Asymptotic Consistency of Data-Driven Distributionally Robust Optimization via Reference-Distribution Convergence and Ambiguity-Set Shrinkage

Qiaoyi Wang*[1], Wei Liu†[2], Feiyu Chen[1], Minxin Shu[1], Zongxin Zhang[1], and Bo Yu‡[1]

[1]National Center for Applied Mathematics in Chongqing, Chongqing Normal University, Chongqing 401331, China
[2]Faculty of Arts and Sciences, Beijing Normal University at Zhuhai, Zhuhai 519000, Guangdong, China
[1]Emails: 2024110518099@stu.cqnu.edu.cn; liuwei@bnuz.edu.cn; fchen@cqnu.edu.cn; shuxin2357@gmail.com; 2484686122@qq.com; yubo@dlut.edu.cn.

Abstract

We study asymptotic consistency of data-driven distributionally robust optimization with shrinking ambiguity sets. The analysis separates reference-distribution convergence from ambiguity-set shrinkage on a prescribed test-function class. Under compactness and continuity assumptions, this yields uniform convergence of robust objectives, optimal-value convergence, and outer convergence of minimizers. For constrained DRO, the same mechanism gives uniform convergence of robust constraints and Painlevé–Kuratowski convergence of feasible regions under a Slater condition. We verify the assumptions for empirical reference measures and generic cost-based optimal-transport ambiguity sets, including the norm-cost case, and extend the framework to MADRO models with uncertain mixture weights.

Keywords: Distributionally robust optimization; data-driven optimization; statistical consistency; ambiguity-set shrinkage; Painlevé–Kuratowski convergence

---

[*] Qiaoyi Wang and Wei Liu contributed equally to this work.

[†] Qiaoyi Wang and Wei Liu contributed equally to this work.

[‡] Corresponding author.

1 Introduction

Distributionally robust optimization (DRO) replaces the unknown data-generating law by an ambiguity set to hedge against sampling error, model misspecification, and distributional shift. Major ambiguity models include moment-based sets [1, 2], ϕ-divergence sets [3], and Wasserstein or optimal-transport sets [4–9]. A basic asymptotic question is whether data-driven DRO converges to the population stochastic optimization problem as the ambiguity set shrinks, including convergence of robust objectives, constraints, feasible regions, optimal values, and minimizers.

Existing analyses are often tied to specific perturbation models. Stochastic programming stability studies perturbations of probability laws [10–14], while metric-specific DRO analyses exploit particular discrepancies and empirical concentration estimates [4, 5, 15]. Set-valued stability further studies convergence of distributional sets under specific metrics [16, 17]. This paper instead gives a test-function-based consistency principle that separates statistical convergence of the reference distribution from geometric shrinkage of the ambiguity set.

The contributions are as follows. First, reference convergence and ambiguity-set shrinkage on a prescribed test-function class are shown to imply uniform convergence of robust objective and constraint functionals. Second, this yields optimal-value convergence, Painlevé–Kuratowski outer convergence of minimizers, and, under a Slater condition, full Painlevé–Kuratowski convergence of constrained feasible regions. Third, the assumptions are verified for empirical reference measures and generic cost-based optimal-transport ambiguity sets, and the same mechanism is extended to Mixture Ambiguity Distributionally Robust Optimization (MADRO) with uncertain mixture weights. The remainder of the paper is organized as follows. Section 2 introduces the framework. Sections 3 and 4 establish unconstrained and constrained consistency. Section 5 verifies the assumptions. Section 6 extends the framework to MADRO. Section 7 concludes.

2 Data-Driven DRO Framework

Let $\Xi \subset \mathbb{R}^n$ be closed with Borel σ-algebra $\mathcal{B}(\Xi)$, and let $P^* \in \mathscr{P}(\Xi)$ be the target law. Samples are realized on $\big(\Xi^\infty, \mathcal{B}(\Xi)^{\otimes\infty}, (P^*)^\infty\big)$. All convergence statements are pathwise on probability-one events; after fixing a sample path, P_N^{ref} and $\mathcal{P}_N$ are deterministic, and measurability of value or solution-set mappings is not considered. For each N, let $P_N^{\mathrm{ref}} \in \mathscr{P}(\Xi)$ be the reference law and $\mathcal{P}_N \subset \mathscr{P}(\Xi)$ the ambiguity set. We use a prescribed test-function class $\mathcal{F}$ to measure reference convergence and ambiguity-set shrinkage.

Assumption 2.1 (Decision set and objective function regularity)**.** The decision set $\mathcal{X} \subset \mathbb{R}^d$ is nonempty, convex, and compact. The loss function $f : \mathcal{X} \times \Xi \to \mathbb{R}$ satisfies:

(i) For every $x \in \mathcal{X}$, the mapping $\xi \mapsto f(x, \xi)$ belongs to the test-function class $\mathcal{F}$.

(ii) There exists a nondecreasing continuous function $\rho_f : [0, \infty) \to [0, \infty)$, with $\rho_f(0) = 0$, such that $|f(x, \xi) - f(x', \xi)| \le \rho_f(|x - x'|)$ for all $x, x' \in \mathcal{X}$ and $\xi \in \Xi$.

Assumption 2.2 (Pointwise integrability on the test class)**.** For every $\psi \in \mathcal{F}$, there is a probability-one event A^I_ψ such that, on A^I_ψ, $\int_\Xi |\psi| \, dP^* < \infty$, $\int_\Xi |\psi| \, dP^{\mathrm{ref}}_N < \infty$ for all N, and $\int_\Xi |\psi| \, dP < \infty$ for all N and all $P \in \mathcal{P}_N$.

Assumption 2.3 (Reference-distribution convergence on the test-function class)**.** For every $\psi \in \mathcal{F}$, there is a probability-one event A^R_ψ such that, on A^R_ψ, $\int_\Xi \psi \, dP^{\mathrm{ref}}_N \to \int_\Xi \psi \, dP^*$.

Assumption 2.4 (Sequential ambiguity-set shrinkage toward the reference distribution)**.** For every $\psi \in \mathcal{F}$, there is a probability-one event A^C_ψ such that, on A^C_ψ, every sequence $P_N \in \mathcal{P}_N$ satisfies $\int_\Xi \psi \, dP_N - \int_\Xi \psi \, dP^{\mathrm{ref}}_N \to 0$.

Assumption 2.5 (Nonemptiness of ambiguity sets)**.** There exists an event $A^\emptyset \in \mathcal{A}$ with $\mathbb{P}(A^\emptyset) = 1$ such that, on $A^\emptyset$, $\mathcal{P}_N \neq \emptyset$ for every N.

Assumption 2.6 (Eventual finite-valuedness of robust functionals)**.** There exists an event $A^F \in \mathcal{A}$ with $\mathbb{P}(A^F) = 1$ such that, on A^F, the robust objective $\mathcal{R}_N(f)$ is finite-valued on $\mathcal{X}$ for all sufficiently large N. In the constrained case, each robust constraint functional $\Phi_{N,j}$ is finite-valued on $\mathcal{X}$ for all sufficiently large N.

Under nonemptiness of $\mathcal{P}_N$, the sequential condition in Assumption 2.4 is equivalent to fixed-test-function uniform shrinkage, as shown later in Lemma 3.2. These assumptions make the ambiguity sets asymptotically indistinguishable from the target distribution on the test-function class used by the optimization model, which is the key input for the unconstrained consistency analysis below.

3 Unconstrained Asymptotic Consistency

We next convert the distributional assumptions into uniform convergence of the robust objective. Under Assumptions 2.1–2.6, we define the robust objective functional $\mathcal{R}_N(f)(x) := \sup_{P \in \mathcal{P}_N} \int_\Xi f(x, \xi) \, dP(\xi)$ and the population objective $\mathcal{R}_\infty(f)(x) := \int_\Xi f(x, \xi) \, dP^\circ(\xi)$.

