theorems*, Cambridge University Press, New York, 1999.
- [16] E. Erdoğan and G. Iyengar, Ambiguous chance constrained problems and robust optimization, *Math. Program.*, 107: 37-61, 2006.
- [17] P. M. Esfahani and D. Kuhn, Data-driven distributionally robust optimization using the wasserstein metric: performance guarantees and tractable reformulations, *Optim. Online*, 2015.
- [18] A. L. Gibbs and F. E. Su, On choosing and bounding probability metrics, *Internat. Statist. Rev.*, 70: 419-435, 2002.
- [19] A. Gut, *Probability: A Graduate Course*, Springer, New York, 2005.
- [20] G. A. Hanasusanto, V. Roitch, D. Kuhn and W. Wiesemann, A distributionally robust perspective on uncertainty quantification and chance constrained programming, *Math. Program.*, 151: 35-62, 2015.
- [21] R. Henrion, Optimierungsprobleme mit wahrscheinlichkeitsrestriktionen: Modelle, struktur, numerik. Lecture Notes, 2010.
- [22] R. Henrion and W. Römisch, Hölder and Lipschitz stability of solution sets in programs with probabilistic constraints, *Math. Program.*, 84: 589-611, 2004.
- [23] R. Henrion and W. Römisch, Metric regularity and quantitative stability in stochastic programs with probabilistic constraints, *Math. Program.*, 100: 55-88, 1999.
- [24] R. Henrion and C. Strugarek, Convexity of chance constraints with dependent random variables: the use of copulae. In M. Bertocchi, G. Consigli, and M.A.H. Dempster, editors, *Stochastic Optimization Methods in Finance and Energy: New Financial Products and Energy Market Strategies*, International Series in Operations Research and Management Science, 427-439, Springer, 2011.
- [25] Z. Hu and L. J. Hong, Kullback-Leibler divergence constrained distributionally robust optimization, *Optim. Online*, 2013.
- [26] R. Jiang and Y. Guan, Data-driven chance constrained stochastic program, *Math. Program.*, 158: 291-327, 2016.
- [27] S. Kullback and R. Leibler, On information and sufficiency, *Ann. Math. Statist.*, 22: 79-86, 1951.
- [28] J. Luedtke and S. Ahmed, A sample approximation approach for optimization with probabilistic constraints. *SIAM J. Optim.*, 19: 674-699, 2008.
- [29] Y. Liu and H. Xu, Entropic approximation for mathematical programs with robust equilibrium constraints, *SIAM J. Optim.*, 24: 933-958, 2014.
- [30] A. Nemirovski and A. Shapiro, Convex approximations of chance constrained programs, *SIAM J. Optim.*, 17: 969-996, 2006.
- [31] 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.
- [32] D. Pallaschke and R. Urbański, *Pairs of compact convex sets*, Kluwer, Dordrecht, 2002.
- [33] A. Prékopa, On probabilistic constrained programming, in: *Proceedings of the Princeton Symposium on Mathematical Programming*, Princeton University Press, Princeton, 113-138, 1970.
- [34] A. Prékopa, *Stochastic Programming*, Kluwer, Dordrecht, Boston, 1995.
- [35] A. Prékopa, Probabilistic programming, in: A. Ruszczyński and A. Shapiro, eds., *Stochastic Programming*, Elsevier, Amsterdam, 267-351, 2003.
- [36] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer-Verlag, Berlin, 1998.
- [37] W. Römisch, Stability of stochastic programming problems, in: A. Ruszczyński and A. Shapiro, eds., *Stochastic Programming*, Elsevier, Amsterdam, 483-554, 2003.
- [38] W. Römisch, R. Schultz, Stability analysis for stochastic programs, *Ann. Oper. Res.*, 30: 241-266, 1991.
- [39] M. Rosenblatt, Remarks on some nonparametric estimates of a density function, *Ann. Math. Stat.*, 27: 832-837, 1956.
- [40] 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.
- [41] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*, SIAM, Philadelphia, 2009.
- [42] H. Sun and H. Xu, Convergence analysis for distributionally robust optimization and equilibrium problems, *Math. Oper. Res.*, 41: 377-401, 2015.
- [43] J. Wolfowitz, Generalization of the theorem of Glivenko-cantelli, *Ann. Math. Statist.*, 25: 131-138, 1954.
- [44] 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.
- [45] H. Xu, Y. Liu and H. Sun, Distributionally robust optimization with matrix moment constraints: lagrange duality and cutting plane methods, *Optim. Online*, 2015.
- [46] W. Yang and H. Xu, Distributionally robust chance constraints for non-Linear uncertainties, *Math. Program.*, 155: 231-165, 2016.
- [47] J. Zhang, H. Xu and L. Zhang, Quantitative stability analysis of stochastic quasi-variational inequality problems and application, *Optim. Online*, 2015.
- [48] C. Zhao and Y. Guan, Data-driven risk-averse stochastic optimization with wasserstein metric, *Optim. Online*, 2015.

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