Lemma 3.1 (Test-function convergence of ambiguity-set selections to the target distribution)**.** *Suppose Assumptions 2.2, 2.3, 2.4, and 2.5 hold. For every $\psi \in \mathcal{F}$, there exists an event $A_\psi := A^I_\psi \cap A^R_\psi \cap A^C_\psi \cap A^\emptyset$ with $\mathbb{P}(A_\psi) = 1$ such that, on A_ψ, for every sequence $P_N \in \mathcal{P}_N$, $\int_\Xi \psi \, dP_N \to \int_\Xi \psi \, dP^\circ$.*

Proof. See Appendix A. □

Lemma 3.2 (Sequential shrinkage implies uniform shrinkage for a fixed test function)**.** *Suppose Assumptions 2.2, 2.4, and 2.5 hold. Fix $\psi \in \mathcal{F}$. On the event $A^I_\psi \cap A^C_\psi \cap A^\emptyset$, one has $\Delta_N(\psi) := \sup_{P \in \mathcal{P}_N} |\int_\Xi \psi \, dP - \int_\Xi \psi \, dP^{\mathrm{ref}}_N| \to 0$.*

Proof. See Appendix A. □

3.1 Uniform Convergence of Robust Objective Functionals

Lemma 3.3 (Inherited modulus continuity of finite-valued robust objectives)**.** *Under Assumptions 2.1 and 2.2:*

(i) If, for some N, $\mathcal{R}_N(f)$ is finite-valued on $\mathcal{X}$, then $|\mathcal{R}_N(f)(x) - \mathcal{R}_N(f)(x')| \leq \rho_f(|x - x'|)$ for all $x, x' \in \mathcal{X}$.

(ii) The population objective $\mathcal{R}_\infty(f)$ satisfies $|\mathcal{R}_\infty(f)(x) - \mathcal{R}_\infty(f)(x')| \leq \rho_f(|x - x'|)$ for all $x, x' \in \mathcal{X}$.

Proof. See Appendix A. ☐

Lemma 3.4 (Convergence of robust objectives at fixed decisions)**.** *Under Assumptions 2.1–2.6, for every $x \in \mathcal{X}$, there exists an event A_x with $\mathbb{P}(A_x) = 1$ such that, on A_x, $\mathcal{R}_N(f)(x) \to \mathcal{R}_\infty(f)(x)$.*

Proof. See Appendix A. ☐

Pointwise convergence alone is not sufficient for stability of minimizers; the compactness and uniform modulus assumptions are used next to obtain uniform convergence.

Theorem 3.5 (Uniform convergence of robust objective functionals)**.** *Under Assumptions 2.1–2.6, there exists an event A_f with $\mathbb{P}(A_f) = 1$ such that, on A_f,*

$$\sup_{x \in \mathcal{X}} |\mathcal{R}_N(f)(x) - \mathcal{R}_\infty(f)(x)| \to 0.$$

Proof. See Appendix A. ☐

Corollary 3.6 (Epigraphical convergence of robust objectives)**.** *Under the assumptions of Theorem 3.5, there exists an event A_f with $\mathbb{P}(A_f) = 1$ such that, on A_f, $\operatorname{epi}_{\mathcal{X}} \mathcal{R}_N(f) \xrightarrow{\text{PK}} \operatorname{epi}_{\mathcal{X}} \mathcal{R}_\infty(f)$, where $\operatorname{epi}_{\mathcal{X}} h := \{(x, \alpha) \in \mathcal{X} \times \mathbb{R} : \ h(x) \leq \alpha\}$.*

Proof. See Appendix A. ☐

3.2 Optimal Values and Minimizer Sets in Unconstrained DRO

Define the unconstrained optimal values and minimizer sets by $\hat{v}_N := \inf_{x \in \mathcal{X}} \mathcal{R}_N(f)(x)$, $v^* := \inf_{x \in \mathcal{X}} \mathcal{R}_\infty(f)(x)$, $\hat{X}_N := \arg\min_{x \in \mathcal{X}} \mathcal{R}_N(f)(x)$, and $X^* := \arg\min_{x \in \mathcal{X}} \mathcal{R}_\infty(f)(x)$.

Theorem 3.7 (Optimal-value and minimizer consistency for unconstrained DRO)**.** *Under the assumptions of Theorem 3.5, on the event A_f, $\hat{v}_N \to v^*$, and in the Painlevé–Kuratowski sense,*

$$\operatorname*{Limsup}_{N \to \infty} \hat{X}_N \subseteq X^*.$$

Proof. See Appendix A. ☐

4 Constrained Asymptotic Consistency

The preceding results establish the basic consistency mechanism for unconstrained DRO. We next show that the same mechanism also controls robust constraints and feasible-set convergence. To analyze the asymptotic behavior of feasible regions and minimizer sets, we use set convergence in the sense of Painlevé–Kuratowski; see [18, Chapter 4].

Definition 4.1 (Painlevé–Kuratowski set convergence)**.** The inner (lower) limit of $\{C_N\}$ is $\operatorname{Liminf}_{N\to\infty} C_N = \{x \in \mathcal{X} : \limsup_{N\to\infty} \operatorname{dist}(x, C_N) = 0\}$, and the outer (upper) limit is $\operatorname{Limsup}_{N\to\infty} C_N = \{x \in \mathcal{X} : \liminf_{N\to\infty} \operatorname{dist}(x, C_N) = 0\}$. We write $C_N \xrightarrow{\mathrm{PK}} C$. if $\operatorname{Limsup}_{N\to\infty} C_N \subseteq C \subseteq \operatorname{Liminf}_{N\to\infty} C_N$.

Let $g_j(x,\xi)$, $j = 1,\dots,J$, be a finite family of constraint functions. Define $\Phi_{\infty,j}(x) := \mathcal{R}_\infty(g_j)(x)$ and $\Phi_{N,j}(x) := \mathcal{R}_N(g_j)(x)$.

Assumption 4.2 (Constraint regularity and Slater condition)**.** Let $\mathcal{X} \subset \mathbb{R}^d$ be nonempty, compact, and convex. For each $j = 1,\dots,J$, let $g_j \colon \mathcal{X} \times \Xi \to \mathbb{R}$ satisfy:

(i) For every $x \in \mathcal{X}$, the mapping $\xi \mapsto g_j(x,\xi)$ belongs to $\mathcal{F}$.

(ii) For every $\xi \in \Xi$, the mapping $x \mapsto g_j(x,\xi)$ is convex on $\mathcal{X}$.

(iii) There exists a nondecreasing continuous function $\rho_{g_j} : [0,\infty) \to [0,\infty)$, with $\rho_{g_j}(0) = 0$, such that $|g_j(x,\xi) - g_j(x',\xi)| \le \rho_{g_j}(|x - x'|)$ for all $x, x' \in \mathcal{X}$ and $\xi \in \Xi$.

Furthermore, there exists $x^\circ \in \mathcal{X}$ such that $\Phi_{\infty,j}(x^\circ) < 0$ for $j = 1,\dots,J$.

Since $x \mapsto g_j(x,\xi)$ is convex for every ξ, the population functional $\Phi_{\infty,j}(x) = \int_\Xi g_j(x,\xi)\, dP^\circ(\xi)$ is convex on $\mathcal{X}$. Furthermore, since integration against a non-negative measure and taking suprema preserve convexity, the empirical robust functionals $\Phi_{N,j}$ are also convex on $\mathcal{X}$ wherever they are finite-valued.

Define the population and empirical feasible regions by

$$\Gamma_\infty := \bigcap_{j=1}^{J} \{x \in \mathcal{X} : \Phi_{\infty,j}(x) \le 0\}, \qquad \Gamma_N := \bigcap_{j=1}^{J} \{x \in \mathcal{X} : \Phi_{N,j}(x) \le 0\}.$$

Define the corresponding constrained optimal values and minimizer sets: $v^c_\infty := \inf_{x\in\Gamma_\infty} \mathcal{R}_\infty(f)(x)$, $\hat{v}^c_N := \inf_{x\in\Gamma_N} \mathcal{R}_N(f)(x)$, $X^*_c := \arg\min_{x\in\Gamma_\infty} \mathcal{R}_\infty(f)(x)$, and $\hat{X}^c_N := \arg\min_{x\in\Gamma_N} \mathcal{R}_N(f)(x)$.

Lemma 4.3 (Uniform convergence of robust constraint functionals)**.** *Under Assumptions 2.2–2.6 and 4.2, there exists an event* A_g *with* $\mathbb{P}(A_g) = 1$ *such that, on* A_g, $\max_{1\le j\le J} \sup_{x\in\mathcal{X}} |\Phi_{N,j}(x) - \Phi_{\infty,j}(x)| \to 0$.

Proof. See Appendix B. □

The following theorem is the set-convergence step needed to transfer objective convergence to constrained optimal values.

Theorem 4.4 (Painlevé–Kuratowski convergence of robust feasible regions)**.** *Under Assumption 4.2 and the assumptions of Lemma 4.3, on the event* A_g*,*

$$\Gamma_N \xrightarrow{\mathrm{PK}} \Gamma_\infty.$$

Proof. See Appendix B. □

Theorem 4.5 (Constrained optimal-value and minimizer consistency)**.** *Under the assumptions of Theorem 3.5 for the objective and the assumptions of Theorem 4.4 for the feasible regions, on* $A_* := A_f \cap A_g$*,* $\hat{v}_N^c \to v_\infty^c$ *and*

$$\operatorname*{Limsup}_{N\to\infty} \hat{X}_N^c \subseteq X_c^*.$$

Proof. See Appendix B. □

The abstract assumptions used so far are high-level; the next section verifies them through empirical reference convergence and cost-based optimal-transport shrinkage.

5 Verification of Assumptions

We now verify the two distributional inputs through empirical reference convergence and OT_c-ambiguity-set shrinkage.

Proposition 5.1 (Summable deviation implies reference-distribution convergence)**.** *Let* $\mathbb{D}$ *be a metric on* $\mathscr{P}(\Xi)$ *such that* $\mathbb{D}(P_n, P^\circ) \to 0$ *implies* $\int_\Xi \psi\, dP_n \to \int_\Xi \psi\, dP^\circ$ *for every* $\psi \in \mathcal{F}$*. Suppose that for every* $\epsilon > 0$*,* $\sum_{N=1}^\infty \mathbb{P}(\mathbb{D}(P_N^{\mathrm{ref}}, P^\circ) > \epsilon) < \infty$*. Then* $\mathbb{D}(P_N^{\mathrm{ref}}, P^\circ) \to 0$ *almost surely, and hence Assumption 2.3 holds.*

Proof. See Appendix C. □

This condition can be checked by standard concentration inequalities for empirical measures under suitable moment or tail assumptions; see, for example, [15].

Proposition 5.2 (OT_c-ambiguity-set shrinkage)**.** *Let* $c : \Xi \times \Xi \to [0, \infty]$ *be a measurable transportation cost satisfying* $c(\xi, \xi) = 0$ *for all* $\xi \in \Xi$*. Define* $\mathrm{OT}_c(P, Q) := \inf_{\gamma \in \Gamma(P,Q)} \int_{\Xi\times\Xi} c(\xi, \zeta)\, d\gamma(\xi, \zeta)$*, where* $\Gamma(P, Q)$ *denotes the set of couplings of* P *and* Q*; see [19]. Let* $\delta_N \downarrow 0$ *and set* $\mathcal{P}_N^c := \{P \in \mathscr{P}(\Xi) : \mathrm{OT}_c(P, P_N^{\mathrm{ref}}) \le \delta_N\}$*. Suppose that every* $\psi \in \mathcal{F}$ *is c-Lipschitz, that is, there exists* $L_\psi < \infty$ *such that* $|\psi(\xi) - \psi(\zeta)| \le L_\psi c(\xi, \zeta)$ *for all* $\xi, \zeta \in \Xi$*. Then, for every* $\psi \in \mathcal{F}$*,*

$$\sup_{P \in \mathcal{P}_N^c} \left| \int_\Xi \psi\, dP - \int_\Xi \psi\, dP_N^{\mathrm{ref}} \right| \le L_\psi \delta_N \to 0.$$

Consequently, Assumptions 2.4 and 2.5 hold.

Proof. See Appendix C. □

In particular, when $c(\xi, \zeta) = \|\xi - \zeta\|$, the c-Lipschitz condition reduces to ordinary Lipschitz continuity with respect to the ambient norm.

Together, Propositions 5.1 and 5.2 verify the two distributional inputs required by the abstract consistency theorems: reference convergence and ambiguity-set shrinkage. We next apply the same mechanism to MADRO.

6 MADRO Extension

Having verified the distributional inputs for transportation-cost ambiguity sets, we now apply the same mechanism to mixture ambiguity sets with uncertain component distributions and uncertain weights.

Let $\mathcal{P}_N^{MA}$ be the mixed ambiguity set, and $P_N^{MA,\mathrm{ref}}$ the mixed reference distribution. The robust objective is evaluated using the componentwise robust values $v_{N,k}(x)$ and the outer weight aggregation formula: $\sup_{\boldsymbol{w} \in \mathcal{W}_N} \sum_{k=1}^K w_k v_{N,k}(x)$. Componentwise shrinkage can be verified, for example, by taking $\mathcal{P}_{N,k} = \{P : \mathrm{OT}_{c_k}(P, P_{N,k}^{\mathrm{ref}}) \leq \delta_{k,N}\}$, where $\delta_{k,N} \downarrow 0$, $c_k(\xi, \xi) = 0$, and the relevant componentwise test functions are c_k-Lipschitz. For fixed componentwise robust values $v_{N,k}(x)$, the outer maximization over uncertain mixture weights admits a finite-dimensional LP dual reformulation. A complete deterministic reformulation is obtained once the componentwise values $v_{N,k}(x)$ admit tractable dual representations. As derived in Appendix D, the final LP dual reformulation is:

$$\inf_{x \in \mathcal{X}} \inf_{\eta, \nu, \boldsymbol{t}} \left\{ \eta + \nu(\delta_{w,N} - 1) + \sum_{k=1}^K \hat{w}_{N,k} t_k : \begin{array}{l} t_k + \eta - \nu \geq v_{N,k}(x), \quad k = 1, \ldots, K, \\ 0 \leq t_k \leq 2\nu, \quad k = 1, \ldots, K, \\ \nu \geq 0, \quad \eta \in \mathbb{R} \end{array} \right\}. \tag{1}$$

The consistency analysis mirrors the abstract framework: componentwise reference convergence and componentwise shrinkage control the distributions, while shrinkage of the weight set controls the mixture coefficients.

Assumption 6.1 (Componentwise reference-distribution convergence)**.** For every $k = 1, \ldots, K$ and every $\psi \in \mathcal{F}$, there exists an event $A_{\psi,k}^R$ with $\mathbb{P}(A_{\psi,k}^R) = 1$ such that, on $A_{\psi,k}^R$, $\int_\Xi \psi \, dP_{N,k}^{\mathrm{ref}} \to \int_\Xi \psi \, dP_k^\circ$.

Assumption 6.2 (Componentwise ambiguity-set shrinkage)**.** For every $k = 1, \ldots, K$ and every $\psi \in \mathcal{F}$, there exists an event $A_{\psi,k}^C$ with $\mathbb{P}(A_{\psi,k}^C) = 1$ such that, on $A_{\psi,k}^C$, $\mathcal{P}_{N,k} \neq \emptyset$ for every N, and $\sup_{P_k \in \mathcal{P}_{N,k}} | \int_\Xi \psi \, dP_k - \int_\Xi \psi \, dP_{N,k}^{\mathrm{ref}} | \to 0$.

Assumption 6.3 (Weight ambiguity shrinkage)**.** There exists an event A^W with $\mathbb{P}(A^W) = 1$ such that, on A^W, $\mathcal{W}_N \neq \emptyset$, $\hat{\boldsymbol{w}}_N \in \mathcal{W}_N$, $\sup_{\boldsymbol{w}_N \in \mathcal{W}_N} \|\boldsymbol{w}_N - \hat{\boldsymbol{w}}_N\|_1 \to 0$, and $\hat{\boldsymbol{w}}_N \to \boldsymbol{w}^\circ$.

Lemma 6.4 (Reference-distribution convergence of the mixed reference distribution)**.** *Under Assumptions 6.1 and 6.3, for every $\psi \in \mathcal{F}$, on an event $A_\psi^{MA,R}$, $\int_\Xi \psi \, dP_N^{MA,\mathrm{ref}} \to \int_\Xi \psi \, dP^{MA,\circ}$, where $P^{MA,\circ} := \sum_{k=1}^K w_k^\circ P_k^\circ$.*

Proof. See Appendix D. □

Theorem 6.5 (Ambiguity-set shrinkage of the mixed ambiguity set)**.** *Under Assumptions 6.1–6.3, for every* $\psi \in \mathcal{F}$*, on an event* $A_\psi^{MA,C}$, $\sup_{P^{MA}\in\mathcal{P}_N^{MA}} |\int_\Xi \psi\, dP^{MA} - \int_\Xi \psi\, dP_N^{MA,\mathrm{ref}}| \to 0$.

Proof. See Appendix D. □

Corollary 6.6 (Consistency of MADRO)**.** *Under Assumptions 2.1, 2.2, 2.5, 2.6, and 6.1–6.3, the mixed reference distribution satisfies Assumption 2.3, and the mixed ambiguity set satisfies Assumptions 2.4 and 2.5. Consequently, Theorems 3.5, 3.7, 4.4, and 4.5 yield uniform convergence of MADRO robust objectives, optimal-value convergence, feasible-region PK convergence under the Slater condition, and PK outer convergence of minimizer sets.*

7 Conclusion

We developed a test-function-based consistency framework for data-driven DRO with shrinking ambiguity sets. By separating reference-distribution convergence from ambiguity-set shrinkage, the framework yields uniform convergence of robust objectives and constraints, optimal-value convergence, feasible-region convergence under Slater conditions, and outer convergence of minimizers. The assumptions were verified for empirical reference measures and OT_c ambiguity sets and extended to MADRO with uncertain mixture weights.

A Proofs for Unconstrained DRO Consistency

Proof of Lemma 3.1. Fix $\psi \in \mathcal{F}$ and define $A_\psi := A_\psi^I \cap A_\psi^R \cap A_\psi^C \cap A^\emptyset$. Then $\mathbb{P}(A_\psi) = 1$. For every $\omega \in A_\psi$ and every sequence $P_N \in \mathcal{P}_N(\omega)$, we decompose $\int_\Xi \psi\, dP_N - \int_\Xi \psi\, dP^\circ$ as

$$\left(\int_\Xi \psi\, dP_N - \int_\Xi \psi\, dP_N^{\mathrm{ref}}\right) + \left(\int_\Xi \psi\, dP_N^{\mathrm{ref}} - \int_\Xi \psi\, dP^\circ\right).$$

The first term converges to zero by Assumption 2.4, and the second term converges to zero by Assumption 2.3. □

Proof of Lemma 3.2. Fix $\omega \in A_\psi^I \cap A_\psi^C \cap A^\emptyset$. We suppress ω from the notation. By Assumption 2.2, all integrals in the definition of $\Delta_N(\psi)$ are finite real numbers for individual $P \in \mathcal{P}_N$, although the supremum may a priori be $+\infty$. Suppose for contradiction that $\Delta_N(\psi) \not\to 0$. Then there exists $\epsilon > 0$ and a subsequence $N_k \to \infty$ such that $\Delta_{N_k}(\psi) > \epsilon$. For each k, we can choose $P_{N_k} \in \mathcal{P}_{N_k}$ such that

$$\left|\int_\Xi \psi\, dP_{N_k} - \int_\Xi \psi\, dP_{N_k}^{\mathrm{ref}}\right| > \epsilon.$$

However, completing the sequence P_{N_k} with arbitrary elements from $\mathcal{P}_N$ for indices N not in the subsequence (which is possible since $\mathcal{P}_N \neq \emptyset$ by Assumption 2.5), Assumption 2.4 implies that for the full sequence P_N,

$$\left| \int_{\Xi} \psi \, dP_N - \int_{\Xi} \psi \, dP_N^{\mathrm{ref}} \right| \longrightarrow 0.$$

This contradicts the inequality along the subsequence N_k. Hence $\Delta_N(\psi) \to 0$. $\square$

Proof of Lemma 3.3. Fix $x, x' \in \mathcal{X}$. For every $P \in \mathcal{P}_N$, Assumption 2.1(ii) gives

$$\int_{\Xi} f(x,\xi) \, dP(\xi) \leq \int_{\Xi} f(x',\xi) \, dP(\xi) + \rho_f(|x - x'|).$$

Taking the supremum over $P \in \mathcal{P}_N$ gives

$$\mathcal{R}_N(f)(x) \leq \mathcal{R}_N(f)(x') + \rho_f(|x - x'|).$$

Interchanging x and x' yields

$$\mathcal{R}_N(f)(x') \leq \mathcal{R}_N(f)(x) + \rho_f(|x - x'|).$$

Hence

$$|\mathcal{R}_N(f)(x) - \mathcal{R}_N(f)(x')| \leq \rho_f(|x - x'|).$$

The proof for $\mathcal{R}_\infty(f)$ is identical, with P° in place of P. The required integrals are finite by Assumption 2.2, since $f(x,\cdot) \in \mathcal{F}$ for all $x \in \mathcal{X}$. $\square$

Proof of Lemma 3.4. Fix $x \in \mathcal{X}$ and set $\psi_x(\xi) := f(x,\xi)$. By Assumption 2.1, $\psi_x \in \mathcal{F}$. Let

$$A_x := A^I_{\psi_x} \cap A^R_{\psi_x} \cap A^C_{\psi_x} \cap A^{\emptyset} \cap A^F.$$

Then $\mathbb{P}(A_x) = 1$. Fix $\omega \in A_x$, and suppress ω from the notation. By Assumption 2.6, $\mathcal{R}_N(f)(x)$ is finite for large N. Thus,

$$\begin{aligned} |\mathcal{R}_N(f)(x) - \mathcal{R}_\infty(f)(x)| \leq {} & \sup_{P \in \mathcal{P}_N} \left| \int_{\Xi} f(x,\xi) \, dP(\xi) - \int_{\Xi} f(x,\xi) \, dP_N^{\mathrm{ref}}(\xi) \right| \\ & + \left| \int_{\Xi} f(x,\xi) \, dP_N^{\mathrm{ref}}(\xi) - \int_{\Xi} f(x,\xi) \, dP^\circ(\xi) \right|. \end{aligned}$$

The first term tends to zero by Lemma 3.2, and the second term tends to zero by Assumption 2.3. Hence $\mathcal{R}_N(f)(x) \to \mathcal{R}_\infty(f)(x)$. $\square$

Proof of Theorem 3.5. Since $\mathcal{X}$ is compact metric, it is separable. Fix a countable dense subset $D = \{x_m\}_{m \in \mathbb{N}} \subset \mathcal{X}$. For each m, by Lemma 3.4, there exists an event A_{x_m} with

$\mathbb{P}(A_{x_m}) = 1$ such that $\mathcal{R}_N(f)(x_m) \to \mathcal{R}_\infty(f)(x_m)$ on A_{x_m}. Define

$$A_f := A^\emptyset \cap A^F \cap \bigcap_{m=1}^{\infty} A_{x_m}.$$

Then $\mathbb{P}(A_f) = 1$. Fix $\omega \in A_f$, and suppress ω from the notation. By Assumption 2.6, for all sufficiently large N, $\mathcal{R}_N(f)$ is finite-valued on all of $\mathcal{X}$. For these sufficiently large N, Lemma 3.3 applies, and $\mathcal{R}_N(f)$ and $\mathcal{R}_\infty(f)$ share the modulus ρ_f. Given $\epsilon > 0$, choose $\delta > 0$ such that $\rho_f(\delta) < \epsilon/4$. Since $\mathcal{X}$ is compact, the open balls $B(x_m, \delta)$ cover $\mathcal{X}$. Extract a finite subcover corresponding to indices $m_1, \ldots, m_L$. Since $\mathcal{R}_N(f)(x_{m_\ell}) \to \mathcal{R}_\infty(f)(x_{m_\ell})$ for each $\ell \in \{1, \ldots, L\}$, there exists N_0 such that for all $N \geq N_0$ and all $\ell \in \{1, \ldots, L\}$,

$$|\mathcal{R}_N(f)(x_{m_\ell}) - \mathcal{R}_\infty(f)(x_{m_\ell})| < \epsilon/2.$$

For any $x \in \mathcal{X}$, there exists ℓ such that $|x - x_{m_\ell}| < \delta$. Then for $N \geq N_0$,

$$\begin{aligned} |\mathcal{R}_N(f)(x) - \mathcal{R}_\infty(f)(x)| &\leq |\mathcal{R}_N(f)(x) - \mathcal{R}_N(f)(x_{m_\ell})| \\ &\quad + |\mathcal{R}_N(f)(x_{m_\ell}) - \mathcal{R}_\infty(f)(x_{m_\ell})| \\ &\quad + |\mathcal{R}_\infty(f)(x_{m_\ell}) - \mathcal{R}_\infty(f)(x)| \\ &\leq 2\rho_f(\delta) + \epsilon/2 < \epsilon. \end{aligned}$$

Taking the supremum over $x \in \mathcal{X}$ yields

$$\sup_{x \in \mathcal{X}} |\mathcal{R}_N(f)(x) - \mathcal{R}_\infty(f)(x)| \to 0.$$

□

Proof of Corollary 3.6. By Theorem 3.5, on A_f,

$$\sup_{x \in \mathcal{X}} |\mathcal{R}_N(f)(x) - \mathcal{R}_\infty(f)(x)| \to 0.$$

Extend $\mathcal{R}_N(f)$ and $\mathcal{R}_\infty(f)$ to the ambient Euclidean space by setting them equal to $+\infty$ outside $\mathcal{X}$. Since $\mathcal{X}$ is compact, uniform convergence on $\mathcal{X}$ yields epi-convergence of these extended functions [18, Chapter 7]. In particular, $\mathcal{R}_N(f) \xrightarrow{\text{epi}} \mathcal{R}_\infty(f)$. By the definition of epi-convergence, this is equivalent to the Painlevé–Kuratowski convergence of the epigraphs:

$$\operatorname{epi}_{\mathcal{X}} \mathcal{R}_N(f) \xrightarrow{\text{PK}} \operatorname{epi}_{\mathcal{X}} \mathcal{R}_\infty(f).$$

□

Proof of Theorem 3.7. For every $x \in \mathcal{X}$, the uniform error bound implies

$$\mathcal{R}_\infty(f)(x) - \sup_{z \in \mathcal{X}} |\mathcal{R}_N(f)(z) - \mathcal{R}_\infty(f)(z)| \leq \mathcal{R}_N(f)(x) \leq \mathcal{R}_\infty(f)(x) + \sup_{z \in \mathcal{X}} |\mathcal{R}_N(f)(z) - \mathcal{R}_\infty(f)(z)|.$$

Taking the infimum over $x \in \mathcal{X}$ yields

$$v^* - \sup_{z\in\mathcal{X}} |\mathcal{R}_N(f)(z) - \mathcal{R}_\infty(f)(z)| \leq \hat{v}_N \leq v^* + \sup_{z\in\mathcal{X}} |\mathcal{R}_N(f)(z) - \mathcal{R}_\infty(f)(z)|.$$

By Theorem 3.5, the right-hand side converges to 0, and therefore $\hat{v}_N \to v^*$. We next prove the outer inclusion for minimizer sets. By Theorem 3.5 and the eventual finite-valuedness proved via Assumption 2.6, $\mathcal{R}_N(f)$ is finite-valued and continuous on $\mathcal{X}$ for all sufficiently large N. Hence $\hat{X}_N \neq \emptyset$ for all sufficiently large N. Also, $X^* \neq \emptyset$ by compactness of $\mathcal{X}$ and continuity of $\mathcal{R}_\infty(f)$. Discarding finitely many indices, let $\{N_k\}$ be an arbitrary subsequence and let $x_{N_k} \in \hat{X}_{N_k}$. By compactness of $\mathcal{X}$, passing to a further subsequence if necessary, we may assume that $x_{N_k} \to \bar{x} \in \mathcal{X}$. Choose any $x^* \in X^*$. Since $x_{N_k} \in \hat{X}_{N_k}$, we have $\mathcal{R}_{N_k}(f)(x_{N_k}) = \hat{v}_{N_k} \leq \mathcal{R}_{N_k}(f)(x^*)$. Hence we can bound the difference as

$$\begin{aligned}\mathcal{R}_\infty(f)(x_{N_k}) - \mathcal{R}_\infty(f)(x^*) &\leq |\mathcal{R}_\infty(f)(x_{N_k}) - \mathcal{R}_{N_k}(f)(x_{N_k})| + |\mathcal{R}_{N_k}(f)(x^*) - \mathcal{R}_\infty(f)(x^*)| \\ &\leq 2\sup_{z\in\mathcal{X}} \left|\mathcal{R}_{N_k}(f)(z) - \mathcal{R}_\infty(f)(z)\right|.\end{aligned}$$

Taking the limit superior gives $\limsup_{k\to\infty}\big(\mathcal{R}_\infty(f)(x_{N_k}) - \mathcal{R}_\infty(f)(x^*)\big) \leq 0$. By continuity of $\mathcal{R}_\infty(f)$, passing to the limit along $x_{N_k} \to \bar{x}$ yields $\mathcal{R}_\infty(f)(\bar{x}) \leq \mathcal{R}_\infty(f)(x^*) = v^*$. Since v^* is the infimum, $\mathcal{R}_\infty(f)(\bar{x}) = v^*$, which proves that $\bar{x} \in X^*$. Since the subsequence and cluster point were arbitrary, we conclude that $\operatorname{Limsup}_{N\to\infty} \hat{X}_N \subseteq X^*$. $\square$

B Proofs for Constrained DRO Consistency

Proof of Lemma 4.3. Fix $j \in \{1, \ldots, J\}$. Since $\mathcal{X}$ is compact metric, let $D = \{x_m\}_{m\in\mathbb{N}}$ be a countable dense subset of $\mathcal{X}$. For each m, set

$$\psi_{j,m}(\xi) := g_j(x_m, \xi).$$

By the constraint regularity assumption, $\psi_{j,m} \in \mathcal{F}$. Define

$$A_{g,j} := \bigcap_{m=1}^{\infty} \left(A^I_{\psi_{j,m}} \cap A^R_{\psi_{j,m}} \cap A^C_{\psi_{j,m}}\right), \qquad A_g := A^\emptyset \cap A^F \cap \bigcap_{j=1}^{J} A_{g,j}.$$

Then $\mathbb{P}(A_g) = 1$. Fix $\omega \in A_g$, and suppress ω from the notation. For fixed j and m, Lemma 3.4 and Assumption 4.2 imply

$$\Phi_{N,j}(x_m) \to \Phi_{\infty,j}(x_m).$$

By Assumption 2.6, for each fixed j, $\Phi_{N,j}$ is finite-valued on $\mathcal{X}$ for all sufficiently large N. The same modulus argument as in Lemma 3.3 gives, for all sufficiently large N,

$$|\Phi_{N,j}(x) - \Phi_{N,j}(x')| \leq \rho_{g_j}(|x - x'|), \qquad \forall x, x' \in \mathcal{X},$$

and

$$|\Phi_{\infty,j}(x) - \Phi_{\infty,j}(x')| \leq \rho_{g_j}(|x - x'|), \qquad \forall x, x' \in \mathcal{X}.$$

Using a finite δ-net in D, exactly as in the proof of Theorem 3.5, yields

$$\sup_{x \in \mathcal{X}} |\Phi_{N,j}(x) - \Phi_{\infty,j}(x)| \to 0.$$

Since $J < \infty$, taking the maximum over $j = 1, \dots, J$ gives

$$\max_{1 \leq j \leq J} \sup_{x \in \mathcal{X}} |\Phi_{N,j}(x) - \Phi_{\infty,j}(x)| \to 0.$$

□

Proof of Theorem 4.4. Define $\varepsilon_N := \max_{1 \leq j \leq J} \sup_{x \in \mathcal{X}} |\Phi_{N,j}(x) - \Phi_{\infty,j}(x)|$. By Lemma 4.3, $\varepsilon_N \to 0$.

We first prove the outer inclusion ($\operatorname{Limsup}_{N \to \infty} \Gamma_N \subseteq \Gamma_\infty$). Let $\{N_k\}$ be an arbitrary subsequence, and let $x_{N_k} \in \Gamma_{N_k}$. By compactness of $\mathcal{X}$, passing to a further subsequence if necessary, we assume $x_{N_k} \to \bar{x} \in \mathcal{X}$. Since $x_{N_k} \in \Gamma_{N_k}$, we have $\Phi_{N_k,j}(x_{N_k}) \leq 0$ for $j = 1, \dots, J$. Hence, for each j,

$$\Phi_{\infty,j}(x_{N_k}) = \Phi_{\infty,j}(x_{N_k}) - \Phi_{N_k,j}(x_{N_k}) + \Phi_{N_k,j}(x_{N_k}) \leq |\Phi_{\infty,j}(x_{N_k}) - \Phi_{N_k,j}(x_{N_k})| \leq \varepsilon_{N_k}.$$

Taking the limit as $k \to \infty$ and using the continuity of $\Phi_{\infty,j}$ yields $\Phi_{\infty,j}(\bar{x}) \leq 0$. Therefore, $\bar{x} \in \Gamma_\infty$. This proves $\operatorname{Limsup}_{N \to \infty} \Gamma_N \subseteq \Gamma_\infty$.

For the inner inclusion ($\Gamma_\infty \subseteq \operatorname{Liminf}_{N \to \infty} \Gamma_N$), let $x \in \Gamma_\infty$. We distinguish two cases. First, suppose $\Phi_{\infty,j}(x) < 0$ for all $j = 1, \dots, J$. Set $\eta := -\max_{1 \leq j \leq J} \Phi_{\infty,j}(x) > 0$. Since $\varepsilon_N \to 0$, there exists N_0 such that for all $N \geq N_0$, $\varepsilon_N < \eta/2$. Then, for every $j = 1, \dots, J$ and every $N \geq N_0$,

$$\Phi_{N,j}(x) \leq \Phi_{\infty,j}(x) + |\Phi_{N,j}(x) - \Phi_{\infty,j}(x)| \leq -\eta + \varepsilon_N < -\eta/2 < 0.$$

Thus $x \in \Gamma_N$ for all $N \geq N_0$, which implies $x \in \operatorname{Liminf}_{N \to \infty} \Gamma_N$.

Next, suppose $\max_{1 \leq j \leq J} \Phi_{\infty,j}(x) = 0$. Let x° be the Slater point from Assumption 4.2 satisfying $\Phi_{\infty,j}(x^\circ) < 0$ for all $j = 1, \dots, J$. For $\lambda \in (0, 1)$, define $x^\lambda := (1 - \lambda)x + \lambda x^\circ$. Since $\mathcal{X}$ is convex, $x^\lambda \in \mathcal{X}$. By convexity of each $\Phi_{\infty,j}$,

$$\Phi_{\infty,j}(x^\lambda) \leq (1 - \lambda)\Phi_{\infty,j}(x) + \lambda \Phi_{\infty,j}(x^\circ) < 0$$

for $j = 1, \dots, J$. Fix $\lambda \in (0, 1)$. By the first case (strictly feasible interior point), there exists N_λ such that $x^\lambda \in \Gamma_N$ for all $N \geq N_\lambda$. Consequently, $\operatorname{dist}(x, \Gamma_N) \leq |x - x^\lambda|$ for all $N \geq N_\lambda$. Taking the limit superior as $N \to \infty$ yields

$$\limsup_{N\to\infty} \operatorname{dist}(x, \Gamma_N) \leq |x - x^\lambda|.$$

Since $x^\lambda = (1-\lambda)x + \lambda x^\circ$, we have $|x^\lambda - x| = \lambda |x^\circ - x| \to 0$ as $\lambda \downarrow 0$. Hence $\limsup_{N\to\infty} \operatorname{dist}(x, \Gamma_N) = 0$. By the distance characterization of the inner limit, $x \in \operatorname{Liminf}_{N\to\infty} \Gamma_N$. Combining the two inclusions yields $\Gamma_N \xrightarrow{\text{PK}} \Gamma_\infty$. □

Proof of Theorem 4.5. Fix $\omega \in A_*$. Then

$$\sup_{x\in\mathcal{X}} |\mathcal{R}_N(f)(x) - \mathcal{R}_\infty(f)(x)| \to 0$$

by Theorem 3.5, and

$$\max_{1\leq j\leq J} \sup_{x\in\mathcal{X}} |\Phi_{N,j}(x) - \Phi_{\infty,j}(x)| \to 0$$

by Lemma 4.3. Moreover,

$$\Gamma_N \xrightarrow{\text{PK}} \Gamma_\infty$$

by Theorem 4.4.

By the Slater condition, there exists $\bar{x} \in \mathcal{X}$ and $\eta > 0$ such that

$$\Phi_{\infty,j}(\bar{x}) \leq -\eta, \qquad j = 1, \dots, J.$$

Since

$$\max_{1\leq j\leq J} |\Phi_{N,j}(\bar{x}) - \Phi_{\infty,j}(\bar{x})| \to 0,$$

we have $\Phi_{N,j}(\bar{x}) \leq -\eta/2$ for all j and all sufficiently large N. Hence $\bar{x} \in \Gamma_N$ eventually, so $\Gamma_N \neq \emptyset$ eventually.

We first establish the upper bound ($\limsup_{N\to\infty} \hat{v}_N^c \leq v_\infty^c$). Let $x^* \in X_c^* \subset \Gamma_\infty$ (which is valid since $X_c^* \neq \emptyset$). By Theorem 4.4, $\Gamma_\infty \subseteq \operatorname{Liminf}_{N\to\infty} \Gamma_N$, hence there exists a sequence $x_N \in \Gamma_N$ such that $x_N \to x^*$. Since $\hat{v}_N^c = \inf_{x\in\Gamma_N} \mathcal{R}_N(f)(x)$, we have $\hat{v}_N^c \leq \mathcal{R}_N(f)(x_N)$. Therefore,

$$\begin{aligned}\limsup_{N\to\infty} \hat{v}_N^c &\leq \limsup_{N\to\infty} \mathcal{R}_N(f)(x_N) \\ &\leq \limsup_{N\to\infty} \Big(\mathcal{R}_\infty(f)(x_N) + \sup_{z\in\mathcal{X}} |\mathcal{R}_N(f)(z) - \mathcal{R}_\infty(f)(z)| \Big) = \mathcal{R}_\infty(f)(x^*) = v_\infty^c.\end{aligned}$$

For the lower bound ($\liminf_{N\to\infty} \hat{v}_N^c \geq v_\infty^c$), let $\{N_m\}$ be a subsequence such that $\hat{v}_{N_m}^c \to \liminf_{N\to\infty} \hat{v}_N^c$. For each m, choose $\hat{x}_{N_m} \in \hat{X}_{N_m}^c$ satisfying $\mathcal{R}_{N_m}(f)(\hat{x}_{N_m}) = \hat{v}_{N_m}^c$. By compactness of $\mathcal{X}$, passing to a subsequence if necessary, we assume $\hat{x}_{N_m} \to \bar{x} \in \mathcal{X}$. Since $\hat{x}_{N_m} \in \Gamma_{N_m}$, the outer inclusion $\operatorname{Limsup}_{N\to\infty} \Gamma_N \subseteq \Gamma_\infty$ implies $\bar{x} \in \Gamma_\infty$. Hence

$v^c_\infty \le \mathcal{R}_\infty(f)(\bar{x})$. Moreover,

$$\begin{aligned}\left|\hat{v}^c_{N_m} - \mathcal{R}_\infty(f)(\bar{x})\right| &= |\mathcal{R}_{N_m}(f)(\hat{x}_{N_m}) - \mathcal{R}_\infty(f)(\bar{x})| \\ &\le \sup_{z\in\mathcal{X}} |\mathcal{R}_{N_m}(f)(z) - \mathcal{R}_\infty(f)(z)| + |\mathcal{R}_\infty(f)(\hat{x}_{N_m}) - \mathcal{R}_\infty(f)(\bar{x})|\,.\end{aligned}$$

The right-hand side converges to 0. Therefore, $\hat{v}^c_{N_m} \to \mathcal{R}_\infty(f)(\bar{x})$, and thus $v^c_\infty \le \liminf_{N\to\infty} \hat{v}^c_N$. Combining the two bounds yields $\hat{v}^c_N \to v^c_\infty$.

For the outer inclusion of constrained minimizers, let $\{N_k\}$ be an arbitrary subsequence and let $\hat{x}_{N_k} \in \hat{X}^c_{N_k}$. By compactness of $\mathcal{X}$, passing to a further subsequence if necessary, we assume $\hat{x}_{N_k} \to \bar{x} \in \mathcal{X}$. Since $\hat{x}_{N_k} \in \Gamma_{N_k}$, Theorem 4.4 implies $\bar{x} \in \Gamma_\infty$. Moreover, since $\hat{x}_{N_k} \in \hat{X}^c_{N_k}$, $\mathcal{R}_{N_k}(f)(\hat{x}_{N_k}) = \hat{v}^c_{N_k}$. Since $\hat{v}^c_{N_k} \to v^c_\infty$,

$$\begin{aligned}|\mathcal{R}_{N_k}(f)(\hat{x}_{N_k}) - \mathcal{R}_\infty(f)(\bar{x})| \le &\sup_{z\in\mathcal{X}} |\mathcal{R}_{N_k}(f)(z) - \mathcal{R}_\infty(f)(z)| \\ &+ |\mathcal{R}_\infty(f)(\hat{x}_{N_k}) - \mathcal{R}_\infty(f)(\bar{x})| \to 0.\end{aligned}$$

Hence $\mathcal{R}_{N_k}(f)(\hat{x}_{N_k}) \to \mathcal{R}_\infty(f)(\bar{x})$. Combining the limits yields $\mathcal{R}_\infty(f)(\bar{x}) = v^c_\infty$. Since $\bar{x} \in \Gamma_\infty$, we conclude $\bar{x} \in X^*_c$. As the subsequence and cluster point were arbitrary, $\mathrm{Limsup}_{N\to\infty} \hat{X}^c_N \subseteq X^*_c$. □

C Verification Proofs

Proof of Proposition 5.1. Fix $\epsilon > 0$. Let

$$E_N(\epsilon) := \{\mathbb{D}(P^{\mathrm{ref}}_N, P^\circ) > \epsilon\}.$$

By assumption, $\sum_{N=1}^\infty \mathbb{P}(E_N(\epsilon)) < \infty$. By the Borel-Cantelli lemma,

$$\mathbb{P}\left(\limsup_{N\to\infty} E_N(\epsilon)\right) = 0,$$

where

$$\limsup_{N\to\infty} E_N(\epsilon) = \bigcap_{n=1}^\infty \bigcup_{k=n}^\infty E_k(\epsilon).$$

Let

$$\mathcal{N}_\epsilon := \limsup_{N\to\infty} E_N(\epsilon).$$

Then $\mathbb{P}(\mathcal{N}_\epsilon) = 0$. For any sample path $\omega \notin \mathcal{N}_\epsilon$, there exists a finite integer $N_0(\omega, \epsilon)$ such that

$$\mathbb{D}(P^{\mathrm{ref}}_N(\omega), P^\circ) \le \epsilon$$

for all $N \geq N_0(\omega, \epsilon)$. To establish almost sure convergence independent of ϵ, consider the countable sequence $\epsilon_m = 1/m$, $m \in \mathbb{N}$, and define

$$\mathcal{N} := \bigcup_{m=1}^{\infty} \mathcal{N}_{1/m}.$$

By countable subadditivity,

$$\mathbb{P}(\mathcal{N}) \leq \sum_{m=1}^{\infty} \mathbb{P}(\mathcal{N}_{1/m}) = 0.$$

For any $\omega \notin \mathcal{N}$ and any $\epsilon > 0$, choose $m \in \mathbb{N}$ such that $1/m < \epsilon$. Since $\omega \notin \mathcal{N}_{1/m}$, there exists an integer $N_0(\omega, 1/m)$ such that for all $N \geq N_0(\omega, 1/m)$,

$$\mathbb{D}(P_N^{\text{ref}}(\omega), P^\circ) \leq \frac{1}{m} < \epsilon.$$

Therefore,

$$\mathbb{D}(P_N^{\text{ref}}(\omega), P^\circ) \to 0.$$

Since $\mathbb{P}(\mathcal{N}) = 0$, we conclude that

$$\mathbb{D}(P_N^{\text{ref}}, P^\circ) \to 0$$

almost surely. By the condition given in the proposition, we obtain

$$\int_\Xi \psi \, dP_N^{\text{ref}} \to \int_\Xi \psi \, dP^\circ$$

almost surely for every $\psi \in \mathcal{F}$. □

Proof of Proposition 5.2. Fix $\psi \in \mathcal{F}$ and $P \in \mathcal{P}_N^{\text{c}}$. Take $L_\psi < \infty$ from the c-Lipschitz condition. For every $\varepsilon > 0$, choose $\gamma_\varepsilon \in \Gamma(P, P_N^{\text{ref}})$ such that

$$\int_{\Xi \times \Xi} c(\xi, \zeta) \, d\gamma_\varepsilon(\xi, \zeta) \leq \mathrm{OT}_c(P, P_N^{\text{ref}}) + \varepsilon \leq \delta_N + \varepsilon.$$

Use the marginals of γ_ε to write the difference of integrals as one joint integral:

$$\int_\Xi \psi \, dP - \int_\Xi \psi \, dP_N^{\text{ref}} = \int_{\Xi \times \Xi} (\psi(\xi) - \psi(\zeta)) \, d\gamma_\varepsilon(\xi, \zeta).$$

Apply the c-Lipschitz bound:

$$\left| \int_\Xi \psi \, dP - \int_\Xi \psi \, dP_N^{\text{ref}} \right| \leq \int_{\Xi \times \Xi} |\psi(\xi) - \psi(\zeta)| \, d\gamma_\varepsilon(\xi, \zeta) \leq L_\psi \int_{\Xi \times \Xi} c(\xi, \zeta) \, d\gamma_\varepsilon(\xi, \zeta) \leq L_\psi(\delta_N + \varepsilon).$$

Let $\varepsilon \downarrow 0$ to obtain:

$$\left| \int_\Xi \psi \, dP - \int_\Xi \psi \, dP_N^{\mathrm{ref}} \right| \le L_\psi \delta_N.$$

Take the supremum over $P \in \mathcal{P}_N^c$:

$$\sup_{P \in \mathcal{P}_N^c} \left| \int_\Xi \psi \, dP - \int_\Xi \psi \, dP_N^{\mathrm{ref}} \right| \le L_\psi \delta_N \to 0.$$

Thus Assumption 2.4 holds.

To prove $P_N^{\mathrm{ref}} \in \mathcal{P}_N^c$, use the diagonal coupling $(\mathrm{id}, \mathrm{id})_\# P_N^{\mathrm{ref}}$. Since $c(\xi, \xi) = 0$,

$$\mathrm{OT}_c(P_N^{\mathrm{ref}}, P_N^{\mathrm{ref}}) \le \int_\Xi c(\xi, \xi) \, dP_N^{\mathrm{ref}}(\xi) = 0 \le \delta_N.$$

Hence $P_N^{\mathrm{ref}} \in \mathcal{P}_N^c$, so $\mathcal{P}_N^c \neq \emptyset$. Conclude Assumptions 2.4 and 2.5. □

D MADRO Proofs

MADRO Dual Reformulation Derivation. For fixed x, the outer aggregation problem over $\boldsymbol{w} \in \mathcal{W}_N$ is

$$\sup_{\boldsymbol{w} \in \mathcal{W}_N} \sum_{k=1}^K w_k \, v_{N,k}(x). \tag{2}$$

Recall that the ℓ_1-weight ambiguity set is given by $\mathcal{W}_N = \{\boldsymbol{w} \in \Delta_K : \|\boldsymbol{w} - \hat{\boldsymbol{w}}_N\|_1 \le \delta_{w,N}\}$. Introduce auxiliary variables $u_k \ge 0$ to linearize $|w_k - \hat{w}_{N,k}|$ such that

$$-w_k + \hat{w}_{N,k} \le u_k, \qquad w_k - \hat{w}_{N,k} \le u_k, \qquad k = 1, \dots, K.$$

Then (2) is equivalent to the primal linear program:

$$\begin{aligned} \sup_{\boldsymbol{w}, \boldsymbol{u}} \quad & \sum_{k=1}^K v_{N,k}(x) w_k \\ \text{s.t.} \quad & \sum_{k=1}^K w_k = 1, \\ & w_k - u_k \le \hat{w}_{N,k}, \qquad k = 1, \dots, K, \\ & -w_k - u_k \le -\hat{w}_{N,k}, \qquad k = 1, \dots, K, \\ & \sum_{k=1}^K u_k \le \delta_{w,N}, \\ & w_k \ge 0, \quad u_k \ge 0, \qquad k = 1, \dots, K. \end{aligned} \tag{3}$$

The feasible set is nonempty because $\hat{\boldsymbol{w}}_N \in \Delta_K$ and $\delta_{w,N} \ge 0$, with feasible point $\boldsymbol{w} = \hat{\boldsymbol{w}}_N, \boldsymbol{u} = \mathbf{0}$. It is also compact. Hence, if $v_{N,k}(x) < \infty$ for all k, the primal linear program has a finite optimal value, and finite-dimensional LP strong duality applies.

We form the Lagrangian. Let $\eta \in \mathbb{R}$ be the dual variable for $\sum_k w_k = 1$, $\alpha_k \geq 0$ for $w_k - u_k \leq \hat{w}_{N,k}$, $\beta_k \geq 0$ for $-w_k - u_k \leq -\hat{w}_{N,k}$, and $\nu \geq 0$ for $\sum_k u_k \leq \delta_{w,N}$. The Lagrangian is:

$$\begin{aligned}
\mathcal{L}(\boldsymbol{w}, \boldsymbol{u}, \eta, \boldsymbol{\alpha}, \boldsymbol{\beta}, \nu) &= \sum_{k=1}^{K} v_{N,k}(x) w_k + \eta\Big(1 - \sum_{k=1}^{K} w_k\Big) + \nu\Big(\delta_{w,N} - \sum_{k=1}^{K} u_k\Big) \\
&\quad + \sum_{k=1}^{K} \alpha_k(\hat{w}_{N,k} - w_k + u_k) + \sum_{k=1}^{K} \beta_k(-\hat{w}_{N,k} + w_k + u_k) \\
&= \eta + \delta_{w,N}\nu + \sum_{k=1}^{K} \hat{w}_{N,k}(\alpha_k - \beta_k) \\
&\quad + \sum_{k=1}^{K} w_k\big(v_{N,k}(x) - \eta - \alpha_k + \beta_k\big) + \sum_{k=1}^{K} u_k\big(-\nu + \alpha_k + \beta_k\big).
\end{aligned}$$

Since $w_k \geq 0$ and $u_k \geq 0$, the supremum of the Lagrangian over $(\boldsymbol{w}, \boldsymbol{u})$ is finite if and only if:

$$v_{N,k}(x) - \eta - \alpha_k + \beta_k \leq 0, \qquad -\nu + \alpha_k + \beta_k \leq 0, \qquad k = 1, \ldots, K.$$

Equivalently,

$$\alpha_k - \beta_k + \eta \geq v_{N,k}(x), \qquad \alpha_k + \beta_k \leq \nu, \qquad k = 1, \ldots, K.$$

Therefore the dual problem is

$$\begin{aligned}
\inf_{\eta, \nu, \boldsymbol{\alpha}, \boldsymbol{\beta}} \quad & \eta + \delta_{w,N}\nu + \sum_{k=1}^{K} \hat{w}_{N,k}\alpha_k - \sum_{k=1}^{K} \hat{w}_{N,k}\beta_k \\
\text{s.t.} \quad & \alpha_k - \beta_k + \eta \geq v_{N,k}(x), \qquad k = 1, \ldots, K, \\
& \alpha_k + \beta_k \leq \nu, \qquad k = 1, \ldots, K, \\
& \alpha_k \geq 0,\ \beta_k \geq 0, \qquad k = 1, \ldots, K, \\
& \nu \geq 0, \qquad \eta \in \mathbb{R}.
\end{aligned} \tag{4}$$

Equivalently, introduce the substitution

$$t_k := \alpha_k - \beta_k + \nu, \qquad k = 1, \ldots, K. \tag{5}$$

Since $\alpha_k, \beta_k \geq 0$ and $\alpha_k + \beta_k \leq \nu$, we have

$$0 \leq t_k \leq 2\nu, \qquad k = 1, \ldots, K.$$

Moreover, $\alpha_k - \beta_k = t_k - \nu$. Substituting this into (4) and using $\sum_{k=1}^{K} \hat{w}_{N,k} = 1$, we obtain the final dual form presented in the main text:

$$\begin{aligned}
\inf_{\eta,\nu,\boldsymbol{t}} \quad & \eta + \nu(\delta_{w,N} - 1) + \sum_{k=1}^{K} \hat{w}_{N,k} t_k \\
\text{s.t.} \quad & t_k + \eta - \nu \geq v_{N,k}(x), \qquad k = 1, \ldots, K, \\
& 0 \leq t_k \leq 2\nu, \qquad k = 1, \ldots, K, \\
& \nu \geq 0, \qquad \eta \in \mathbb{R}.
\end{aligned} \tag{6}$$

□

Proof of Lemma 6.4. Let $A_\psi^{MA,R} := A^W \cap \bigcap_{k=1}^K A_{\psi,k}^R$. On $A_\psi^{MA,R}$, $\hat{\boldsymbol{w}}_N \to \boldsymbol{w}^\circ$ and $\int_\Xi \psi \, dP_{N,k}^{\text{ref}} \to \int_\Xi \psi \, dP_k^\circ$. Because limits commute with finite sums, $\int_\Xi \psi \, dP_N^{MA,\text{ref}} = \sum_{k=1}^K \hat{w}_{N,k} \int_\Xi \psi \, dP_{N,k}^{\text{ref}} \to \sum_{k=1}^K w_k^\circ \int_\Xi \psi \, dP_k^\circ = \int_\Xi \psi \, dP^{MA,\circ}$. □

Proof of Theorem 6.5. Fix $\psi \in \mathcal{F}$ and define

$$A_\psi^{MA,C} := A_\psi^{MA,R} \cap \left(\bigcap_{k=1}^K A_{\psi,k}^C \right) \cap A^W.$$

Then $\mathbb{P}(A_\psi^{MA,C}) = 1$. Fix $\omega \in A_\psi^{MA,C}$ and suppress it. First note that

$$\sup_{\boldsymbol{w}_N \in \mathcal{W}_N} \|\boldsymbol{w}_N - \boldsymbol{w}^\circ\|_1 \leq \sup_{\boldsymbol{w}_N \in \mathcal{W}_N} \|\boldsymbol{w}_N - \hat{\boldsymbol{w}}_N\|_1 + \|\hat{\boldsymbol{w}}_N - \boldsymbol{w}^\circ\|_1 \to 0.$$

For $k = 1, \ldots, K$, set

$$a_{N,k} := \int_\Xi \psi \, dP_{N,k}^{\text{ref}}, \qquad a_k^\circ := \int_\Xi \psi \, dP_k^\circ.$$

On $A_\psi^{MA,R}$, $a_{N,k} \to a_k^\circ$ for every k. Since $K < \infty$, there exists $C_\psi < \infty$ such that

$$\max_{1 \leq k \leq K} |a_{N,k}| \leq C_\psi$$

for all sufficiently large N. Let $P^{MA} = \sum_{k=1}^K w_{N,k} P_{N,k} \in \mathcal{P}_N^{MA}$. The difference is bounded as:

$$\begin{aligned}
\left| \int_\Xi \psi \, dP^{MA} - \int_\Xi \psi \, dP_N^{MA,\text{ref}} \right| &\leq \sum_{k=1}^K w_{N,k} \left| \int_\Xi \psi \, dP_{N,k} - a_{N,k} \right| + \left| \sum_{k=1}^K w_{N,k} a_{N,k} - \sum_{k=1}^K \hat{w}_{N,k} a_{N,k} \right| \\
&\leq \max_{1 \leq k \leq K} \sup_{P_k \in \mathcal{P}_{N,k}} \left| \int_\Xi \psi \, dP_k - a_{N,k} \right| + C_\psi \|\boldsymbol{w}_N - \hat{\boldsymbol{w}}_N\|_1.
\end{aligned}$$

Taking the supremum over $P^{MA} \in \mathcal{P}_N^{MA}$, both terms on the right-hand side converge to

0 by Assumptions 6.2 and 6.3, as well as the fact that $K < \infty$. Hence:

$$\sup_{P^{MA} \in \mathcal{P}_N^{MA}} \left| \int_\Xi \psi \, dP^{MA} - \int_\Xi \psi \, dP_N^{MA,\mathrm{ref}} \right| \to 0.$$

This proves the stated ambiguity-set shrinkage toward $P_N^{MA,\mathrm{ref}}$. □

References

[1] Erick Delage and Yinyu Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3): 595–612, 2010. doi: 10.1287/opre.1090.0741.

[2] Wolfram Wiesemann, Daniel Kuhn, and Melvyn Sim. Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376, 2014. doi: 10.1287/opre.2014.1314.

[3] Aharon Ben-Tal, Dick den Hertog, Anja De Waegenaere, Bertrand Melenberg, and Gijs Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357, 2013. doi: 10.1287/mnsc.1120.1641.

[4] P. Mohajerin Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the wasserstein metric. *Mathematical Programming*, 171(1-2):115–166, 2018.

[5] R. Gao and A. J. Kleywegt. Distributionally robust stochastic optimization with wasserstein distance. *Mathematics of Operations Research*, 48(2):603–655, 2023.

[6] Jose Blanchet and Karthyek Murthy. Quantifying distributional model risk via optimal transport. *Mathematics of Operations Research*, 44(2):565–600, 2019. doi: 10.1287/moor.2018.0936.

[7] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming.* SIAM, 2009.

[8] H. Rahimian and S. Mehrotra. Distributionally robust optimization: A review. *SIAM Review*, 61(4):631–669, 2019.

[9] Jose Blanchet, Daniel Kuhn, Jiajin Li, and Bahar Taşkesen. Unifying distributionally robust optimization via optimal transport theory, 2023.

[10] Jitka Dupačová and Roger J-B Wets. Asymptotic properties of estimators in stochastic programming. *Annals of Operations Research*, 14(1):265–294, 1988.

[11] Alexander Shapiro. Quantitative stability in stochastic programming. *Mathematical Programming*, 67(1-3):99–108, 1994.

[12] Alexander Shapiro and Tito Homem-de Mello. On the rate of convergence of optimal solutions of monte carlo approximations of stochastic programs. *SIAM journal on optimization*, 11(1):70–86, 2000.

[13] Svetlozar T Rachev and Werner Römisch. Quantitative stability in stochastic programming: The method of probability metrics. *Mathematics of Operations Research*, 27(4):792–818, 2002.

[14] Werner Römisch. Stability of stochastic programming problems. In Andrzej Ruszczyński and Alexander Shapiro, editors, *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*, pages 483–554. Elsevier, 2003. doi: 10.1016/S0927-0507(03)10008-4.

[15] N. Fournier and A. Guillin. On the rate of convergence in wasserstein distance. *Probability Theory and Related Fields*, 162:707–738, 2015.

[16] Haikui Sun and Huifu Xu. Convergence of empirical distributionally robust optimization and equilibrium problems. *Mathematics of Operations Research*, 41(1): 78–106, 2016.

[17] Alois Pichler and Huifu Xu. Quantitative stability analysis for minimax distributionally robust risk optimization. *Mathematical Programming*, 171(1-2):221–247, 2018.

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

[19] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.